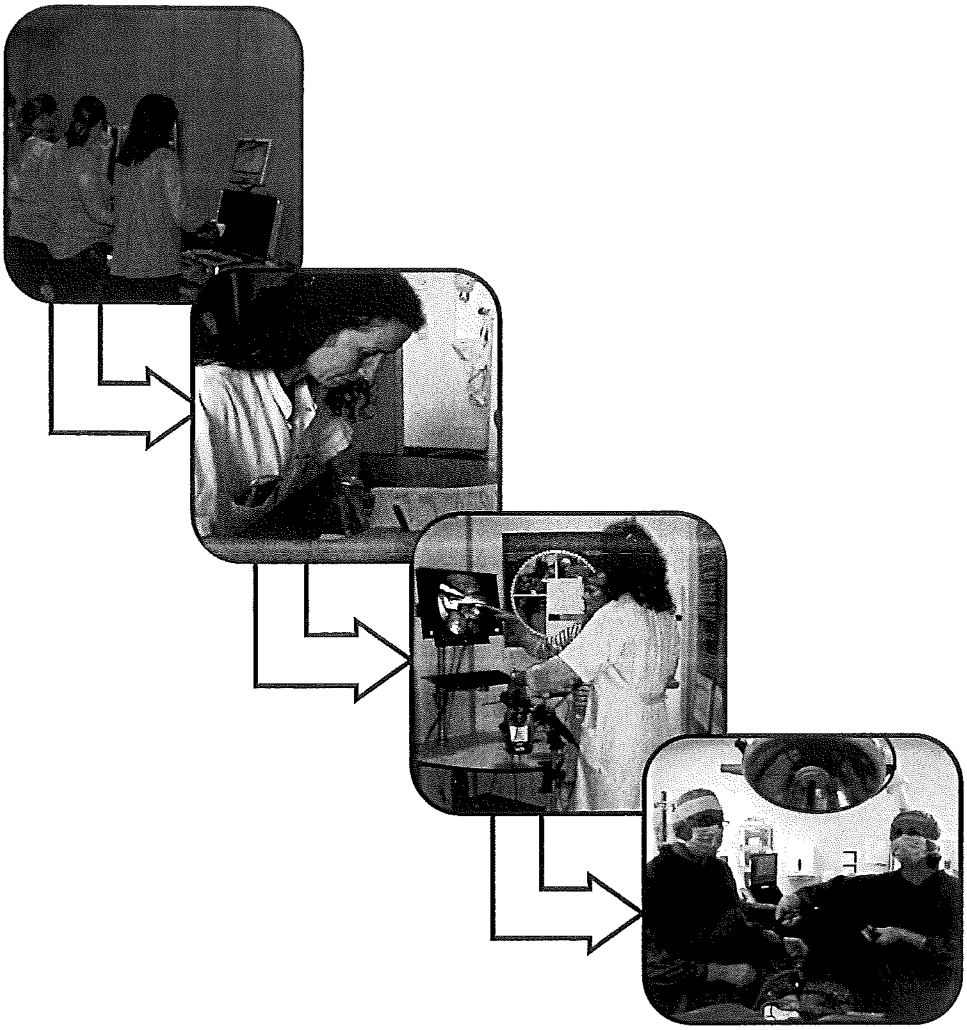
FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Jeanett Strandbygaard

Development and validation of a structured curriculum in basic laparoscopy

- A four-step model



Proper Planning Prevents Poor Performance

Development and validation of a structured curriculum in basic laparoscopy

- A four-step model

PhD thesis

Jeanett Strandbygaard

2012

Table of Contents

Articles

This PhD thesis is based on four articles generated from four studies:

I. *'Development and validation of a theoretical test in basic laparoscopy'.*
Strandbygaard J, Maagaard M, Larsen CR, Schouenborg L, Ottosen C, Ringsted C, Grantcharov T, Ottesen B, Sorensen JL.
The study is accepted in Surgical Endoscopy, September, 2012

II. *'Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial'.*
Strandbygaard ~~Oestergaard~~ J, Bjerrum F, Maagaard M, Winkel P, Larsen CR, Ringsted C, Gluud C, Grantcharov T, Ottesen B, Sorensen JL.
The study is accepted in Annals of Surgery, September, 2012

III. *'Can both residents and chief physicians assess surgical skills?'*
Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL.
Surgical Endoscopy 2012 Jul;26(7):2054-60

IV. *'A structured four-step curriculum in basic laparoscopy: development and validation'.*
Strandbygaard J, Bjerrum F, Maagaard M, Larsen CR, Ottesen B, Sorensen JL
The study was submitted to Journal of Surgical Education, November, 2012

The articles are referenced with Roman numbers. The previously published article is reproduced with permission from Springer inc. (Surgical Endoscopy).

Academic advisors

- **Bent Ottesen**, Professor, DMSc, Managing Director, Juliane Marie Centre, Rigshospitalet, Copenhagen University Hospital, Denmark
- **Jette Led Sørensen**, MD, MMEd, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet, Copenhagen University Hospital, Denmark
- **Christian Rifbjerg Larsen**, MD, PhD, Department of Obstetrics and Gynecology, Hillerød Hospital, Hillerød, Denmark
- **Teodor Grantcharov**, MD, PhD, FACS, Department of Surgery, St. Michael's Hospital, Toronto Canada

Assessment committee

- **Kári Mikines** (chairman), MD, DMSc, Department of Urology, Herlev Hospital, Copenhagen University Hospital, Denmark

- **Berit Eika**, MD, Professor, PhD, MHPH, Associate Dean of Education, Aarhus University, Denmark

- **Richard Reznick**, MD, Med, FRCSC, FACS, FRCSEd (hon), FRCSI (hon), Dean of Faculty of Health Sciences, CEO at Southeastern Ontario Academic Medical Organization, Canada

Summary in English

Aim

The overall aim of this PhD thesis was to develop and validate a four-step curriculum in basic laparoscopy for residents in obstetrics and gynecology.

Background

Important work has been done to validate laparoscopic simulators as viable methods for teaching technical skills outside the operating room. The next step is to integrate simulation training into a curriculum in order to plan and standardize education. However, integration of structured simulation training remains a problem. This is mainly due to lack of knowledge on best practice within training methods and the ideal design of a curriculum.

Through four collective studies this thesis aimed to develop and validate a four-step curriculum in basic laparoscopy that adhered to the current standards of proficiency-based training and distributed and deliberate practice. The core competencies revolved around basic laparoscopic theory and practice, integrating a knowledge component, a technical skills component and, as a novelty compared with the existing literature, an operational component.

In the first study, a validation study, a test containing 37 multiple-choice questions was developed through interviews with four experts in laparoscopy and subsequently through a Delphi audit involving regional laparoscopic surgeons. The test showed good construct validity and no evidence of differential item functioning, meaning no questions had to be excluded before the test could be taken into use.

The second study, a randomized trial with 99 participants, demonstrated that instructor feedback significantly increased efficiency when training a complex operational task on a virtual reality simulator. The intervention group, who received standardized instructor feedback for 20–30 minutes, used half the amount of time and number of repetitions to reach proficiency level compared with the control group. Three participants in the control group dropped out due to the frustration of not being able to complete the task. The control group, however, reached the proficiency level with a higher performance score than the intervention group did.

The third study, a cohort study, demonstrated that senior residents and chief physicians in gynecology were equally able to assess laparoscopic operations on the basis of an assessment scale. Neither group revealed difficulties when using the assessment scale.

In the fourth study, a prospective observational study, a curriculum with four steps was developed: Step 1) 1-day course in basic laparoscopy, Step 2) A multiple-choice test in basic laparoscopy, Step 3) Structured virtual reality simulation training, and Step 4) Operation on a patient with subsequent formative assessment.
A cohort of 52 first-year residents signed up for the study. There was 100% attendance during Step 1, and 55% completed all four steps. There were several reasons for dropping out; the main reason was the voluntary nature of the study. Additionally, some logistical problems were identified for Step 4.

Conclusion

We propose a flexible four-step curriculum in basic laparoscopy containing a course component, a knowledge component, a virtual reality training component involving instructor feedback, and an operational component with subsequent formative assessment.

Dansk resumé

Formål
Afhandlingens overordnede formål var at udvikle og validere et struktureret træningsprogram i basal kikkertkirurgi for læger under uddannelse til speciallæger i gynækologi og obstetrik.

Baggrund
Gennem de sidste ti år er brugen af virtual reality simulation inden for optræning af kikkertkirurgiske færdigheder markant øget. Denne udvikling er sket på baggrund af flere randomiserede forsøg, der har vist, at man kan overføre færdigheder opnået på en virtual reality simulator til egentlige operationer på patienter. Således behøver man ikke længere oplæres ved operationer på patienter, det kan gøres på en simulator. Det medfører en del udfordringer når simulationstræning skal implementeres i trænings- og uddannelsesprogrammer, idet der endnu foreligger tilstrækkelig viden om omfang, træningsmetoder og kompetencevurdering.

Forskningsstrategi
Gennem fire delstudier blev der udviklet og valideret et struktureret træningsprogram i basal kikkertkirurgi indeholdende en kursuskomponent, en teoretisk komponent, en færdighedstrænings komponent samt en operations komponent. Studierne inddrager aktuelle standarder inden for udvikling af træningsprogrammer samt veldokumenteret teori inden for oplæring af motoriske færdigheder.

I det første studie, et valideringsstudie, blev der gennem interview med fire eksperter i kikkertkirurgi og samt efterfølgende Delphi audit blandt 12 overlæger i gynækologi fra Region Hovedstanden samt Region Sjælland, udviklet en test med 37 multiple-choice spørgsmål. Testen viste at kunne skelne mellem medicinstuderende og læger på forskelligt uddannelsesniveau, det vil sige, at der var god construct validitet. Desuden viste testen ikke tegn på differential item functioning, hvilket betyder at ingen af spørgsmålene skulle ekskluderes før testen kunne tages i brug.

I studie to, et prospektivt randomiseret forsøg, fandt vi at 20-30 minutters feedback fra en instruktør fordobler effektiviteten i forhold til tid og antal repetitioner ved træning af et kikkertkirurgisk operationsmodul på en virtual reality simulator. Dette vurderet i forhold til kontrolgruppen som ikke modtog feedback. Der var tre personer fra kontrolgruppen der udgik på grund af frustration over ikke at kunne gennemføre operationsmodulet uden feedback. Kontrolgruppen opnåede en højere samlet 'udførelsesscore' i forhold til interventionsgruppen.

I det tredje studie, et prospektivt kohortestudie, fandt vi at læger med forskelligt kirurgisk kompetenceniveau (10 førstereservelæger og 12 overlæger) evaluerede kikkertkirurgiske operationer ens ud fra en valideret kompetencevurderingsskala. Dog evaluerede de ikke med samme resultat som to eksperter.

I det fjerde studie, et prospektivt observationsstudie, blev et træningsprogram bestående af fire trin udviklet. De fire trin var følgende: trin 1) Et ét-dagskursus i basal kikkertkirurgi, trin 2) En teoretisk test i basal kikkertkirurgi, trin 3) Struktureret træning af basale koordinerings moduler samt operationsmoduler på en virtual reality

simulator og trin 4) En operation på en patient med efterfølgende kompetenceevaluering.
En kohorte på 52 introduktionslæger i obstetrik og gynækologi deltog, og 100% gennemførte trin 1 og 55% gennemførte alle fire trin. Der var flere grunde til frafald; hovedårsagen var at studiet var frivilligt, og at der dermed ikke blev givet fri fra afdelinger til deltagelse. Desuden var der nogle logistiske problemer i trin fire i form af manglende planlægning omkring at få en operation.

Konklusion
Vi foreslår at en fleksibel model til et struktureret træningsprogram i basal kikkertkirurgi bestående af en kursus komponent, en teoretisk komponent, en praktisk komponent samt en operationel komponent bliver indført i uddannelsen af læger i gynækologi og obstetrik.

Introduction

Laparoscopic surgery has become a standard operation method in many gynecologic procedures. The challenges when learning to master laparoscopy relate to both psychomotor skills and the perceptual nature. The laparoscopic technique is ideally taught in a simulated environment, where simulation training has shown to deliver a fast and patient-safe path to technical competence in the operating room [1-5]. Despite convincing research achievements proving the worth of laparoscopic simulators, development of structured training curricula is challenging. This is mainly due to lack of knowledge regarding the best methods for training and the best design of a curriculum [6-8].

The overall aim of this thesis was to develop and validate a curriculum in basic laparoscopy for residents in obstetrics and gynecology.

Background

The background chapter is divided into three sections. The first section focuses on the laparoscopic technique and current opportunities to train this technique in a simulated environment. The next section explains some educational theories applicable in simulation training. The last section elucidates some of the most important components in an educational curriculum.
In the following thesis the terms resident, learner and trainee are used interchangeably, based on the assumption that the trainee and resident are always the learner.

The laparoscopic technique

Minimally invasive surgery has become the new gold standard in many operations within general surgery and gynecology. As the term minimally invasive surgery indicates, it is characterized by being less invasive compared with open surgery. The method has several advantages including shorter hospitalization, faster postoperative recovery, less postoperative pain and improved cosmetics.
Laparoscopy is a minimally invasive surgery technique widely used in many gynecologic procedures, such as salpingectomy, hysterectomy, cyst removal, and endometriosis, and it is increasingly used in surgical oncology.

Laparoscopic surgery is challenging to learn. In contrast to open surgery, where it is possible to touch the organs with the hands, laparoscopic surgery allows handling of organs only with long instruments managed through small holes in the abdomen, which reduces the maneuverability. Furthermore, the tip of the instrument is moved in the opposite direction to the surgeon's hand; a counterintuitive hand movement, and the surgeon has to navigate the instruments in a three-dimensional room using only two-dimensional visual feedback on a screen placed at a distance.

The novice surgeon requires practice to master the challenges of the laparoscopic technique. The learning curve is typically steep, and in many specialties, such as general surgery [9-11], urology [12-14], pediatrics [15] and gynecology [16], it has been shown that during the initial part of learning there is an increase in the number of complications and the training time.
It is now possible to do the initial training in a simulated setting using laparoscopic simulators and without any risk to patients. Laparoscopic surgical simulators provide trainees with the opportunity to enhance motor skills in a standardized, controlled environment, where tasks can be repeated until proficiency is reached. There are several kinds of laparoscopic simulator, both high-fidelity virtual reality simulators and low-fidelity bench models. Bench models are designed to deliver only didactic training, whereas virtual reality simulators are designed to measure various facets of performance such as instrument handling and effectiveness, time to complete tasks, blood loss, errors and use of electrocautery.

In modern surgical residency initial skills training has transitioned from the operating room to the surgical skills laboratory. This is due to extensive research, both randomized trials and systematic reviews, on the effectiveness of simulators [1-5,17-20]

Since the first randomized trials involving virtual reality training were conducted ten years ago by pioneers such as Grantcharov [4] and Seymour [2], simulators have undergone rapid development. The currently available virtual reality simulators combine procedural tasks with didactic learning programs on diagnosis, indications and contraindications for an operation as well as anatomy in the surgical field. Additionally, these features are portrayed with very realistic graphics, Figure 1 and 2. There are several commercially available simulators on the market. In this thesis the LapSim®Gyn from Surgical Science, Goteborg, Sweden was used.

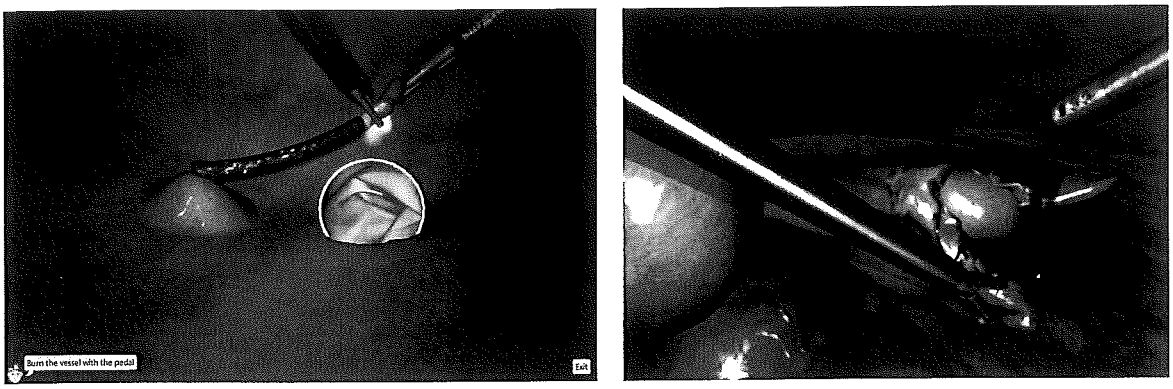


Figure 1 and 2. Exmples of the graphical representation of a basic skills task (left) and the operational task 'salpingectomy' (right) from the virtual reality simulator LapSim®.

The educational perspective

Education within the medical world—and in general—is life long. It cannot be staged to a certain point in time; however, it is necessary to ensure that certain surgical standards are reached before operating in order to minimize the risk of unnecessary harm to patients.
Historically, the operating room has served as the educational setting for novice surgeons, and skills were learned via the Halstedian model, better known as 'see one, do one, teach one' [21]. This is no longer acceptable due to ethical considerations toward patients, the demand for operating room efficiency, and the advances in educational theory and technology [22].

Unlike the clinical environment, simulation is designed for the benefit of the learner. Skills laboratories support a learner-centered approach to education compared with the clinical setting, the operation room, which is patient-centered. Barbara Schout, who recently published a PhD concerning virtual reality training in urology, described the characteristics and differences of a skills laboratory and the clinical settings, Table 1 [23].

Table 1. Characteristics of skills laboratory settings and clinical settings. Simulator training in a skills laboratory accommodates several of the requirements demanded in the learning theory literature.

Skills laboratory setting	Clinical setting
Learner centered	Patient centered
Repeated practice (deliberate practice)	One case at a time
Mistakes are allowed	Mistakes are unacceptable
Safe environment for learner	Less safe environment for learner
24-hour availability of learning facilities	Learning opportunities dependent on patient care
Variation of cases and exercises	Cases dependent on patient care
Possibility for peer teaching (learning from colleagues in the same learning phase)	No possibility for peer teaching
Standardized assessment	Assessment dependent on patient case

From B. Schout's PhD thesis 'Training in Urology, From Virtual to Reality' [23]

As Table 1 shows, there are several advantages of performing initial skills training in a skills laboratory, especially the point concerning mistakes, which are acceptable in the simulated setting but unacceptable in clinical practice.

Several psychomotor learning approaches and educational frameworks have been described and are clarified separately below in the light of simulation training. To understand what simulation can achieve and what its limitations are, educational theory can be helpful [24]. In this thesis three frameworks are reflected on as they contribute to understanding the assessment of skills performance and learning of psychomotor skills.

Miller's pyramid of competence and assessment

In 1990 Miller described a framework for clinical competence assessment with increasingly complex levels of skills performance [25]. The different layers in Miller's model are built up like a pyramid with four levels, Figure 3, forming a horizontally layered hierarchical categorization of clinical competence. All layers are needed and have their own impact on clinical competence. The bottom level of the pyramid is the *knows* level as Miller realized that learning of skills, any skill, begins with a knowledge component which is a prerequisite for carrying out professional functions. The next two layers *knows how* and *shows how* follow upon *knows* and are often used interchangeably; however, *knows how* as the procedural knowledge means that a physician can apply knowledge in concrete situations, while *shows how* is the competence and thereby the ability to use this knowledge to perform concrete actions. The top layer *does* focuses on what occurs in practice rather than what happens in the artificial setting and is the actual doing in day-to-day practice. Reaching the *does* level regarding laparoscopic surgery requires both knowledge and motor skills.
Miller's framework is applicable to skills training since all levels can be connected in a curriculum outside the operation room. However, the highest level of the pyramid collects information about doctors' performance in their daily practice and is a continuous process.
The importance of Miller's framework lies in its ability to identify learning objectives (what) and link them with appropriate testing contexts (where) and instruments (how). This is illustrated in Figure 3.

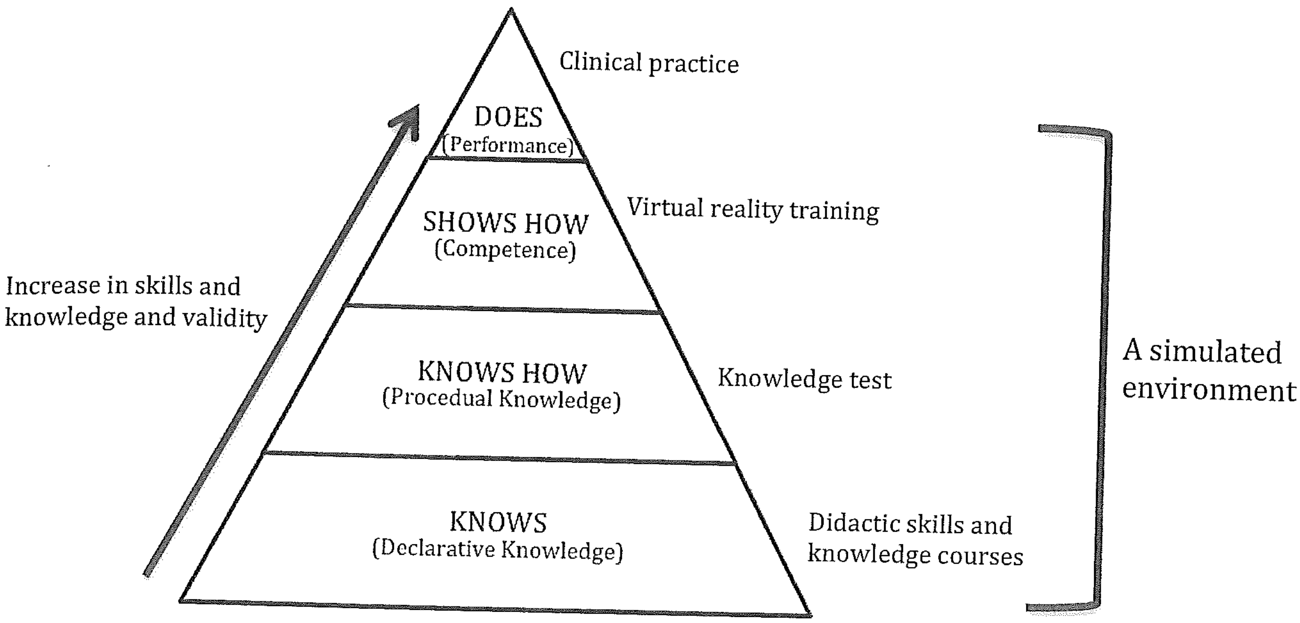


Figure 3. Miller's framework of competence and assessment in education with application of how and where the different levels can be assessed. The arrow on the left illustrates the increase in skills and knowledge, and the validity of results.

Kirkpatrick's evaluation model

Perhaps the best-known evaluation methodology for judging learning processes is Donald Kirkpatrick's Four Level Evaluation Model first published in a series of articles in 1959. The series was later compiled and published as the article, 'Techniques for Evaluating Training Programs' from 1975 [26], and it has been republished several times since [27]. Even though the model has recently been criticized for being outdated and simplistic, it remains one of the corner stones of the evaluation theory.

Kirkpatrick's four levels of evaluation include: *reaction*, *learning*, *behavior* and *results*, and all four levels can be reflected in a laparoscopic curriculum. The importance of Kirkpatrick's model lies in its ability to identify a range of dimensions that need to be evaluated in order to inform about the educational quality of a specific program or curriculum [27]. It is an outcome-focused model; however, it provides limited information about the individual learner.

The *reaction level* measures how the participants react to the training or learning activity. The next level, the *learning level*, measures what the participants have actually learned. This is followed by the *behavior level*, which measures whether what is learned is being applied on the job. Last is the *result level* measuring whether the application of training is achieving results.

Each level of evaluation builds upon the evaluation of the previous level, and the multiple measures of evaluation add precision to the measure of training effectiveness; however, the analysis becomes increasingly extensive and thereby time consuming and costly.

Criticism of the model has been directed at the oversimplified view of the training effectiveness, which does not consider individual or contextual influences in the evaluation of training [28]. Bates argues that it is likewise important to consider fundamental, ethical questions about training evaluations, for example "Are we doing the right thing, and are we doing it well?" [28]. We acknowledge the criticism and take note of it throughout the thesis.

Fitts and Posner's three-stage theory of motor skills acquisition

Unlike Miller's and Kirkpatrick's frameworks, which focus on evaluation of a program or curriculum, Fitts and Posner have suggested a skills learning theory that focuses on the individual learner's movement through specific stages when acquiring a new skill [29]. The theory was originally based on motor skills acquisition but can also be applied to surgical skills [30-32].

The learning process is sequential and involves three stages starting with *the cognitive phase*, which is the identification and development of the component parts of the skill involving the formation of a mental picture of the skill. This phase is followed by the *associative phase* linking the component parts into a smooth action involving practicing the skill and using feedback to perfect the skill. The third phase is the *autonomous phase* where the developed and learned skill becomes automatic, involving little or no conscious thought or attention when performing the skill. It is not certain that all performers reach this stage.

The concept of progressing through well-defined stages correlates well with training in a simulated environment since all stages can be achieved, just as in the operation room, but without jeopardizing patient safety. The theory is useful regarding the learning of psychomotor skills, such as laparoscopic operations, and identifies that the relevant movements must be assembled, component-by-component. This can be accomplished on a virtual reality simulator where the unskilled performance with many errors shifts to a skilled performance with few, minor errors.

Developing a curriculum in basic laparoscopy

Important work has been done to validate simulators as viable methods for teaching technical skills outside the operating room, and the next step is to integrate simulation training into a curriculum in order to standardize and plan education. Although virtual reality simulators and bench models have been studied thoroughly and have proven their worth, integration in the actual training curriculum remains a problem [6,7,33]. This is mainly due to lack of knowledge on best practice within training methods and the ideal design of a curriculum [7,8,34].

Stefanidis and Heniford write: *'A successful laparoscopic skills curriculum depends on many factors including participants' motivation, available resources and personnel, and trainee and faculty commitment. It should encompass goal-oriented training, sensitive and objective performance metrics, appropriate methods of instruction and feedback, deliberate, disturbed and variable practice, an amount of overtraining, maintenance training, and a cognitive component. A curriculum that follows these principles is likely to spark trainee interest...'* [7].

Development of a curriculum requires a comprehensive approach. It has to meet demands from stakeholders at many levels, such as the organizational level, the faculty level and the student level. Furthermore, as stated by Stefanidis and Heniford, a curriculum encompasses many aspects, such as learning opportunities and outcomes, educational strategies, assessment, educational environment and content [35]. Thus, the overall aim of any curriculum is to ensure the learning of an agreed standard of skills or knowledge [36].

Psychomotor learning theory emphasizes the integration of a cognitive stage—where a novice develops a mental picture of the motor task—as the first step in learning a procedure. To do this correctly, the trainee must understand the steps of the operation in the correct order and must know how to troubleshoot unexpected developments. Following the cognitive component is the rehearsal of a task in a simulated environment, which requires both technical and cognitive knowledge. These components, knowledge and skills, are essential in the development of technical skills curricula [7,37-39].
The importance of cognitive training has been highlighted in a study by Tang et al., where they demonstrate that cognitive errors, such as a lack of understanding of the correct sequence of steps in an operation, trigger the majority of mistakes in a procedural task rather than technical errors [40].

Within laparoscopic curricula the Fundamental of Laparoscopic Surgery (FLS) is predominant, especially in North America. The FLS program addresses both technical skills using bench models along with laparoscopic knowledge using a multiple-choice test [41]. The performance metrics including the cognitive multiple-choice test have been extensively validated for surgical residents [42-45], but currently the cognitive test is not optimal for gynecologic residents [46,47]. Additionally, it is limited to teaching basic component laparoscopic skills and not full procedures [48].
Other curricula on virtual reality simulators have been described for basic laparoscopy, but many do not contain a cognitive component [49-51].

The most convincing way to scientifically justify a curriculum's existence would be to set up a randomized trial. This was recently done in colorectal surgery by Palter and Grantcharov, showing that the intervention group improved technical proficiency in the operation room and improved cognitive knowledge related to performing an advanced procedure compared with the control group, who underwent traditional training [52].
In the current thesis, we decided it was ethically unacceptable to conduct a randomized study where residents used patients as training models (traditional training), due to the mounting evidence in favor of simulation training. This is also argued by Seymour in a review from 2008 [53].

Ample evidence demonstrates that simulation is an efficient method of teaching laparoscopic skills [54], and virtual reality training has shown to improve operative performance [1-4]. However, mere surgical skills are not sufficient to ensure proper laparoscopic knowledge, there is also a need for a cognitive component. As Roger Kneebone writes: '*Without a curriculum, there is a danger that technology will dominate learning and that educational goals will be eclipsed. Clear vision is essential to make wise use of simulation's enormous potential, and to ensure an equal partnership between pedagogy and technology*' [55]. Furthermore, as this thesis shows, there is much to gain from discussing both components in relation to well-evidenced psychomotor learning theories.

The aim of this thesis was to develop and validate a step-wise curriculum in basic laparoscopy for first-year residents in obstetrics and gynecology with clear educational objectives that mirrored actual practice. The curriculum is inspired by the FLS program and the curriculum literature [7,36,41,56-59] and adheres to the current standards of proficiency-based training and distributed and deliberate practice [3,60-62]. Additionally, the curriculum acknowledges the psychomotor learning theory [25,27,29,62-64]. The core competencies revolve around basic laparoscopy, integrating a knowledge component, a technical skills component and, as a novelty compared with the existing literature, an operational component.

Aims

The current thesis consists of four studies all aiming to develop and validate individual steps in a four-step curriculum in basic laparoscopy for residents in obstetrics and gynecology.

Study I
The aim of the study was to develop and validate a knowledge test in basic laparoscopy.

Study II
The aim of the study was to investigate the impact of instructor feedback when training a complex operational task on a laparoscopic virtual reality simulator.

Study III
The aim of the study was to investigate whether doctors on different educational levels can assess laparoscopic operations on the basis of an assessment scale.

Study IV
The aim of the study was to develop and validate a four-step curriculum in basic laparoscopy for first-year residents in obstetrics and gynecology.

Research questions

Study I
1a) Is it possible to develop and validate a multiple-choice test in basic laparoscopy on the basis of a regional consensus panel?
1b) Can the multiple-choice test distinguish between different educational levels?

Study II
What is the impact of instructor feedback when training a complex operational task on a virtual reality simulator?

Study III
Can doctors at different educational levels assess laparoscopic operations on the basis of an assessment scale?

Study IV
1a) Is it possible to develop a four-step curriculum in basic laparoscopy based on validated constituent elements?
1b) Is a four-step curriculum in basic laparoscopy usable for first-year residents in obstetrics and gynecology?

Study I. The test study

'Development and validation of a theoretical test in basic laparoscopy'.
Strandbygaard J, Maagaard M, Larsen CR, Schouenborg L, Ottosen C, Ringsted C, Grantcharov T, Ottesen B, Sorensen JL.
Accepted in Surgical Endoscopy, September, 2012
DOI: 10.1007/s00464-012-2615-7

Surg
DOI

De
lap

Jear
Lar
Ben

Recei
© Sp

Abst
Back
pone
struc
the s
need
basic
gyne
valid
multi
Meth
deteri
views
vance
metho

J. Strai
C. Otto
Depart
Rigsho
Blegda
e-mail:

C. R. L
Departi
Dyreha

C. Ring
Centre
Copenh
2100 C

T. Gran
Departn
Toronto

B. Ottes
Juliane
Reprodu

Surg Endosc
DOI 10.1007/s00464-012-2615-7

Development and validation of a theoretical test in basic laparoscopy

Jeanett Strandbygaard · Mathilde Maagaard · Christian Rifbjerg Larsen · Lars Schouenborg · Christian Ottosen · Charlotte Ringsted · Teodor Grantcharov · Bent Ottesen · Jette Led Sorensen

Received: 24 May 2012 / Accepted: 14 September 2012

Abstract

Background Testing of knowledge is an important component in a successful skills curriculum. Nonetheless, structured testing of basic procedure-relevant knowledge in the surgical domains is not ordinary practice. A regional need assessment showed insufficient knowledge regarding basic laparoscopy for first-year residents in obstetrics and gynecology. This study therefore aimed to develop and validate a framework for a theoretical knowledge test, a multiple-choice test, in basic theory related to laparoscopy.

Methods The content of the multiple-choice test was determined by conducting informal conversational interviews with experts in laparoscopy. The subsequent relevance of the test questions was evaluated using the Delphi method involving regional chief physicians. Construct validity was tested by comparing test results from three groups with expected different clinical competence and knowledge levels: senior medical students, first-year residents, and chief physicians.

Results The four conversational interviews resulted in the development of 47 test questions, which were narrowed down to 37 test questions after two Delphi rounds involving 12 chief physicians. Significant differences were found between the test scores from the senior medical students ($n = 14$) and the first-year residents ($n = 52$) (median test scores, 18 vs. 24, respectively; $p = 0.001$), and between the first-year residents and the chief physicians ($n = 12$) (median test scores, 24 vs. 33, respectively; $p = 0.001$). Internal consistency (Cronbach's alpha) was 0.82. There was no evidence of differential item functioning between the three groups tested.

Conclusions A newly developed knowledge test in basic laparoscopy proved to have content and construct validity. The formula for the development and validation of a theoretical test could potentially be used for any topics that require structured testing of knowledge.

Keywords Assessment · Curriculum · Laparoscopy · Multiple-choice test · Theoretical test

A cognitive knowledge component is an important part of a laparoscopic curriculum and is expected to provide the physician with basic understanding to handle unexpected events [1, 2]. A supporting argument for this was demonstrated in a randomized study that found cognitive skills training enhancing the ability to execute a surgical task correctly [3]. Nonetheless, structured testing of basic procedure-relevant knowledge in the surgical domains is not ordinary practice.

J. Strandbygaard (✉) · M. Maagaard · L. Schouenborg · C. Ottosen · J. L. Sorensen
Department of Obstetrics and Gynecolology, 4074, Rigshospitalet, University Hospital of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark
e-mail: jeanett.oestergaard@rh.regionh.dk

C. R. Larsen
Department of Obstetrics and Gynaecology, Hillerød Hospital, Dyrehavevej 29, 3400 Hillerød, Denmark

C. Ringsted
Centre of Clinical Education, University of Copenhagen, Copenhagen and Capital Region, Blegdamsvej 9, 2100 Copenhagen, Denmark

T. Grantcharov
Department of Surgery, St. Michael's Hospital, 30 Bond Street, Toronto, ON M5B 1W8, Canada

B. Ottesen
Juliane Marie Centre, Centre for Children, Women, and Reproduction, Blegdamsvej 9, 2100 Copenhagen, Denmark

In Northern America, the validated program, fundamentals of laparoscopic surgery (FLS) [4], developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES), currently is widely accepted as a requirement during surgical residency. The FLS is based on examination of both theoretical knowledge in basic laparoscopy using a multiple-choice test and technical skills performance. Several aspects of the FLS program, both theoretically and technically, can be adapted to other endoscopic specialties, but it is a costly program, and the cognitive component currently is not optimal for residents in gynecology [2]. On this basis, the current study aimed to develop and validate a multiple-choice test on basic theory related to laparoscopy for first-year residents in gynecology.

A primary purpose of testing is to communicate what is important, and any test should be preceded by analysis of knowledge gaps (e.g., a need analysis of the relevant subject). Tests and assessment are powerful motivators for learning and can encourage trainees to study on their own and participate in available educational opportunities [5]. Resident curricula may vary from hospital to hospital, and it therefore is important to develop tests that promote knowledge and aid in the accomplishment of key curricular goals [5]. This can facilitate homogeneous learning, both regionally and nationally.

Several factors must be considered in the development of test questions; purpose, type, objectives, content, construction, format, validity, and reliability [5, 6]. Furthermore, educational impact, cost effectiveness, and acceptability need to be taken into account [7]. Although test development often is considered a simple task, designing a valid and reliable test is comprehensive.

Research has shown that the format of the question is of limited importance. It is the content of the question that determines what the question tests [7, 8]. Multiple-choice questions (MCQs) are useful for testing cognitive knowledge and efficient for examination of large groups and large subject areas. The scoring of MCQs is time saving, and effective MCQs can be reused (i.e., MCQs are cost effective). The main advantages of the MCQ format are its high reliability and its ability to cover a broad domain of knowledge. If constructed well, MCQs can test more than simple recall of facts [7].

Materials and methods

Context of the study

Danish specialty training in obstetrics and gynecology requires 5 years followed by 12 months of obligatory internship. There are no postgraduate examinations to pass along the way, just in-training and work-based assessment. The current test was designed for use in a regional lapa roscopic curriculum for first-year residents in gynecolog to standardize the in-training and work-based assessments

Structure of the MCQ development

The development and validation of the MCQs include four steps (Fig. 1) based on recommendations from severa articles [6, 7, 9–11]. Haladyna et al. have formulated guideline with 31 basic principles of effective MCQ writ ing [12], which was endorsed in this study together with th item-writing guide by Case and Swanson [5]. The firs author was the primary creator of MCQs, supported by th last author.

The question consisted of a stem (e.g., a clinical cas presentation) and a lead in question, followed by a series o choices: one correct answer and three distractors. Th distractors were not completely wrong but less correct tha the correct answer. The examinee was instructed to selec the "most likely diagnoses." The emphasis was put o writing questions that required examinees to problem solv using clinical judgment and operational decision makin rather than simple recollection.

Step 1: need analysis of the relevant subject

A prior need analysis conducted among eight educationa chief physicians representing all departments in th Copenhagen and Zealand Region identified a need fo structured testing and training on the topic "basic lapa roscopy" for first-year residents in obstetrics an gynecology.

Step 2a: content validity 1

One investigator (the first author) performed an inform 2-h conversational interview with four experts in lapar scopic gynecology. There was no predetermination o questions except from the topic: identification of releva basic laparoscopy knowledge for first-year residents i gynecology. The interviews were recorded using notes. Th four experts, representing three different university hosp tals, were chosen due to their extensive knowledge an technical skills in laparoscopic gynecology.

Step 2b: content validity 2

Based on the conversational interviews, MCQs wi matching answers and distractors (incorrect options) we developed and subsequently distributed to chief physicia in gynecology representing eight departments in the Ze land Region of Denmark. The chief physicians were chos on the basis of their operational expertise.

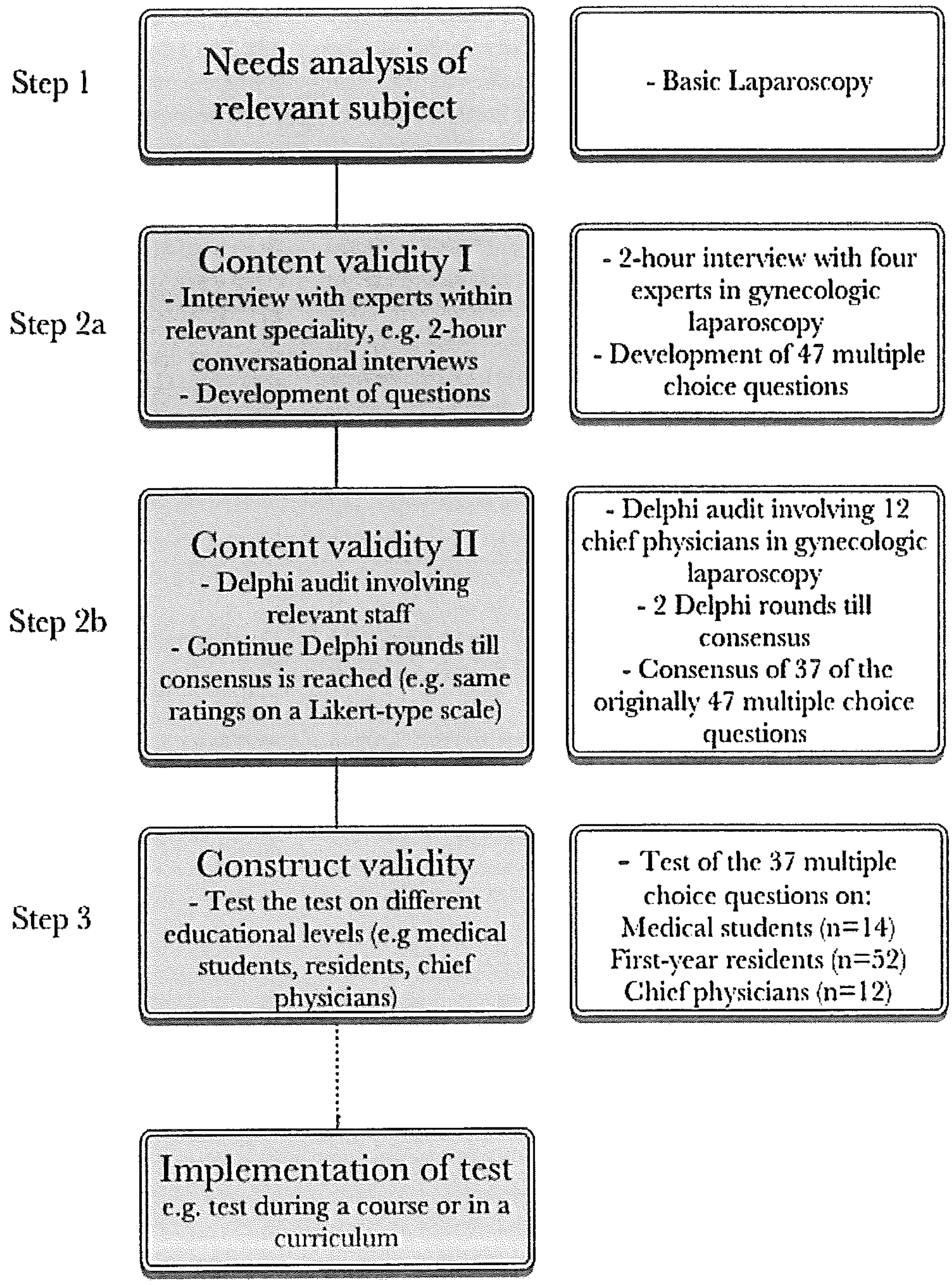


Fig. 1 Steps of test development and validation. The *right side* presents data from the current study

A modified Delphi method was applied to achieve consensus on the relevance of the MCQs. The Delphi technique is an anonymous process whereby responses are collected and analyzed until consensus is achieved [13]. Each question was rated on a scale of 1 (extreme relevance) to 5 (no relevance at all) concerning its relevance for residents. Additionally, the chief physicians were asked to answer the MCQs the first time they received the test. Their answers were used in step 3.

If a question was rated above 3 by more than 20 % of the chief physicians, the question was excluded in the final version. If a new question was recommended by more than 20 %, the question was included for the next round. The questions were sent out until consensus was achieved (i.e., all the chief physicians rated all MCQs below 3).

Step 3: construct validity

The construct validity of the MCQs was tested by comparison of three groups with expected different clinical competence and knowledge levels. The high-competence group consisted of the chief physicians who participated in the Delphi evaluation (step 2b). The intermediary competence group consisted of first-year residents in obstetrics and gynecology who had been in the specialty for 4–6 months. They were tested in the morning on the day of an obligatory course held at a university hospital. The low-competence group consisted of medical students in their final year of medical school who recently had finished a 4-week clinical stay at an obstetrics and gynecology department. They were tested on their final day of the stay, with no remedies (e.g., Internet or books) allowed during the test. There was no time constraint for answering.

Ethics

Participation was voluntary, and no patients were involved. Therefore, ethical approval was not required according to Danish regulations.

Statistical analysis

Data were processed using SPSS 19.0 for Windows (SPSS, Chicago, IL, USA). Cronbach's alpha was calculated for internal consistency.

Differential item functioning was tested with logistic regression [14, 15]. Differential item functioning occurs when examinees from different groups show differing probabilities of scoring on questions after matching on the underlying ability that the question is intended to measure [15]. It is an important statistical aspect in the development of tests because if differential item functioning is apparent for a question, then bias is present, and it needs to b considered whether the questions should be removed fro the test.

Nonparametric tests were used because distributions o the obtained answers did not justify the general use o parametric tests. The Mann–Whitney *U* test was used t compare test scores of medical students with those of firs year residents and chief physicians. All *p* values lower tha 0.05 were considered significant. Scores are presented a medians and minimum and maximum scores.

Results

Step 1: A regional need analysis of the relevant subject

This analysis was performed as described in the Materia and methods section.

Content validity

Figure 1 presents obtained results. Based on the conve sational interview with experts, 47 MCQs and matchin answers were developed. The MCQs were distributed to 1 chief physicians with prior experience in the range o 250–1,000 independent laparoscopic procedures.

After the first Delphi round, 10 MCQs were exclude None was added. The modified 37 MCQs were sent out fo a second evaluation. After the second round, no MCQ

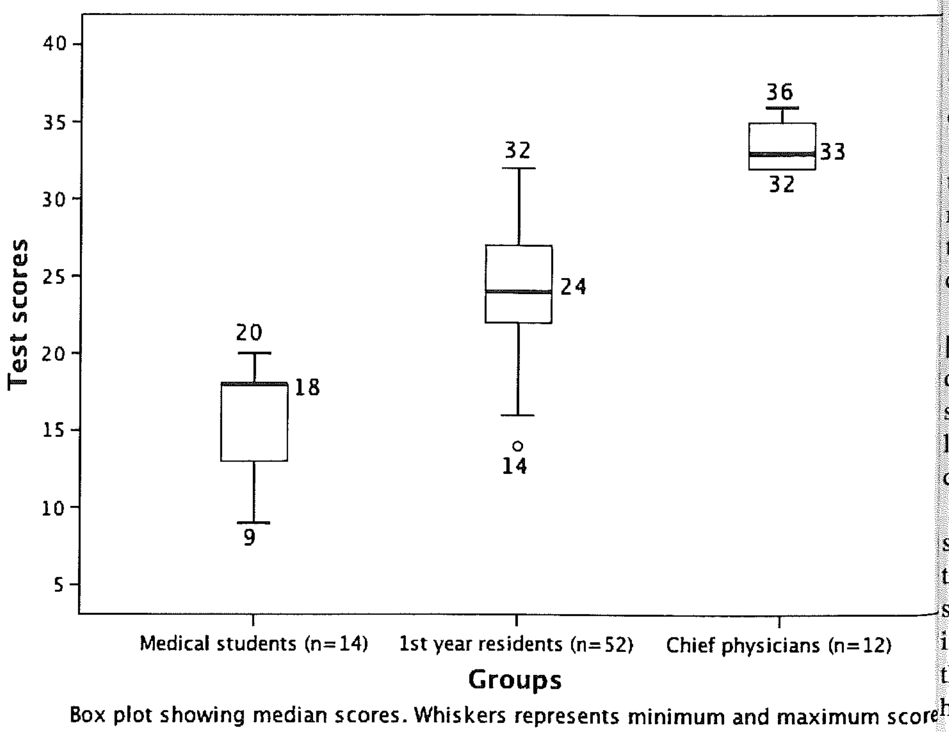


Fig. 2 MCQ test scores from medical students, first-year residents, and chief physicians; significant differences were found between MCQ test scores in the three groups

were rated not relevant (i.e., consensus was obtained). The response rate was 100 % for both rounds. In the end, the total number of questions was 37. Four of these questions consisted of illustrative drawings.

Construct validity

Significant differences were found between the MCQ test scores from the medical students ($n = 14$) and the first-year residents ($n = 52$) ($p = 0.001$), and between the first-year residents and the chief physicians ($n = 12$) ($p = 0.001$) (Fig. 2). The internal consistency (Cronbach's alpha) for the test, including the scores from the medical students, first-year residents, and chief physicians, was 0.82. There was no evidence of differential item functioning between the three groups tested.

Discussion

This report describes the development and validation of a theoretic knowledge test in basic laparoscopy.

Validation of a theoretical test

The content of the theoretical test was based on informal conversational interviews of experts. The interview format has both strengths and weaknesses [16]. Its strengths are that the interview can be matched to individuals and circumstances, and the non-predetermination of topics allows the person being interviewed to associate freely with observations from his or her professional life. The potential weaknesses of this format are the lack of taxonomy and the comprehensive analysis.

Although analysis can be comprehensive, it is important that the sample of questions be broad. Otherwise, the test results can be biased and thereby not representative [5]. If the sample is too small, the test results may not be sufficiently precise to ensure that they reflect true proficiency.

Content validation is based on the judgment of experts [11]. Thus, the definition and selection of experts to validate test items is controversial. To overcome potential selection bias, a validation process needs a sufficiently large and diverse sample of doctors within a given specialty [9, 17].

In the current study, experts in laparoscopy and a subsequent consensus panel with chief physicians evaluated the content of the MCQ test. It is therefore reasonable to state that the content sampled was within the domain of interest and thereby valid. We cannot, however, rule out that there was a response bias (i.e., different experts would have produced different results).

One of the predominant factors for diverging opinions among surgeons is the different perception of "correct surgery," and it can be difficult to obtain proper evidence on those matters. The 10 MCQs omitted in the first Delphi round were mainly concerning degree of difficulty and not diverging operational methods.

For tests on which high-stakes decisions are based, reliability above 0.8 is desirable, indicating that a retest (with similar but not identical content) should be correlated about 0.8 with the original test score [18]. For most tests, the reliability score increases with test length. A well-written MCQ test with 100 questions often generates a reliability of 0.8.

The MCQ test in this study generated good reliability and demonstrated construct validity based on only 37 questions. Most probably this was due to the rather focused content area: knowledge related to laparoscopy. In addition, the thorough test development process contributed to the good validity and reliability. Supporting this was the finding of no differential item functioning among the test questions (i.e., it was not necessary to exclude any of the questions).

MCQ as a test format

The MCQ test format was chosen due to the following parallels with surgical decision making: recognition, distinguishing of the correct solutions, and characteristics of selection, estimation, prediction, and categorization [5, 6]. Writing effective, creative, and challenging MCQs is difficult and time consuming, and an important beginning is to select and sample the correct topics. The test writer cannot ask everything. Furthermore, it is of great importance to have content expertise in reviewing the questions [6]. However, expertise is not sufficient to ensure valid questions. Training in question writing is necessary. The MCQs showed good results regarding reliability and validity, and the main reason for this is presumably due to the broad sampling of questions.

The designers of the FLS program prepared their test in cooperation with the Joint Committee on Standards of the American Educational Research Association, The American Psychological Association, and The National Council on Measurement in Education [4]. Involvement of these authorities is undoubtedly an advantage, especially in national standardized programs, but as shown in this article, a panel of experts within a given topic combined with instructions from MCQ guidelines [5, 12] can provide a solid basis for a valid multiple-choice test.

As mentioned, MCQs are useful and cost effective for examination of large groups and large subject areas due to ease of scoring and reuse. In contrast short- and long-answer essay responses are time consuming and difficult to score accurately and reliably [6]. Although the clinical reasoning process might be easier to review in an essay, a format that allows a broad reasoning across a larger number of problems, such as MCQs, is more efficient for testing knowledge domains [6, 18]. It is beyond the scope of this article to thoroughly discuss differences between MCQs and other test formats.

Schuwirth and van der Vleuten [7] created five criteria that can be used to evaluate the advantages and disadvantages of written assessment methods: reliability, validity, educational impact, cost effectiveness, and acceptability. The current study met the majority of these demands and provided useful information on how to develop and validate a test with relatively simple tools that can be applied within different areas of knowledge.

A randomized study showed that cognitive skills training enhanced the ability to execute a surgical task correctly [3] and that a cognitive component should accompany skills training in a successful curriculum [1, 4, 19]. However, estimating concurrent validity is not straightforward. In surgery, performance on examinations testing knowledge has correlated poorly with assessments of technical skills and operative performance [20, 21]. This indicates that assessment strategies targeting performance in practice are necessary supplements to the cognitive tests in order to test medical competence comprehensively.

Several surgical organizations and colleges have developed knowledge-based tests on an individual scale, but currently, not much literature is published on the matter. The FLS knowledge-based test has been investigated thoroughly [2, 4, 22], but whether the test results extrapolate well in real-world clinical situations is yet be investigated.

Limitations and future perspectives

The current report provides a practical formula for the development and validation of a cognitive test component. Nevertheless, whether ensuring sufficient knowledge before clinical skills training has an effect on the clinical performance curve needs further research. Thus, we believe that a validated knowledge test is the first step toward examining transferability of training to the operation room.

Research has shown that students tested on material remember that material better than if they were not tested on the material. This is called "the testing effect" [23]. Whether the MCQ test can be used as a progress test, taking advantage of the testing effect remains to be studied.

Knowledge tests in laparoscopy must go hand in hand with skills training and evaluation of operative performance [24]. The MCQ test developed in this study is the cognitive element in a comprehensive skills curriculum for first-year residents in gynecology that also consists of technical skills element and an operational element.

Acknowledgments We thank Karl Bang Christensen, Associa Professor, Department of Public Health, Unit of Biostatistics, f statistical help. We also thank Dorte Nielsen, Chief Physicia Department of Obstetrics and Gynecology, Rigshospitalet, Universi Hospital of Copenhagen; Jens Jørgen Kjer, Chief Physician, Depar ment of Obstetrics and Gynecology, Rigshospitalet, University Ho pital of Copenhagen; and Helle V. Clausen, Chief Physicia Department of Obstetrics and Gynecology, Herlev Hospital, Unive sity Hospital of Copenhagen for their participation in the intervie part of the study. Furthermore, we thank all the chief physicians fro the consensus panel.

Disclosures Jeanett Strandbygaard, Mathilde Maagaard, Christi Rifbjerg Larsen, Lars Schouenborg, Christian Ottosen, Charlo Ringsted, Teodor Grantcharov, Bent Ottesen, and Jette Led Sorens have no conflicts of interest or financial ties to declare.

References

1. Stefanidis D, Heniford BT (2009) The formula for a success laparoscopic skills curriculum. Arch Surg 144:77–82
2. Zheng B, Hur H-C, Johnson S, Swanstrom LL (2010) Validity using Fundamentals of Laparoscopic Surgery (FLS) program assess laparoscopic competence for gynecologists. Surg Endo 24:152–160
3. Kohls-Gatzoulis JA, Regehr G, Hutchison C (2004) Teachi cognitive skills improves learning in surgical skills courses: blinded, prospective, randomized study. Can J Surg 47:277–2
4. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin L Schirmer B, Hofmann K, SAGESFLS Committee (200 Development and validation of a comprehensive program education and assessment of the basic fundamentals of lapar scopic surgery. Surgery 135:21–27
5. Case SM, Swanson DB (1998) Constructing written test questio for the basic and clinical sciences, 2nd edn. National Board Medical Examiners, Philadelphia
6. Downing SM, Yudkowsky R (2009) Written tests. Assessment health professions education, 1st edn. Routledge, New Yo pp 149–184
7. Schuwirth LWT, van der Vleuten CPM (2003) ABC of learni and teaching in medicine: written assessment. BMJ 22:643–6
8. van der Vleuten CPM, Schhwirth LWT (2009) Written asse ments: a practical guide for medical teachers, 3rd edn. Church Livingstone, London, pp 323–331
9. Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm Kiessling C (2008) A situational judgement test of professio behaviour: development and validation. Med Teach 30:528–5
10. Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BL Scherpbier AJJA (2010) Validation and implementation of su gical simulators: a critical review of present, past, and futu Surg Endosc 24:536–546
11. Norman GR, Streiner DL (2008) Validity. Health measureme scales: a practical guide to their development and use., 4th ed Oxford University Press, Oxford, pp 247–276
12. Haladyna T, Downing S, Rodriguez M (2002) A review multiple-choice item-writing guidelines for classroom asse ment. Appl Meas Educ 15:309–334
13. Graham B, Regehr G (2003) Delphi as a method to establi consensus for diagnostic criteria. J Clin Epidemiol 56:1150–11

14. Swaminathan H (1990) Detecting differential item functioning using logistic regression procedures. Educ Meas 27:361–370
15. Zumbo B (1999) A handbook on the theory and methods of differential item functioning (DIF). National Defense Headquarters, Ottawa
16. Cohen L, Manion L, Morrison K (2000) Interviews. Research methods in education, 5th edn. Routledge, New York, pp 265–292
17. Wass V, Van der Vleuten C, Shatzer J, Jones R (2001) Assessment of clinical competence. Lancet 24:945–949
18. Hawkins RE, Swanson DB (2008) Using written examinations to assess medical knowledge and its application: practical guide to the evaluation of clinical competence, 1st edn. Mosby, Philadelphia, pp 42–59
19. Fried GM (2006) Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. Gastrointest Endosc Clin N Am 16:425–434
20. Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Jones DB (2000) Evaluating surgical competency with the American board of surgery in-training examination, skill testing, and intraoperative assessment. Surgery 128:613–622
21. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE (1995) The objective structured clinical examination: the new gold standard for evaluating postgraduate clinical performance. Ann Surg 222:735–742
22. Swanstrom LL, Fried GM, Hoffman KI, Soper NJ (2006) Beta test results of a new system assessing competence in laparoscopic surgery. J Am Coll Surg 202:62–69
23. Larsen DP, Butler AC, Roediger HL (2008) Test-enhanced learning in medical education. Med Educ 42:959–966
24. Wentink M, Stassen LPS, Alwayn I, Hosman RJAW, Stassen HG (2003) Rasmussen's model of human behavior in laparoscopy training. Surg Endosc 17:1241–1246

Study II. The feedback study

'Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial'.

~~Oestergaard~~ J, Bjerrum F, Maagaard M, Winkel P, Larsen CR, Ringsted C, Gluud C, Grantcharov T, Ottesen B, Sorensen JL.

Accepted in Annals of Surgery, September 2012

Abstract

Objective: To investigate the impact of instructor feedback versus no instructor feedback when training a complex operational task on a laparoscopic virtual reality simulator.
Background: Simulators are now widely accepted as a training tool, but there is insufficient knowledge about how much feedback is necessary, which is useful for sustainable implementation.
Method: A randomized trial following CONSORT Statement. All participants had to reach a predefined proficiency level for a complex operational task on a virtual reality simulator. The intervention group received standardized instructor feedback a maximum of three times. The control group did not receive instructor feedback.
Participants: Senior medical students without prior laparoscopic experience (n=99). Outcome measures were time, repetitions and performance score to reach a predefined proficiency level. Furthermore, influence of sex and perception of own surgical skills were examined.
Results: Time (min.) and repetitions (rep.) were reduced in the intervention group (162 min. vs. 342 min., $p<0.005$) and (29 rep. vs. 65 rep., $p<0.005$). The control group achieved a higher performance score (%) than the intervention group (57% vs. 49 %, $p=0.004$). Men used less time (min.), than women, $p=0.037$, but no sex difference was observed for repetitions, $p=0.20$.
Participants in the intervention group had higher self-perception regarding surgical skills after the trial, $p=0.011$.
Conclusions: Instructor feedback increases the efficiency when training a complex operational task on a virtual reality simulator; time and repetitions used to achieve a predefined proficiency level were significantly reduced in the group that received instructor feedback compared with the control group.
Trial registration number:NCT01497782

Introduction

For virtual reality simulation the benefits are clear; the drawbacks are less clear. Throughout the last decade several studies have found a positive effect of surgical virtual reality training on the learning curve and improvement of basic psychomotor skills in the operating room.(1-5) Despite the now well-established advantages of virtual reality simulators, the majority of surgical and gynecological departments encounter hurdles when implementing simulator training in surgical practice.(6) This is mainly due to concern about the time and human resources needed to train novice surgeons to an adequate level,(7) along with lack of knowledge on how to design a training program.(8) Questions have especially been asked regarding frequency, amount and type of feedback, i.e. simple versus escalating feedback, to obtain the best learning outcomes in complex operational virtual reality tasks.(9-12)

Feedback can be defined as the provision or return of performance-related information to the performer, and is an important part of learning in medical education.(13) In the laparoscopic virtual reality setting no studies or trials have investigated how instructor feedback affects learning to a predefined proficiency level in complex operational tasks, however, in more basis tasks, such as coordination and instrument navigation, no advantages of instructor feedback has been found.(14) Training of complex surgical tasks is a necessary prerequisite in an advanced surgical training program; therefore, focus needs to be drawn on whether feedback is a requirement when learning these operational skills.

Current literature suggests a predefined proficiency level based on experts' performance as an endpoint for novice training rather than a fixed training time.(15-18) Having attained a proficiency level on a virtual reality simulator has thoroughly been investigated and many studies have shown improvement of performance in the operation theatre.(2,3,5) These studies all included use of feedback, but their aim was to investigate transfer of skills from the virtual environment to the clinical setting, and not how instructor feedback impacts performance during simulator training. At present it is relevant to target investigation on how instructor feedback influences learning, additionally to investigate amount of feedback needed to optimize learning. Within motor skills learning, researchers have demonstrated that participants who self-direct their access to instruction or feedback during practice learn more than those whose access is controlled externally,(19,20) but it is uncertain whether these results apply in the surgical training environment.

With a worldwide proliferation of simulation centers, it is essential to explore the optimal circumstances for simulator training and investigate different learning approaches, e.g., a self-directed approach and the impact and type of feedback. We aimed to investigate the following in a randomized trial: the impact of instructor feedback versus no instructor feedback when training a complex operational task on a virtual reality simulator. Self-directed practice regarding when to receive instructor feedback was applied. Additionally, sex differences, computer gaming skills, and self-perception were examined during simulator training.

Methods and materials

The protocol for this trial has previous been published.(21)

Participants

Medical students in their 4th to 6th year (o of 6 years) recruited through adve tisements on websites at the surgical a anesthesiological student associations at t Copenhagen University Medical Scho Copenhagen, Denmark.(21)

Inclusion criteria: 1) Medical bachel degree (completion of the first three yea out of six at University of Copenhage Medical School). 2) Informed consent befo enrolment. 3) Attendance at an introducto meeting before the trial.

Exclusion criteria: 1) Independe experience with more than thr laparoscopic procedures. 2) Pri experience with virtual reality simulation. Not fluent in the Danish language. 4) Lack informed consent.

The virtual reality simulator task and equipment

The virtual reality simulator was LapSim®, version 2010, produced b Surgical Science, Sweden. The virtual reali task was a right-side laparoscop salpingectomy due to an ectopic pregnan At the end of each completed task, t virtual reality simulator summed up tim spent and quality of performance, table and presented automated feedback, i. performance score, available for a participants. The virtual reality simulat electronically recorded data from eve repetition. These electronic data w transferred to a secure database by a independent investigator.

The intervention group and the control grou

Both the intervention group and the contr group had to reach a predefined proficien level. The predefined proficiency level w defined and validated in a previous study b the same research group,(22) and referred to as the 'expert level' in this articl

The 'expert level' had to be reached twice within five consecutive repetitions.
All participants were informed about the operational technique of a salpingectomy during the obligatory introduction meeting.(21) Furthermore, all participants were instructed in use of the virtual reality simulator and shown the instruction video on how to perform the operational task and how to interpret the automated feedback generated by the simulator before the practice sessions began.
Practice sessions were maximum three hours per day; the virtual reality training was not full three hours, but it allowed for breaks during training. Participants were instructed to finish the practicing within a two-month period.

The intervention group had an obligatory feedback session placed on the first virtual reality operation and could additionally request two instructor feedback sessions. One instructor (the first author) provided the standardized feedback and used the same template for every participant. The instructor feedback was 10 to 12 minutes, and the template used feedback was standardized and consisted of the following: how to hold the laparoscopic instruments and thereby minimize instrument movements, how to optimally use electric cautery, and how to remove the fallopian tube. The feedback was not tailored to individual needs.

The control group did not receive any instructor feedback.

Both groups were asked to complete a questionnaire before and after the trial pertaining to perception of own surgical skills on a Likert-type scale from 1 to 5.

Outcome measures
Primary outcome measures: number of repetitions and total time (min.) used to reach 'expert level'.

Secondary outcome measure: the performance score (%) obtained when 'expert level' was reached. The performance score was automatically generated by the virtual reality simulator and based on performance recorded during the task, table 1.
In post hoc analyses the effect of sex and computer gaming skills were explored.

Randomization and participation
The randomization was performed by a computer randomization at the Copenhagen Trial Unit. The randomization procedure was concealed and executed by using the participants' unique personal identification number; the central personal registration number. We followed the CONSORT Statement for randomized trials.(21)

Sample size calculation
Based on data in a previous trial,(4) it was assumed that participants in the intervention group and control on average would use 30 and 40 repetitions respectively to reach 'expert level', i.e., a minimal relevant difference of 10 repetitions. The standard deviation was set to 15. With type I error set at 0.05 and power set at 0.90, the sample size added up to 96 participants, 48 in each group.

Statistical analysis
The data were analyzed using SPSS (Chicago, IL), version 15.0. Two sided significance tests were used with a level of significance = 0.05. The distributions of each outcome measure were compared between the intervention group and the control group using the general linear uni-variate model. The analyses were repeated with the co-variate semester number and the two protocol specified co-factors (sex, and computer game experience) included.
If the assumptions of the model (normally distributed residuals and variance homogenecity of the groups compared) could not be fulfilled using simple trans-formations of the data, the distributions of

the intervention group and the control group were compared using a non-parametric test (Mann-Whitney).
To adjust the p values for multiple testing Holm's procedure was used.(23)
All of the above-specified analyses were complete participant analyses. If a significant effect of the intervention after adjustment for multiple testing was noted, three sensitivity analyses were carried out with a Mann-Whitney test using mean values of the primary and secondary outcome: (1) the worst case scenario (missing value imputed by most pessimistic value from opposite group), (2) strong bias (missing value imputed by most pessimistic value in the group to which it belonged), (3) mild bias (missing value imputed by mean value found during the complete participant analysis of the group to which the missing value did not belong). These had increasing degree of bias by replacing missing values with constructed ones reflecting the degree of skepticism of the observed effect.

Figures are presented in error plots showing mean and 95% confidence intervals.
Post hoc analysis examined if sex and/or computer game experience interacted with the intervention. Questionnaire replies concerning perception of own surgical skills after the trial was analyzed with a Mann-Whitney test.

Results

Primary and secondary outcome measures
Participant enrolment and demographics are presented in figure 1. All participants reached the 'expert level' within a two-month period. Average time per training session was two hours and twenty minutes.

Table 2 shows the primary outcomes; number of repetitions and time (min.) to reach the 'expert level'. It was necessary to log transform the data in order to obtain Gaussian distributions and variance homogencity. Mean of ln(repetitions) (mean of the control group minus mean of the intervention group) was 0.80 (95% CI 0.5 to 1.03, $p < 0.0005$), table 2. Mean (ln(min.) (mean of the control group minu mean of the intervention group was 0.73 (95% CI 0.544 to 0.924, $p < 0.0005$)), table Both repetitions and time were significant higher in the in control group than in th intervention group.
The performance score (%) *at* 'expert lev was significantly lower in the interventio group (mean of control group minus mea of intervention group) 7.23 (95% CI 2.34 t 12.1, p=0.004), table 2.
Similar results were obtained when th analysis was repeated with the protoc specified co-factors included. Results wer still highly significant when adjusted fo multiple testing.
Three participants dropped out from th control group in frustration. However, th primary outcomes, repetitions and tim both still differ highly significantly betwee the two groups in all three sensitivit analyses (p=0.005 in all three analyses). Th analysis of the performance score (% obtained at 'expert level' was not significa in the worst-case scenario (p=0.24) an strong bias scenario (p=0.083), bu significant in mild bias (p=0.034) an original data set scenarios (p=0.004).

Responses from the questionnaire showe that participants' perception of own surgic skills after the trial was significantly highe in the intervention group (p=0.011).

Post hoc analysis regarding sex and comput gaming skills
Post hoc analysis showed that men i general used significantly less time to reac 'expert level' than women (p=0.037), but n significant difference was observed i regards to number of repetitions (p=0.20 There was significant interaction betwee sex and the intervention (p=0.044); th effect of intervention seems to be mo pronounced in women than in men, figure Borderline significance (p=0.051) wa

shown in regards to repetitions in favor of men.

No significant interaction was found regarding computer game experience and the outcome measures time and repetitions (p=0.83 and p=0.88). Participants with computer game experience had a significantly higher performance score (%) (p=0.011).

Feedback requests

Two participants did not request the second feedback session; one feedback session was sufficient. The second feedback session was requested by 46 participants (96%) and placed around the tenth repetition (range 6 to 11 repetitions). The third feedback session was requested by 17 participants (35%) and was unevenly distributed from the 14^{th} to the 50^{th} repetition. The average time spent on the first, second, and third feedback session was 11, 11, and 8 minutes.

Discussion

Impact of instructor feedback

This randomized trial revealed that instructor feedback increases the efficiency regarding the amount of time and number of repetitions needed to reach a predefined proficiency level for a complex simulated operation. The intervention group used less time and fewer repetitions, and results were overall more homogeneous compared with the control group. However, the majority of the control group succeeded in reaching the predefined proficiency level, i.e., 'expert level' without instructor feedback. We expected that for this group of novices practicing a complex task the cognitive load would be overwhelming. That was also the case for three participants in the control group who dropped out due to frustration of not being able to reach proficiency level. The control group, however, reached a significantly higher performance scores than the intervention group, although at a significantly slower pace. This finding was not unexpected; the performance score is a measure based on time and accuracy and increases with training, and in average, the control group used twice the amount of time training.

The control group assessed their own surgical skills significantly lower after the trial compared with the intervention group. The validity of self-assessment is disputable and there is a difference between confidence and competence; improvement in confidence is not necessarily translated into better competence and better outcome.(24,25) Nonetheless, the point of extensive skills training is not for the trainees to feel a decline of skills, which is an additional argument for supplementary feedback. Since there were no dropouts from the intervention group it is likely that feedback would have allowed dropouts from the control group to reach 'expert level'.

Learning of simple and basic laparoscopic tasks, such as coordination and instrument navigation, without instructor provided augmented feedback has previously been demonstrated in a randomized trial to be more effective than instructor controlled learning.(14) These basic tasks are relatively easy to accomplish intuitively, which could explain the different findings in our randomized trial where the task was a complex operation involving both knowledge and motor skills. It is feasible that trainees who initially train on the virtual reality simulator could learn simple tasks without instructor feedback. However, on the basis of our findings, when proceeding to more complex tasks we recommend feedback from an instructor to ensure time-efficient and correct learning.

Boyle et al. found that standardized feedback was associated with significantly fewer errors and improved learning curve when performing a hand-assisted laparoscopic colectomy on a the hybrid simulator ProMIS (a hybrid simulator is a video-trainer which provides feedback) among 3-5-year surgical trainees.(9) However, Boyle et al. did not report how time-consuming or how frequent feedback was provided. Boyle et al. argues that feedback could be provided by non-surgical staff, given the facilitator is sufficiently familiar with the procedure and simulator,(9) which is an argument we endorse but need to explore further. Another randomized trial where the intervention group had access to virtual reality training (but no feedback) and the control group did not train on a virtual reality simulator, indicated that simulator training in a non-supervised setting may not be sufficient to increase laparoscopic suturing skills.(26)

In contrast with previous studies,(27,28) computer gaming experience was not a significant predictor of time and number of repetitions.

No baseline on the virtual reality simulator between the two groups was assessed since this familiarity with the virtual reality simulator might have contaminated the outcome. Additionally, due to the nature of the randomization and the fact that there was no difference in operational experience between the two groups, it is reasonable believe that the results reflect the true observed differences. All participants received the same standardized information regarding the operational technique for a salpingectomy at the pre-trial introductory meeting along with review of the instruction video generated by the virtual reality simulator. However, whether the two groups understanding of the procedure was the same is unknown, nevertheless, since it is a homogeneous group of participants it is reasonably fair to state that they had the same minimum knowledge level.
We did fully obtain the planned sample size, but few participants had missing values during the trial. The deviation of participants was very small and we do not believe it affects our results.

Participants indicated that instructor feedback positively affected the trainees' self-perception of surgical skills. This finding i in accordance with other studies focusing o reactions from trainees using simulatio based training. Among the advantages foun were: improved self-confidence and sel efficacy, and improved feeling of bein proficient.(29,30) However, there were als some indications of drawbacks, whic include high levels of anxiety an stress.(31,32) Little is known about whic personal and contextual factors facilitate o impair transfer of learning from simulation-based setting to the clinic setting, and the next ideal step would be t demonstrate improved clinical performanc when having received standardize feedback on a virtual reality simulator. On randomized trial within endoscopic trainin of general surgery trainees proved bette outcome on colonoscopy when systemat feedback on colonoscopy performance wa applied,(33). It is unknown whether th skills required for colonoscopy an laparoscopy are the same; nonetheless, it i reasonable to extrapolate the use of sy tematic feedback to other specialties.

Sex differences in surgical virtual realit training

Men spent significantly less time reachin the predefined proficiency level tha women, but no difference was foun regarding number of repetition Additionally, we found that the feedbac influenced the women's performance mo compared with that of men. These finding are in accordance with several other studie and yet this difference remains largel unrecognized and, consequently, unaddres sed.(27,34,35) In an environment wher laparoscopic training is increasing, it i important that surgical curricula acknow ledge differences between sex to ensure fai and personalized training opportunities.

Predefined proficiency level – is the bar s high enough or should we 'over train'?

Interestingly, we found that the contr group reached the 'expert level' with significantly higher performance score compared with the intervention group. One interpretation of this is that 'practice makes perfect' since on average the control group performed twice as many operation modules and used double amount of time practicing.

Since the introduction of virtual reality simulators in surgical training ten years ago, the optimal endpoint for simulation training has been a predefined proficiency level based on that of expert surgeons, i.e., surgeons who have performed numerous operations.(15-17) The current standard is that trainees have to accomplish this level once or twice, depending on the set up; in our trial passing twice was applied. 'Overtraining' of intracorporeal suturing has shown to be beneficial on the learning curve,(36) and an important determinant of skills and knowledge retention is the amount of 'over learning' or additional training beyond that required for proficiency.(37) Speculations could be made on whether a pre-set proficiency level based on experts' performance is a sufficient level for training. Based on our findings, we suggest that when planning a surgical virtual reality curriculum an amount of over training or continuous training should be considered.

Self-directed feedback

A self-directed learner takes responsibility for knowledge production by becoming behaviorally and metacognitively active, and increased autonomy probably allows the participant to tailor knowledge production to his or her specific needs.(19,38) Self-directed access to instruction or feedback has within the sports domain demonstrated higher learning outcomes compared with participants whose access is controlled externally.(19,20,39) On the basis of this, we let the intervention group decide themselves when they requested feedback. Contrary to the motor skills research in sports, which found scattered feedback requests, our results showed feedback requests clustered around the tenth repetition of training. The difference could be explained by the fact that an operation is considered high stakes and calls for feedback sooner, whereas in basketball (19) it is easier to do 'trial and error' without it having perceived consequences. Surprisingly, only one third of the participants wanted the third feedback session; two sessions appeared sufficient. To our knowledge, no prior studies have focused on the optimal time to provide feedback in surgical virtual reality training. Our findings with feedback placed around the first and tenth repetition could be used as a guideline, though it needs to be examined further whether this result reflects a general tendency in simulator training.

Study strengths and limitations

One strength of this randomized trial is the relatively large sample size compared with previous studies on efficiency of virtual reality simulator training. This reduces the risk of random errors.(40) Another strength is the conduct of central randomization as well as blinded outcome assessment, i.e., outcome produced by the virtual reality simulator, which reduces the risk of selection bias and assessment bias.(40) Instructor feedback was standardized to ensure consistency and reproducibility, and can therefore be replicated.

The major limitation is that the sample was senior medical students and hence generalizability of results to first-year trainees to whom a virtual reality-training curriculum would apply could be a problem. The participating senior medical students had completed all anatomy courses and the mandatory six-month surgical stay placed towards the end of medical school, and in several contexts the trial participants actually resemble first year trainees since they often have no prior laparoscopic training either. Moreover, the participants were recruited from special interest groups. In comparison with a previous trial by the same author group where first and second year residents performed the same operational virtual reality task (and also received feedback), the average number of repetitions used to reach the predefined proficiency level is almost identical to the average number used by the intervention group.(4) Whether the senior medical students is a true resemblance to first-year residents is unknown, thus we feel confident that motivation to learn this kind of complex technical skills was equally high among the students as in first-year postgraduate trainees.

Future perspectives

Research within the sports domain has shown that augmented feedback can have a dramatic effect during training. Yet, whether this effect reflects a sustainable capacity to perform the task needs to be demonstrated in retention or transfer studies, e.g., performance of a similar but different tasks or performance in clinical practice on real patients. Both these aspects are topics for future research.

Conclusion

Instructor feedback increases efficiency when training a complex operational task on a virtual reality simulator; time and repetitions used to achieve a predefined proficiency level were significantly reduced in the group that received instructor feedback compared with the control group.

Acknowledgements: We thank Birgitte Schneider, Rigshospitalet Copenhagen University Hospital, for thoroughly designing a secure database.

Contributers: JO (principal investigator) acquired the data, drafted the paper, and obtained funding. JO, TG, MM, CRL, CR, JLS, BO and CG contributed to the trial design. PW, JO and CG conducted the statistical analyses. FB acquired data and monitored trial participants. All authors provided administrative support, analyzed and interpreted the data, and critically revised the manuscript.

Funding: The Copenhagen Universi Hospital Rigshospitalet, and TrygFonden, non-profit organization, supported t project. The contents of the manuscript a solely the responsibility of the authors a do not necessarily represent the offic views of Rigshospitalet or TrygFonden.

Competing interests: None declared.

Ethical approval: The trial complies wi the Helsinki Declaration on biomedic research. The Danish National Committee o Biomedical Research Ethics evaluated a approved the trial (journal number: H- 2010-082). The Danish Data Protectio Agency approved collection, analysis a storage of data, approval code 2007-5 0015/30-0996.

All participants were provided writt information on the trial. Participation w voluntary; no material goods were given participants. The trial is registered clinicaltrials.gov with trial registrati number: NCT01497782.

References

1. Aggarwal R, Ward J, Balasundaram I et al. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg.* 2007;246:771–779.

2. Ahlberg G, Enochsson L, Gallagher AG et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg.* 2007;193:797–804.

3. Grantcharov TP, Kristiansen VB, Bendix J et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg.* 2004;91:146–150.

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

5. 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;236:458–63.

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

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

8. Schout BMA, Hendrikx AJM, Scheele F et al. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surg Endosc.* 2010;24:536–546.

9. Boyle E, Al-Akash M, Gallagher AG et al. Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgrad Med J.* 2011;87:524–528.

10. O'Connor A, Schwaitzberg SD, Cao CGL. How much feedback is necessary for learning to suture? *Surg Endosc.* 2008;22:1614–1619.

11. Sewell C, Morris D, Blevins NH et al. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg.* 2008;13:63–81.

12. Van Sickle KR, Gallagher AG, Smith CD. The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. *Surg Endosc.* 2007;21:220–224.

13. Ende J. Feedback in clinical medical education. *JAMA* 1983;250:777–781.

14. Snyder CW, Vandromme MJ, Tyra SL et al. Proficiency-based laparoscopic and endoscopic training with virtual reality simulators: a comparison of proctored and independent approaches. *J Surg Educ.* 2009;66:201–207.

15. Aggarwal R, Grantcharov T, Moorthy K et al. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *Am J Surg.* 2006;191:128–133.

16. Brunner WC, Korndorffer JR, Sierra R et al. Laparoscopic virtual reality training: are 30 repetitions enough? *J Surg Res.* 2004;122:150–156.

17. Gallagher AG, Ritter EM, Champion H et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg.*

2005;241:364–372.

18. Brinkman WM, Buzink SN, Alevizos L et al. Criterion-based laparoscopic training reduces total training time. *Surg Endosc.* 2012;26:1095-101.

19. Wulf G, Raupach M, Pfeiffer F. Self-controlled observational practice enhances learning. *Res Q Exerc Sport* 2005;76:107–111.

20. Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. *Med Educ.* 2010;44:75–84.

21. Oestergaard J, Bjerrum F, Maagaard M et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized educational trial. *BMC Med Educ.* 2012;12:7.

22. Larsen CR, Grantcharov T, Aggarwal R et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. *Surg Endosc.* 2006;20:1460–1466.

23. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979;6:65-70.

24. Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. *Adv Health Sci Educ Theory Pract.* 2002;7:63–80.

25. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. *JAMA.* 2006;296:1137–9.

26. Halvorsen FH, Fosse E, Mjåland O. Unsupervised virtual reality training may not increase laparoscopic suturing skills. *Surg Laparosc Endo Percutan Tech.* 2011;21:458–461.

27. Grantcharov TP, Bardram L, Fund Jensen P et al. Impact of han dominance, gender, and experien with computer games on performan in virtual reality laparoscopy. *Su Endosc.* 2003;17:1082–1085.

28. Van Hove C, Perry KA, Spight DH et Predictors of technical sk acquisition among resident trainees a laparoscopic skills educatio program. *World J Surg.* 2008;32:191 1921.

29. Cass GKS, Crofts JF, Draycott TJ. Th use of simulation to teach clinic skills in obstetrics. *Semin Perinat* 2011;35:68–73.

30. Zigmont JJ, Kappus LJ, Sudikoff S Theoretical foundations of learnin through simulation. *Semin Perinat* 2011;35:47–51.

31. Bong CL, Lightdale JR, Fredette ME al. Effects of simulation vers traditional tutorial-based training o physiologic stress levels amon clinicians: a pilot study. *Simul Health* 2010;5:272–278.

32. Harvey A, Nathens AB, Bandiera G al. Threat and challenge: cogniti appraisal and stress responses simulated trauma resuscitations. *Me Educ.* 2010;44:587–594.

33. Harewood GC, Murray F, Winder S al. Evaluation of formal feedback o endoscopic competence amon trainees: the EFFECT trial. *Ir J Med S* 2008;177:253–256.

34. Elneel FHF, Carter F, Tang B et Extent of innate dexterity an ambidexterity across handedness an gender: Implications for training

laparoscopic surgery. *Surg Endosc.* 2008;22:31–37.

35. Thorson CM, Kelly JP, Forse RA et al. Can we continue to ignore gender differences in performance on simulation trainers? *J Laparoendosc Adv Surg Tech A.* 2011;21:329–333.

36. Kolozsvari NO, Kaneva P, Brace C et al. Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention. *Surg Endosc.* 2011;25:2063–2070.

37. Kneebone RL, Scott W, Darzi A et al. Simulation and clinical practice: strengthening the relationship. *Med Educ.* 2004;38:1095–1102.

38. Brydges R, Carnahan H, Safir O et al. How effective is self-guided learning of clinical technical skills? It's all about process. *Med Educ.* 2009;43:507–515.

39. Chiviacowsky S, Wulf G. Self-controlled feedback is effective if it is based on the learner's performance. *Res Q Exerc Sport* 2005;76:42–48.

40. Keus F, Wetterslev J, Gluud C et al. Evidence at a glance: error matrix approach for overviewing available evidence. *BMC Med Res Methodol.* 2010;10:90.

Table 1
Based on 11 variables listed below the virtual reality simulator generates a performance score (%) on the operational task; *right-side salpingectomy*, which is available for all participants after each repetition. When all variables are within the passing range the predefined proficiency level, i.e. 'expert level', is reached. 'Expert level' is set and validated in a previous study (22). The performance score is calculated on the basis of weighting of the different parameters; the better the individual parameter, the better performance score.

Variable	**Passing range**	**Weight in calculating performance score**
Total time	>280 (s)	15
Blood loss	>180 (ml)	15
Pool volume	>10 (ml)	0
Ovary diathermy damage	>3 (s)	5
Tube cut: uterus distance	>4 (mm)	5
Bleeding vessel cut	0	Fail if performed
Evacuation from body	>1	Fail if not performed
Left instrument path length	>2 (m)	15
Left instrument angular path	>350 (degrees)	15
Right instrument path length	>3 (m)	15
Right instrument angular path	>450 (degrees)	15

Table 2
Mean differences in primary and secondary outcome measures between the intervention group and the control group

	Intervention group (n=48)	**Control group** (n=43)	P value
Mean number (#) of repetitions used to reach the 'expert level' (CI)	29 (23.9 to 33.5)	65 (53.9 to 75.5)	p<0.0005
Mean time (minutes) used to reach the 'expert level' (CI)	162 (140 to 183)	342 (285 to 398)	p<0.0005
Performance score (%) *at* 'expert level' (CI)	49 (45 to 53)	57 (53 to 60)	P=0.004

CI=95% Confidence Interval

Figure 1
Participant enrolment (following the CONSORT Statement) and demographics. The distribution of sex did not differ significantly between dropouts and those participating in the trial (p=0.46) nor did presence of computer game experience (p=0.25).
VR = virtual reality

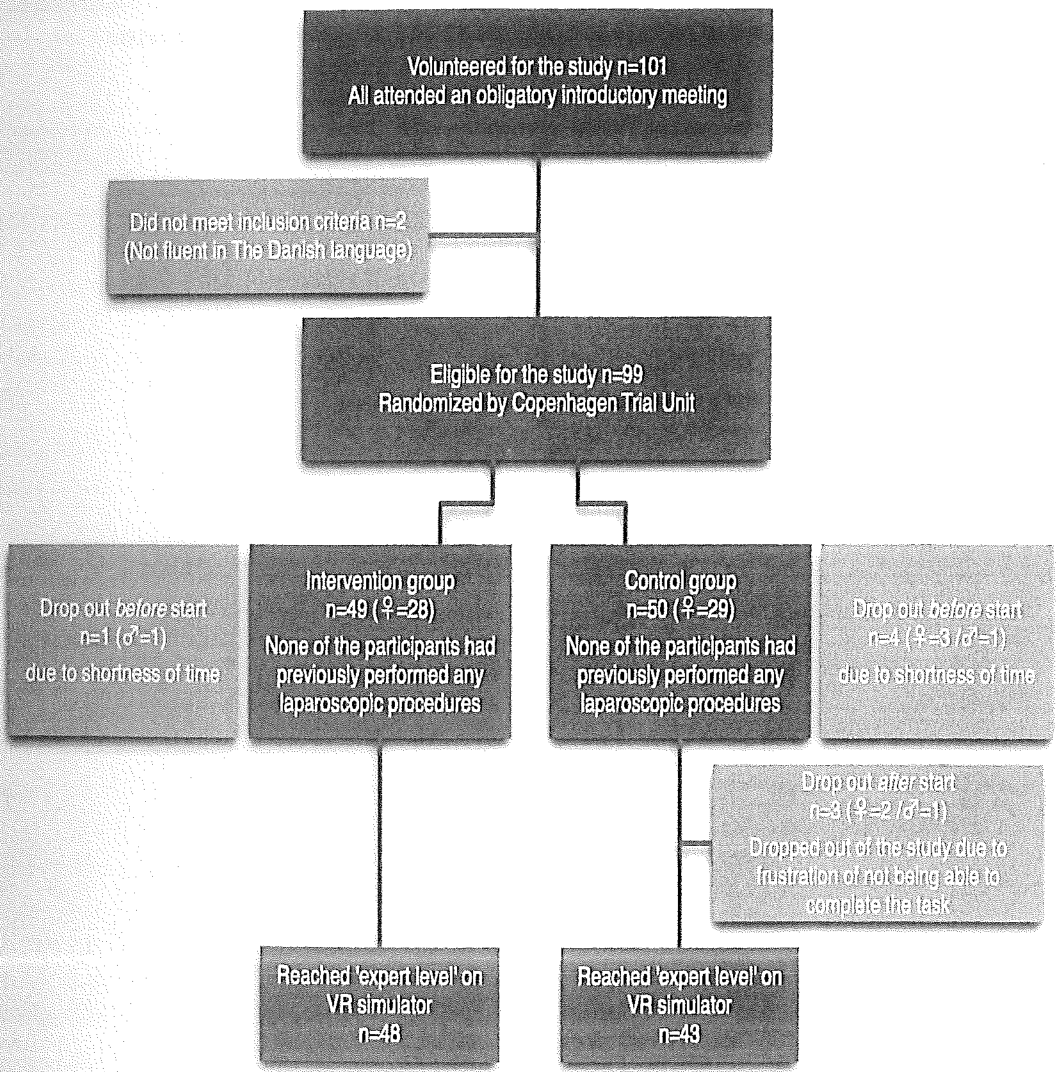
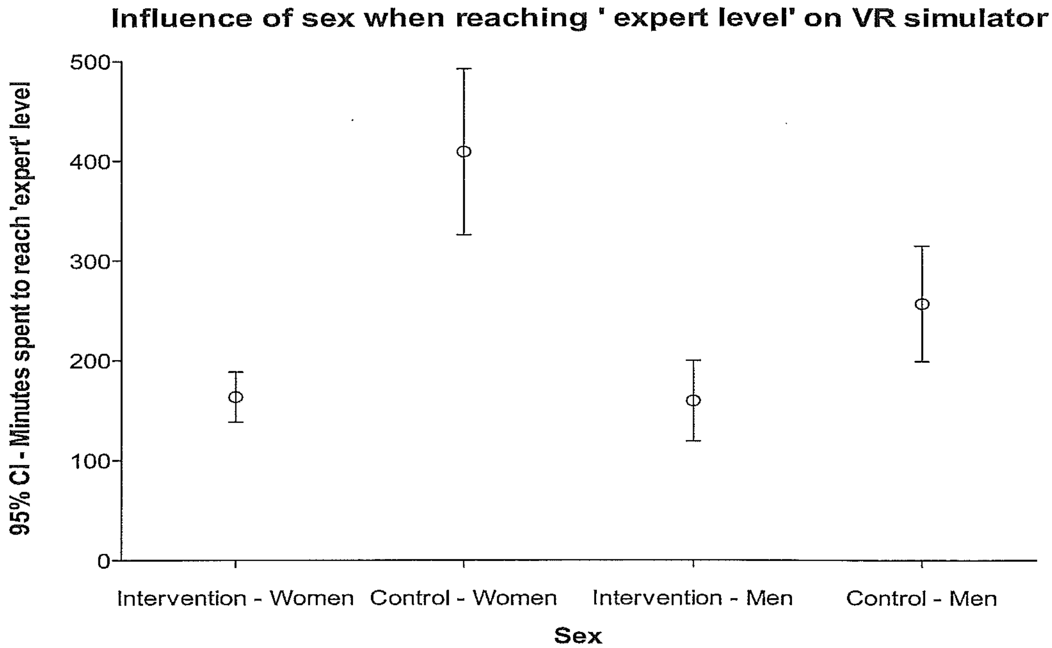


Figure 2
Comparison of time used to reach 'expert level' between men and women. The effect of intervention is significantly better in females than in males (p=0.044). Women in the intervention group: mean 164, 95% CI 139 to 189, women in the control group: mean 409, 95% CI 326 to 493. Men in the intervention group: mean 159, 95% CI 119 to 200, men in the control group: mean 257, 95% CI 199-315.
VR = virtual reality



Study III. The assessment study

'Can both residents and chief physicians assess surgical skills?'
Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL.
Surgical Endoscopy 2012 Jul;26(7):2054-60

Surg Endosc
DOI 10.1007/s00464-012-2155-1

Can both residents and chief physicians assess surgical skills?

Jeanett Oestergaard · Christian Rifbjerg Larsen · Mathilde Maagaard · Teodor Grantcharov · Bent Ottesen · Jette Led Sorensen

Received: 8 August 2011 / Accepted: 22 December 2011

Abstract
Background It is known that structured assessment of an operation can provide trainees with useful knowledge and potentially shorten their learning curve. However, methods for objective assessment have not been widely adopted into the clinical setting. This might be because of a lack of expertise using an assessment tool. The aim of this present study was to investigate if a validated laparoscopic procedure-specific assessment tool could be used by doctors with different levels of experience.
Methods The study was conducted as an observer-blinded, prospective cohort study. Three video recordings of a right-side laparoscopic salpingectomy were distributed to ten chief physicians, eight residents (fourth year trainees), and two expert assessors (all in gynecology) in order to be assessed using a validated procedure-specific assessment tool. The three salpingectomies were selected because they easily showed the different operational levels: novice, intermediate, and expert. The two expert assessors, i.e., our gold standard, were familiar with the OSA-LS assessment scale, but the chief physicians and the residents were not. All participants were blinded to the fact that surgeons with different experience had performed the salpingectomies.
Results No significant differences between the residents and chief physicians were observed in any of the three assessed operations: novice, $p = 0.63$; intermediate, $p = 0.93$; and expert, $p = 0.93$. The chief physicians and residents matched our gold standard in assessing the intermediate operation ($p = 0.177$), but not the novice operation ($p = 0.005$) or the expert operation ($p = 0.001$).
Conclusions Residents and chief physicians generated similar performance scores when assessing operations using a laparoscopic procedure-specific assessment scale, and they could distinguish performance levels between the surgeons. They matched the assessment score of our expert on the intermediate operation. We conclude that a procedure-specific assessment scale can be used by both residents and chief physicians when giving formative feedback.

Keywords Assessment · Procedure-specific rating scale · Gynecology · Laparoscopy · Salpingectomy · OSATS

J. Oestergaard (✉) · C. R. Larsen · M. Maagaard · B. Ottesen · J. L. Sorensen
Department of Gynaecology and Obstetrics, The Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University Hospital of Copenhagen, Section 4074, Door 7248, Blegdamsvej 9, 2100 Copenhagen, Denmark
e-mail: jeanett.oestergaard@rh.regionh.dk; jeanett78@gmail.com

T. Grantcharov
Department of Surgery, St. Michael's Hospital, University of Toronto, Toronto, ON, Canada

Published online: 20 January 2012

An increasing number of surgical procedures are currently performed by laparoscopic techniques. This leads to an increasing demand for evidence-based assessment of laparoscopic skills to ensure proper education. Methods for objective assessment are needed to assess the performance of surgeons and are of great importance in an era with intense focus on training [1–3]. This has increased the focus on the need for objective and independent assessment of surgical skills, not only on-site assessment, but also outside the operating room.

A number of different methods for objective assessment of surgical skills have been developed throughout the last decade, and several studies have addressed their validity and reliability [1, 3–5]. It is known that structured assessment of an operation can provide the trainees with useful knowledge on how to improve their skills and potentially shorten their learning curve [6]. However, methods for objective assessment have not been widely adopted into the clinical setting because of a lack of expertise in using an assessment tool, a lack of proper infrastructure for implementation, and cost [3].

We investigated whether doctors with different levels of experience could use a laparoscopic procedure-specific assessment tool. The assessment tool used was developed and validated by Larsen et al. [7] and is a procedure-specific rating scale called objective structured assessment-laparoscopic salpingectomy (OSA-LS). The OSA-LS assessment scale is a procedural modification of the objective structured assessment of technical skills scale (OSATS) [8, 9], specifically developed to assess laparoscopic operations.

Traditionally, surgical skills have been assessed on-site, during the actual operation, by supervision and feedback [4]. An advantage to the OSA-LS assessment scale is that it can be used in both real time and video-based, the latter providing the possibility to assess the performance whenever it is convenient. In addition, this study explored the subjective perceptions in relation to assessment of operations performed by peers.

Material and methods

Eight residents in gynecology (in their fourth year of 5), having performed an average of 36 independent laparoscopic procedures (range = 20–70), and ten chief physicians in gynecology, having performed an average of >450 independent laparoscopic procedures (range = 250–1,000), represented the two groups. The study took place from April to July 2009. The residents group comprised all fourth-year residents in the region, and the chief physicians were pointed out by the local head of the department as requested by the study group.

All participants received three DVDs by mail, each containing an unedited video-recording of a right-side laparoscopic salpingectomy: one from a novice surgeon (defined as having performed <10 individual laparoscopic procedures), one from a surgeon with intermediate experience (defined as having performed 20–50 individual laparoscopic procedures), and one from an expert surgeon (defined as having performed >200 individual laparoscopic procedures). The three salpingectomies were selected because they easily demonstrate the different surgical skill levels: novice, intermediate, and expert. The participants also received three copies of the OSA-LS assessment scale. None of the participants had used the OSA-LS assessment scale prior to the study. The residents and the chief physicians were blinded to the training level of the surgeon performing the salpingectomy. Both groups were requested not to discuss the assessment of the surgical procedur with peers while the study was in progress.

To make sure that there were no mistakes by assessi the same video twice, both the DVD containing the op ation and the OSA-LS assessment scale were match using a color code.

The participants answered a questionnaire with que tions on the following: (1) time spent assessing the thr operations; (2) whether they had learned from watching t videos; (3) confidence with the option to give oral written feedback to the surgeon; (4) applicability of t OSA-LS assessment scale; (5) thoughts on colleagu assessing oneself; and (6) the possibility of improvement their own surgical performance was assessed.

The three video-recorded operations were assessed i previous study by two expert assessors who identified s nificant differences in the levels of surgical performance [The experts' assessments were used as our "gold standar assessment of the three operations. Like the residents a chief physicians, the experts had no knowledge of the d ferent training levels of the surgeons performing the s pingectomies. The two experts were senior chief physicia and codevelopers of the OSA-LS assessment scale [7]. Th used the OSA-LS assessment scale on more than 60 vid recorded salpingectomies. In addition they have perform more than 1,000 laparoscopies each. Consequently, decided to use their assessment as our gold standard.

Like the OSATS [8, 9], the OSA-LS assessment sc contains five items that rate general skills with respect general surgical items and five items on skills specific used in a LS. Each of the ten items is evaluated on a glo 5-point Likert-like scale, where the lowest, middle, highest scores are defined by explicit descriptions of p formances (Table 1). Larsen et al. [7] demonstrated that OSA-LS assessment scale was feasible with high constr validity and a high interrater reliability of 0.83.

In two of the operations, item 10 on the OSA assessment scale (removal of the dissected fallopian t from the abdomen) was not on the video-recorded op tion and was therefore removed from the assessment sc on all three operations. Consequently, the maximum sc of each operation could sum up to 45 points.

Statistics

For each operation the assessment scores of the resid group were compared to the assessment scores of chief physicians group with a nonparametric two-sam Wilcoxon test. When comparing the two groups with gold standard, one-sample Wilcoxon was used. A pa Wilcoxon test was used to evaluate whether the gro could distinguish between the different surgical per mance levels.

Table 1 The OSA-LS assessment scale, BJOG 2008 115:908–16

OSA-LS general skills	1	2	3	4	5
1. Economy of movements	Many unnecessary moves		Efficient motion but some unnecessary moves		Maximum economy of movements
2. Confidence of movements: instrument handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
3. Economy of time	Too long time used to perform sufficiently		Intermediate time used to perform sufficiently		Minimal time used to perform sufficiently
4. Errors: respect for tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling if tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
5. Flow of operation/ operative technique	Imprecise, wrong technique in approaching the operative interventions		Careful technique with occasional errors		Fluent, secure and correct technique in all stages of the operative procedure
OSA-LS specific skills	**1**	**2**	**3**	**4**	**5**
1. Presentative of anatomic structures	Poor retraction & exposure of fallopian tube and round ligament		Satisfactory retraction & exposure of fallopian tube and round ligament		Expert retraction & exposure of fallopian tube and round ligament
2. Dissection of fallopian tube	Inadequate dissection of fallopian tube, additional damage or insufficient tube cut distance		Identified fallopian tube Adequate dissection little damage of other structures		Clearly identified fallopian tube, perfectly dissected no additional damage.
3. Use of diathermy	Using diathermy too close to healthy ovarian or other tissue, risk of damage		Mostly safe use minimal risk of damage		Perfectly safe use of diathermy, no risk of damage
4. Care for Ovarian artery and pelvic wall	Using diathermy or cutting too close to ovarian artery high risk of bleeding or occlusion of vessel.		Mostly safe use of instruments, low risk of arterial damage		Perfectly safe use instruments, no risk of cauterizing or cutting the ovarian artery
5. Extraction of fallopian tube	Clumsily done with major difficulty to catch the tissue, retract or get the tissue in the Endobag		Minor difficulty retracting or getting the tissue in the bag		Perfect retraction or, easy placement of tube in bag

In the questionnaire, ratings were measured with a Likert-like scale ranging from 1 to 5, where 1 = strongly disagree, 2 = partly disagree, 3 = neither, 4 = partly agree, and 5 = strongly agree. When comparing the questionnaire responses a Mann–Whitney test was used.

Because of the sample size and the nature of the results, a Gaussian distribution could not be expected.

A p value (two-tailed) of <0.05 was considered to be statistically significant. Analysis was performed using the Statistical Package for Social Sciences for Windows (SPSS Inc., Chicago, IL, USA). Graphics was made using Graphpad Prism 4.0 for Windows (Graphpad Software Inc., San Diego, CA, USA).

Ethics

All the participants volunteered and no ethical approval was necessary. The video-recorded operations were reused from another study, for which no ethical approval was needed according to the Danish National Committee on Biomedical Research Ethics [7]. No personal data were displayed in the video-recorded operations.

Results

The residents, the chief physicians, and our gold standard experts were able to discriminate the three levels of surgical experience (Fig. 1). For the novice operation, the residents' assessment varied from 11 to 22 points (median = 17, SD = 4.1) and the chief physicians' assessment varied from 12 to 21 points (median = 15.5 points, SD = 2.8). For the intermediate skill level operation, the residents' assessment varied from 21 to 36 points (median = 27.5, SD = 4.6) and the chief physicians' assessment varied from 20 to 33 points (median = 27.5, SD = 4.0). For the expert skill level operation, the residents' assessment varied from 21 to 45 points (median = 36.5, SD = 7.6) and the chief physicians' assessment varied from 21 to 45 (median = 37.5, SD = 8.1). No significant differences between the two groups were

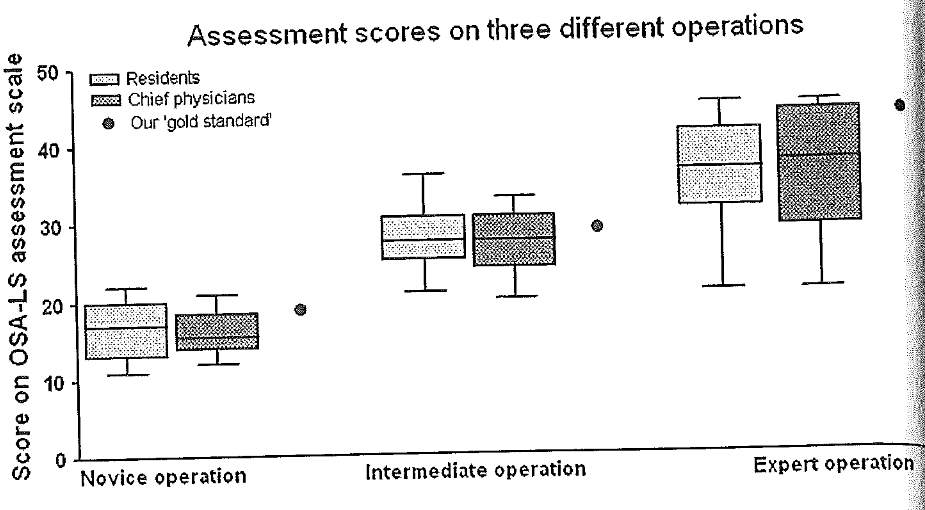


Fig. 1 The residents and the chief physicians generated similar performance scores and could distinguish the different operational levels. They matched the score of our 'gold standard' on the intermediate operation. Box plot median score all groups, *band* represents median; *boxes* represents IQR; *whiskers* represents range

observed in any of the three assessed operations: novice, $p = 0.63$; intermediate, $p = 0.93$; and expert $p = 0.93$.

Our gold standard assessment on all three operations was a mean of the two experts' assessments. The expert assessors had exactly the same assessment score for all three operations: 19 for the novice operation, 29 for the intermediate operation, and 44 for the expert operation.

Since no significant differences were observed between the residents' and the chief physicians' assessment scores, the data were pooled and compared to our gold standard. Both groups (the residents and the chief physicians) matched our gold standard for the intermediate operation ($p = 0.177$) but not for the novice operation ($p = 0.005$) or the expert operation ($p = 0.001$).

Table 2 shows the internal consistency, Cronbach's α coefficient, for each operation assessed by the residents and chief physicians. The internal consistency is calculated using the nine items assessed with the OSA-LS assessment scale.

There was no difference in time used to assess the operations between the residents and chief physicians. The average time (minutes) to assess the novice, intermediate, and expert operations was 31 (20–60), 24 (15–70), and 20 (10–60), respectively. The exact runtime was 22, 11, and 7 min for novice, intermediate, and expert, respectively. The results from the questionnaire are given in Tables 3 and 4.

Table 2 Internal consistency (Cronbach's α) of the three operations using nine items on the OSA-LS assessment scale

Operation	Residents	Chief physicians
Novice	0.79	0.69
Intermediary	0.85	0.84
Expert	0.90	0.95

Discussion

Assessment of operations

We found that residents and chief physicians genera similar assessment scores for a video-recorded lapa scopic operation using the OSA-LS assessment sc Furthermore, they could distinguish operations perform by a novice, an intermediate skill level surgeon, and expert surgeon. This demonstrates that this procedu specific assessment scale is usable for senior residents chief physicians, and it can be a practical tool to facili standardized and specific feedback and assessment.

An uncomplicated LS is a procedure that residents obstetrics and gynecology meet early on in their trai and would be an obvious operation for which to rece feedback. Therefore, it is an advantage that doctors different experience levels can provide structured feedb on technical skills using the OSA-LS assessment scale support of this, reviews have shown the effectivenes peer-assisted teaching, i.e., teaching by peers of sim groupings [10], which, in combination with the OSA assessment scale, could be useful for resident training.

The internal consistency, which estimates the reliab of the test scores, is considered good if above 0.8 excellent if above 0.9. We found high internal consist scores from both the residents and the chief physician the intermediate and expert operations and accept scores for the novice operation. The latter could explained by small sample size or that a novice operati more difficult to assess due to large variation in operati performance.

This study did not investigate the interrater agreem as this has been thoroughly tested in a previous study found to be high [7].

Only the assessment of the intermediate opera matched our gold standard's assessment. For the no

Table 3 Questionnaire concerning assessment of operations on different levels

Operation by	Residents			Chief physicians		
	Novice	Intermediate	Expert	Novice	Intermediate	Expert
Did the assessor learn something about operational procedures and techniques themselves, median (range)	2.0 (1–4) S	3.0 (1–4) S	4.0 (2–4) S	1.0 (1–4) S	1.0 (1–3) S	2.0 (1–5) S
Would the assessor feel self-confident if he/she were to give oral feedback about the operation to the surgeon, median (range)	4.0 (3–5) NS	4.5 (4–5) NS	4.0 (2–5) NS	4.0 (3–5) NS	4.0 (3–5) NS	3.0 (2–5) NS
Would the assessor feel self confident if he/she were to give written feedback about the operation to the surgeon, median (range)	4.0 (3–5) NS	4.0 (4–5) NS	4.0 (2–5) NS	4.0 (3–5) NS	4.0 (3–5) NS	3.0 (2–5) NS

The assessors self-rated the questions on a scale of 1–5, where *1* strongly disagree, *2* partly disagree, *3* neither, *4* partly agree, *5* strongly agree

NS nonsignificant, indicating there is no difference between the two groups, *S* significant, indicating that the two groups are of different conception

Table 4 Questionnaire concerning assessment of operations

	Residents	Chief physicians
Need for instruction of the OSA-LS assessment scale before use, median (range)	1.0 (1–2) NS	1.0 (1–4) NS
The thought of a colleague assessing your operations is unpleasant, median (range)	1.0 (1–4) NS	1.0 (1–4) NS
Your surgical skills will improve if your operations were assessed in a structured way, median (range)	4.0 (4–5) S	3.0 (2–4) S

The assessors self-rated the questions on a scale of 1–5, where *1* strongly disagree, *2* partly disagree, *3* neither, *4* partly agree, *5* strongly agree

NS nonsignificant, indicating there is no difference between the two groups, *S* significant, indicating that the two groups are of different conception

and the expert operations, the residents and the chief physicians gave lower scores. This is not unexpected since our gold standard comprises experienced assessors who are better at using the full scale, giving both minimum and maximum scores. This is consistent with the literature on assessment, where it is reported that inexperienced assessors have a reluctance to use the extreme categories of a scale, so-called "central tendency bias" or "end aversion bias" [11].

A major advantage of the OSA-LS assessment scale is that the assessment is based on real human operations, thereby testing the actual performance of the surgeon. According to Miller's pyramid of competence development, the top level of evaluation takes place at what he refers to as "the Does Level," which occurs only when assessing real (human) operations [12]. The OSA-LS provides us with the ability to test the competence at the Does Level.

The use of the OSA-LS assessment scale must be seen as formative to identify different levels of performance and to monitor the inexperienced surgeon's progress over time. So far, no cutoff values have been defined when using the OSA-LS assessment scale, but this should not prevent its use for specific feedback and discussion during surgical training. Because accurate clarity about the summative function of the scale does not exist, summative assessment is not applicable [13].

Both the residents and the consultants did not express concern about using the OSA-LS assessment scale; they were comfortable using it right away. This must be interpreted as good face validity, which addresses the users' opinions about the functionality and realism of an assessment tool and emphasizes user-friendliness of the scale [13].

Video assessment

Few studies have focused on video assessment, where the methods for assessment are the same as in the live settings. Video tutorials have shown to hasten skills acquisition, teach error avoidance strategies, and help develop reflective practice in trainees [14–17]. Additionally, video assessment can distinguish between individuals with great differences in experience, but subtle variations evidently will not appear [3].

One drawback of video assessment is that it can be rather time consuming, and an argument could be made that the performance when video recorded might have an impact on the assessment outcome that would be different from an on-site assessment [3]. In vascular surgery, Beard [18] and Driscoll et al. [19] established construct validity at level 1b evidence with assessing a saphenofemoral disconnection procedure. However, the Driscoll study deals with only two videos from two surgeons (inexperienced vs. experienced). Another study found good correlation between video and live assessment; therefore, this study was of lower evidence grade: 3b [20].

We believe that there are several advantages to video assessment, e.g., the possibility of providing feedback in calm surroundings and the opportunity to keep track of surgical improvement in a standardized way.

Few studies have systematically investigated whether self-assessment through review of video-recorded operations improves subsequent performance, but it is intuitively plausible [19, 21, 22]. To our knowledge, no study has investigated whether watching video-recorded surgery in general improves the viewer's surgical performance. Results from a questionnaire in our study showed subjective gain when watching the operational videos, but this needs further investigation in the operational setting. Two small studies found that inexperienced surgical trainees significantly accelerated their acquisition of laparoscopic suturing skills simply by reviewing a video of their own performance [22, 23]. It is reasonable to extrapolate these findings to more complex tasks and anticipate a gain from the assessor when assessing videos. It is complicated to prove this conjecture but nonetheless worthwhile to consider and an interesting perspective for further studies.

Limitations

One could argue whether our gold standard was valid enough to be labeled as a true "gold standard," since it consisted of the ratings of only two experts. We believe it is plausible because the main outcome was to investigate whether residents and chief physicians were able to assess a laparoscopic operation using a procedure-specific assessment scale equally well and to compare their results with the assessment by expert assessors. To our knowledge no gold standard exists.

An important aspect of the study was the blinding of the assessors to the training level of the surgeon. Unblinded rating is less objective and can affect the reliability and validity [3]. We believe that our findings are applicable because the assessors represent a variety of different hospitals, with different operational standards, thereby limiting potential bias from one single hospital. Although the data seem consistent, a limitation to this study is the small sample size; consequently, the results have to be inte preted with caution.

Future perspectives

One of the goals in this study was to test the usability of laparoscopic procedure-specific rating scale outside th operating room in order to standardize and structure feed back. Perhaps the findings of this study could be applied other assessment scales such as the Global Operati Assessment of Laparoscopic Skills (GOALS), develope by Vassiliou et al. [24], or the OSATS [8, 9].

Based on the results this study, all first-year residents our facility are now having several operations assessed b peers using the OSA-LS assessment scale to record the surgical progress.

Conclusion

Residents and chief physicians generated similar perfo mance scores when assessing operations using a lapar scopic procedure-specific assessment scale, and they cou distinguish the performance levels of the surgeons. Th OSA-LS procedure-specific assessment scale is usable both residents and chief physicians when giving formati feedback.

Acknowledgments Thanks to Susanne Rosthoej, Department Bio Statistics, Faculty of Health Sciences, University of Copenhag for statistical support.

Disclosure Dr. J. Oestergaard, Dr. C. R. Larsen, Dr. M. Maaga Dr. T. Grantcharov, Dr. B. Ottesen, and Dr. J. L. Sorensen have conflicts of interest or no financial ties to disclose.

Funding Rigshospitalet, University Hospital of Copenhagen in Capital Region of Denmark, funded this study.

References

1. Darzi A, Datta V, Mackay S (2001) The challenge of obje assessment of surgical skill. Am J Surg 181:484–486
2. Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton Hunt EA, Pronovost PJ (2009) Beyond "see one, do one, te one": toward a different training paradigm. Postgrad Med J 244–249
3. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen Dankelman J (2010) Objective assessment of technical sur skills. Br J Surg 97:972–987
4. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic s training and assessment. Br J Surg 91:1549–1558
5. Moorthy K, Munz Y, Sarker SK, Darzi A (2003) Obje assessment of technical skills in surgery. BMJ 327:1032–10
6. Grantcharov TP, Schulze S, Kristiansen VB (2007) The impa objective assessment and constructive feedback on improve

of laparoscopic performance in the operating room. Surg Endosc 21:2240–2243

7. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B (2008) Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG 115:908–916
8. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84: 273–278
9. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W (1997) Testing technical skill via an innovative "bench station" examination. Am J Surg 173:226–230
10. Ten Cate O (2009) AMEE Guide Supplements: peer-assisted learning: a planning and implementation framework. Guide supplement 30.5 – viewpoint. Med Teach 31:57–58
11. Streiner D, Norman GR (2008) Health measurement scales: a practical guide to their development and use, 4th edn. Oxford University Press, New York
12. Miller GE (1990) The assessment of clinical skills/competence/performance. Acad Med 65:S63–S67
13. Wass V, Van der Vleuten C, Shatzer J, Jones R (2001) Assessment of clinical competence. Lancet 357:945–949
14. Scott DJ, Dunnington GL (2008) The new ACS/APDS skills curriculum: moving the learning curve out of the operating room. J Gastrointest Surg 12:213–221
15. Driscoll PJ, Paisley AM, Paterson-Brown S (2008) Video assessment of basic surgical trainees' operative skills. Am J Surg 196:265–272
16. van Det MJ, Meijerink WJ, Hoff C, Middel LJ, Koopal SA, Pierie JP (2011) The learning effect of intraoperative video-enhanced surgical procedure training. Surg Endosc 25:2261–2267
17. Tavakol M, Mohagheghi MA, Dennick R (2008) Assessing the skills of surgical residents using simulation. J Surg Educ 65: 77–83
18. Beard JD (2005) Setting standards for the assessment of operative competence. Eur J Vasc Endovasc Surg 30:215–218
19. Driscoll PJ, Paisley AM, Paterson-Brown S (2008) Video assessment of basic surgical trainees' operative skills. Am J Surg 196:265–272
20. Beard JD, Jolly BC, Newble DI, Thomas WE, Donnelly J, Southgate LJ (2005) Assessing the technical skills of surgical trainees. Br J Surg 92:778–782
21. Backstein D, Agnidis Z, Sadhu R, MacRae H (2005) Effectiveness of repeated video feedback in the acquisition of a surgical technical skill. Can J Surg 48:195–200
22. Jamshidi R, LaMasters T, Eisenberg D, Duh QY, Curet M (2009) Video self-assessment augments development of videoscopic suturing skill. J Am Coll Surg 209:622–625
23. Akl MN, Giles DL, Long JB, Magrina JF, Kho RM (2008) The efficacy of viewing an educational video as a method for the acquisition of basic laparoscopic suturing skills. J Minim Invasive Gynecol 15:410–413
24. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 190: 107–113

Study IV. The curriculum study

'A four-step curriculum in basic laparoscopy: development and validation - A curriculum blueprint'.

Strandbygaard J, Bjerrum F, Maagaard M, Larsen CR, Ottesen B, Sorensen JL
Submitted to Journal of Surgical Education, November, 2012.

Abstract

Objective It is widely accepted that initial laparoscopic skills training should be patient free and take place outside the operation room on validated simulators. However, implementation of structured training curricula is challenging and focus must be on development of practical curriculum models. The objective of this study was to develop a 4-step curriculum in basic laparoscopy consisting of validated modules integrating a procedure component.
Design A four-step curriculum was developed. The methodology was different for each step. Step 1) A 1-day course in basic laparoscopy developed on the background of a regional needs analysis. Step 2) A multiple-choice test, developed and validated through interviews with experts in laparoscopy and subsequently through a Delphi audit involving regional chief physicians. Step 3) Training a procedure (a salpingectomy) on a validated virtual reality simulator. Step 4) An operation on a patient (a salpingectomy) with following formative assessment on a validated assessment scale.
Setting Rigshospitalet, University Hospital of Copenhagen, Denmark
Participants 52 first-year residents in obstetrics and gynecology from 2009-2011.
Results Of 52 possible participants, 52 participated in step 1 and all improved post-course test scores compared with pre-course test scores, p=0.001. Step 2 was completed by 75% (37/52); all further improved test scores after six months, p=0.001. Step 3 was completed by 75%, and participants used in average 238 minutes (range 75-599) and 38 repetitions (range 8-99) to reach proficiency level on virtual reality simulator. Step 4 was completed by 55%. There was no correlation between test scores and simulator training time.
Several reasons for dropout were found, the main reason being decreased motivation due no protected training time.
Conclusion Each step of a four-step curriculum in basic laparoscopy was validated and the model was found to be applicable in residency training. Protected training time correlated with increasing completion rate.

Introduction

Currently it is widely accepted that new residents should undergo initial laparoscopic skills training outside the operating room on validated simulators to increase patient safety [1-4]. However, in spite of convincing research demonstrating the advantages of laparoscopic simulators, including virtual reality simulators [5-8], the implementation of structured training curricula remains challenging [9,10]. This is mainly due to lack of knowledge regarding the best methods for training and ideal design of curricula [9-11]. Evidence-based research on curricula designs and optimization is therefore essential.

A curriculum encompasses many aspects, such as: aims and objectives, learning outcomes, educational strategies, and assessment [12,13]. The overall aim of any curriculum is to ensure structured teaching of an agreed standard of skills or knowledge within a specific field. For a surgical curriculum to be successful it requires several elements, but it is important that it contains a cognitive component, a practice component and subsequent supervised training in the actual clinical setting [10,14,15].

Within laparoscopic curricula the Fundamental of Laparoscopic Surgery (FLS) is predominant, especially in North America. The FLS program, which has been extensively validated [16-19], addresses both technical skills and laparoscopic knowledge using a multiple-choice test [20]. The FLS program utilizes low-fidelity bench models, and skills training is therefore limited to teaching basic skills component and not procedural skills [21]. Currently the cognitive test is not optimal for residents in gynecology [22,23].

The purpose of this study was to develop a structured step-wise curriculum consisting of validated modules in basic laparoscopy for first-year residents in obstetrics and gynecology with educational aims that mirrored actual practice. The core competencies revolved around basic laparoscopy, integrating a knowledge component, a procedural simulation skills component (training a salpingectomy on a virtual reality simulator), and implementation in the clinical setting (performing a salpingectomy on a patient).

Material and methods

Curriculum development was inspired by the FLS program and the curriculum literature [10,13,14,20,24-26] by using a step-wise approach where each step focused on different areas, such as practical training, theoretical education, and assessment of clinical work, and it also adhered to the current principles of proficiency-based training, distributed and deliberate practice [8,27-29]. Additionally, the curriculum acknowledged the psychomotor learning presented by Fitts and Posner[30] and Millers competence assessment theory [31].

Context of the study

Danish specialty training in obstetrics and gynecology is five years and consists of work-based assessment, but does not include postgraduate examinations. The present curriculum was designed for use in a regional setting in order to improve and standardize a basic level for laparoscopic knowledge and technical skills. All gynecological departments in the region were involved in the study. One project investigator (the first author; JS) managed all communication and logistic issues with participants and departments, and was located at the hospital of study origin; Rigshospitalet, Copenhagen University Hospital.

Participants

First-year residents in obstetrics and gynecology in the Capital and Zealand region from June 2009 till November 2011 were invited to participate.

Inclusion criteria: 1) informed consent. Exclusion criteria: 1) not completing the first three steps of the curriculum within a six-month period, 2) performing an actual laparoscopic salpingectomy before completion of step 1, 2 and 3. Previous experience with laparoscopy was registered, but was not an exclusion criterion.

The virtual reality simulator and training modules

The LapSim® from Surgical Science, Göteborg, Sweden. Pre-programmed settings were used for the basic skills modules, which included: camera navigation, instrument coordination, grasping, cutting, clip applying, and lifting and grasping. As examination module the *upgraded salpingectomy* (a salpingectomy due to an ectopic pregnancy) were chosen. A clinically validated proficiency level was used, table 1 [32].

The LapSim software stored all performance parameters e.g. time and repetitions to complete a task, instrument path length during task, and the number of errors. The simulator provided performance feedback on the individual parameters to the participants immediately after each repetition.

Study design

This was an observational study involving four steps, which are described separately in following text. The aims of the study were explained to the participants and informed written consent was obtained. Participants were free to withdraw at any stage.

Step 1: A 1-day course in basic laparoscopy.
A verbal needs analysis among ten regional chief physicians in charge of education focusing on future courses for residents was conducted in 2009. It concluded that a need for a course in basic laparoscopy existed and subsequently three chief physicians, two fellows, and the first author developed a course program. It contained six theoretical topics: laparoscopic anatomy, laparoscopic instrument handling, entry and insufflation techniques, and principles of electro-surgery. All theoretic topics were presented with 30-minutes lectures. Furthermore, the course featured hands-on training on virtual reality simulators and bench models. Course instructors included three chief physicians and two staff specialists. The regional panel of chief physicians responsible for resident education and head of departments in obstetrics and gynecology accepted the content of the course and participants were given protected time to participate.

Step 2: A multiple-choice test in basic laparoscopy.
For the purpose of this curriculum a multiple-choice test containing 37 questions revolving around basic principles in laparoscopy including the procedure for a laparoscopic salpingectomy was developed. Development and validation of the multiple choice test is described in a separate publication [33].
Participants were tested a total of three times with the same test; right before and after the course *Basic Laparoscopy* (Step 1), and again within six months. After the second test, participants were informed of individual test score and incorrectly answered questions. This was done to encourage participants to self-study.
The author group decided on a passing level at 85%, which was based on the average score from chief physicians validating the test.
No remedies, e.g. Internet access or books, were allowed while answering the test and the project investigator monitored the testing. There was no time constraint for answering. If failing the test, the participant was allowed another attempt after a period of minimum two weeks.

Step 3: Structured training on a virtual reality simulator
Participants had to complete all basic task modules before training the examination module; *upgraded salpingectomy*, Table 1. Participants could schedule training time on the simulator by contacting the project investigator, and each training session was limited to three hour. The project investigator provided standardized feed-back regarding instrument handling, use of electro-surgery and operational technique on request from participants.

Step 4: An operation (a salpingectomy)
The last step in the four-step curriculum was to perform a real operation; a laparoscopic salpingectomy. The operation was performed under supervision at the departments where the participants were employed. The operation could be a salpingectomy due to an ectopic pregnancy or part of a larger operation where salpinx was laparoscopically removed. The operation was subsequently assessed using the previously validated assessment scale Objective Structured Assessment of Laparoscopic Salpingectomy (OSA-LS) [34]. The supervising surgeon or a surgeon reviewing the recorded operation performed the assessment. There was no passing level for the operation; the assessment consisted of formative feedback. The fourth step was only allowed if the previous three steps were completed.

Questionnaire
A questionnaire was distributed after the 1-day course concerning satisfaction with all aspects of the course, including being tested with a multiple-choice test. Furthermore, an

anonymized post-study questionnaire concerning the curriculum and participant perceptions was e-mailed to all participants. The questionnaires were scored on a Likert-type scale from one to five, where one represented 'strongly disagree' and five represented 'strongly agree'.

Ethics

Participation was voluntary. The operation in step 4 was performed at seven different educational hospitals, where a laparoscopic salpingectomy potentially can be a part of residents training; therefore protocol did not differ from normal standards. No patient related information was displayed during the recording of the operation.

Due to the fact that participation was voluntary and that patients involved received standard care, no ethical approval was needed according to the Danish national ethics committees.

Statistics

Data was processed using SPSS 19.0 for Windows. Non-parametric tests were used because distributions of the obtained answers did not justify the general use of parametric tests. Wilcoxon test was used to compare the different test scores from participants. Mann-Whitney U was used to compare results from virtual reality simulation training. A significance level of < 0.05 was chosen.

Statistics for development of the multiple-choice test is described separately[33].

Results

Participants

A total of 52 first-year residents employed at seven obstetric and gynecologic departments were invited to participate in the study and all accepted the invitation, Figure 1. Average time in the obstetric and gynecologic specialty was four months. One participant had previous experience with laparoscopy from general surgery.

Step 1: A 1-day course in basic laparoscopy.

A pilot course was held before initiation of the project, and subsequently the course was held twice a year. In the study period three courses were conducted and there was a 100% attendance rate, Figure 1. The questionnaire revealed an overall satisfaction with the course with a mean score of 4.8 out of 5.

Step 2: A multiple-choice test in basic laparoscopy

Participants were all tested before and after the course, and 38 completed the third test within six months, Figure 2. Significant differences were found between test scores at pre-course and post-course, $p=0.001$, and between post-course and third test, $p=0.001$, Figure 2. All Participants who completed the third test passed. There was no difference in test scores between participants who went through all four steps and participants who dropped out after step 3, $p=0.876$. On average participants spent 30 minutes on each test. Fourteen participants dropped out after step 1. Reasons for dropping out are listed in Figure 1.

Step 3: Structured training on a virtual reality simulator

Thirty-eight participants completed the structured simulation training. Time used on basic skills modules and the examination module *upgraded salpingectomy* is shown in Figure 3, the mean time was 238 minutes. The average number of repetitions to reach proficiency level was 38, with a range of 91 (minimum 8, maximum 99). There was no difference in time and repetitions between participants who went through all four steps and participants who dropped out after step 3, $p=0.975$ and $p=0.671$, respectively. Dropouts were the same participants that dropped out in step 2.

Step 4: The operation (salpingectomy)

Twenty-six participants completed the final step. All participants performed a salpingectomy as a part of a larger operation; none performed a salpingectomy due to an ectopic pregnancy. Fourteen participants had the operation assessed immediately by the supervising surgeon and twelve by using the recorded operation. The completion rate was 55% (26/47 - not including dropouts due to change of specialty) of the participants. Participants going on maternity leave were considered as dropouts due to the exclusion criteria.

Questionnaire answers
The post-study questionnaire had a 96% (50/52) response rate. Out of the 38 who completed step 3 and 4, 14 reported they had some protected training time, in average five hours. Participants who had protected training had a significantly higher completion rate, p=0.046. Relevant responses are displayed in Figure 4.

Discussion

We have developed and validated each step in a comprehensive four-step curriculum in basic laparoscopy for first-year residents in obstetrics and gynecology. In contrast with most laparoscopic skills curricula, which mainly focus on basic simulation training and knowledge [20,35-37], we have extended this step-wise curriculum to also include a procedural component allowing participants to perform a laparoscopic salpingectomy after having acquired the necessary skills and knowledge outside the operating room.

Step 1: The 1-day course Basic Laparoscopy.
Intensive short courses can improve both knowledge and skills in medicine [38], and laparoscopic 'hands on' courses have also demonstrated to improve short term knowledge and motor skills [39]. The course content and set up for the 1-day course was approved regionally, which indicates acceptable content validity. Furthermore, the course revealed good face validity due to high ratings on an anonymized questionnaire. All participants significantly increased their multiple-choice test scores from pre-course to post-course, which we interpret as validation of the cognitive part of the course.

Step 2: A multiple-choice test in basic laparoscopy
A cognitive component is an essential part of a curriculum [14,20,40], and it is important to ensure a basic understanding of surgical principles and equipment. Tests and assessment are powerful motivators for learning and can encourage residents to study on their own and participate in available educational opportunities [41,42]. We wanted to make use of this potential motivation to self-study by testing the participants again after returning to their respective departments, and all participants actually increased their test scores the third time. However, we tested the participants several times with the same test, which could be a bias; the participants might improve their performance simply because they recognize the test. Furthermore, being tested three times could have induced some test-enhanced learning, which could have increased test scores. This last-mentioned phenomenon, test-enhanced learning, is not, in our opinion, unwarranted since it has shown to increase knowledge [43,44,44,45].
It would have been optimal to test the MCQ test on a resident control group who did not go through the four-step curriculum to compare progress.
Palter and Grantcharov recently conducted a randomized study assigning the intervention group to a comprehensive curriculum in laparoscopic right hemicolectomy, involving virtual reality training and knowledge testing, and the control group to traditional training. In addition to finding improved technical performance in the actual operation room, the study demonstrated that the curriculum trained group improved cognitive knowledge related to performing the advanced surgical procedure [46].

Step 3: Structured training on a virtual reality simulator

Reaching proficiency level on the operation module *upgraded salpingectomy* improves performance in the operation room [5]. This is substantiated by other studies involving virtual reality simulators [6-8], although some of these studies only tested basic skills modules such as coordination and navigation. We found major individual variability in training time and repetitions to reach the proficiency level, which clearly supports proficiency-based training in contrast to time-based training. The major advantages of proficiency based training over a fixed number of repetitions includes self-paced practice and goal-directed learning, which can lead to a maximal training benefit [35,47]. Additionally, proficiency-based training can be tailored to the individual participant [8,10,27], and there is a consistency of the final result, because all residents are expected to reach the same performance standard.

There is still no evidence on the predictive validity of simulation training, i.e. whether some residents should be advised to change specialty due to poor performance on a simulator. This is a controversial topic and further research is required before the use of simulators for this purpose can be considered [4].

Standardized instructor feedback was provided upon request and was therefore not distributed evenly to participants, which potentially could have contributed to the scattered performance. A randomized study by our research group recently found that instructor feedback reduced training time by half when training operational modules on a virtual reality simulator [48].

Step 4: The operation (salpingectomy)

The purpose with the fourth step in the curriculum was to link the skills and knowledge obtained in the preceding three steps to daily clinical work. It is highly relevant that learners engage in clinical activity after simulation training and receive assessment on performance [49]. The OSATS-based assessment scale OSA-LS is thoroughly validated as an assessment tool [34], and we have previously demonstrated that doctors with different educational levels can use the OSA-LS[50]. Nonetheless, there might be some variability in assessment score depending on the assessors and depending on whether the assessment was performed 'live' i.e., right after the operation or by video assessment. Consequently, we did not choose a cut-off value for the assessment of the operation. Additionally, current assessment literature does not support a cut-off value [51]. Furthermore, the current four-step curriculum was designed as a skills training intervention, not a screening examination.

The participating departments had agreed to provide the participants with an operation after completion of the first three steps; however, several participants were not offered this operation, mainly due to logistic challenges with the operating schedule. Whether this reflects lack of interest to participate, or missing information about the curriculum on all levels of the organization is unknown. Conducting a multi-center study can be difficult, and a careful analysis of the actual logistical conditions is essential when designing and implementing a curriculum.

Participants who *did* complete the operation step reported in the questionnaire that the 'door had opened to more surgery', thereby allowing them to learn laparoscopic surgery that they otherwise would not have performed at their educational level. Furthermore participants mentioned that the reward, i.e. getting the laparoscopic operation, was motivation for participating.

Assessment of the structured four-step curriculum

Several participants dropped out of the study, the majority due to lacking motivation. In an anonymized questionnaire the dropouts informed that the missing support from their departments, e.g., protected training time, was the main reason. In a similar study involving skills training, the

attendance rate leaped from 6% to 71% when time was dedicated specifically to skills training and supervision staff were hired [9]. In the current study completion of all four steps was correlated with pre-determined training hours. This is consistent with other studies that also finds protected time an effective strategy [35,52], and efforts must be made to secure protected laparoscopic training time.
Motivation to participate was addressed in the questionnaire but not investigated in a structured way. Kusurkar et al. reported that existing research suggests the learning environment plays an important role in enhancing motivation[53], and a future study should encompass focus on both intrinsic motivation, such as personal improvement and interest, and extrinsic motivation, such as assessment and feedback.
Several participants did not complete the four-step curriculum due to maternity leave. This will continue to be a challenge in curriculum planning [54] underlining the importance of a flexible curriculum.

Miller has described a framework of clinical competence assessment with increasingly complex levels of skills performance [31]. The pyramid consists of four levels: *knows, knows how, shows how* and *does*. The four-step curriculum is in accordance with Miller's framework, Figure 5. Furthermore, the four-step curriculum supports Fitts and Posner's established theory of motor-skills acquisition involving threes-stages to learning a new skill; *a cognitive phase, a associative phase*, and *a autonomous phase* [30], figure 5. Involving such frameworks in curriculum design provides a structured layout.
The important question is whether the desired goal has the intended outcome, i.e., did the participants actually gain skills and knowledge. It is not an easily answered question, nevertheless, the literature supports the hypothesis that practice is an important determinant of outcome [46,55,56]. The most convincing way to scientifically justify a curriculum's existence would be to set up a randomized trial with an intervention group and a control group. However, in the present study we decided it was ethically unacceptable to conduct a randomized trial with patients as training models (traditional training), due to the mounting evidence in favor of simulation training. We acknowledge that this study does not have the advantage of comparing results to a control group.

We propose a four-step curriculum in basic laparoscopy containing a course component, a knowledge component, a virtual reality training component aided by instructor feedback, and an operational component with subsequent formative assessment. If the laparoscopic procedure, the fourth step, is not completed within six months initial simulation training, training must be repeated [57]. Surgical curriculum development is a top priority for the 21st century [58], and we believe this blueprint for a step-wise curriculum is applicable in other specialties.

Future perspectives
The four-step curriculum is now a part of regional residency training and continuous efforts will be made to make it obligatory.

Acknowledgements

We thank Karl Bang Christensen, PhD in biostatistics, associate professor, Department of Public Health, Unit of Biostatistics, University of Copenhagen, Copenhagen, Denmark for statistical help.

Disclosures: None declared. All authors have signed the ICMJE Form for Disclosure of Potential Conflicts of Interests.

References

1. Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. *Acta Obstet Gynecol Scand.* 2012;91:1015–1028.

2. Reznick RK, Macrae H. Teaching surgical skills--changes in the wind. *N. Engl. J. Med.* 2006;355:2664–2669.

3. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev.* 2009:CD006575.

4. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World J Surg.* 2008;32:182–188.

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

6. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg.* 2004;91:146–150.

7. 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;236:458–63.

8. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *Am. J. Surg.* 2007;193:797–804.

9. 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;65:4–7.

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

11. Korndorffer JR, Stefanidis D, Scott DJ. Laparoscopic skills laboratories: current assessment and a call for resident training standards. *Am. J. Surg.* 2006;191:17–22.

12. Dent J, Harden RM. *A Practical Guide for Medical Teachers.* 3rd ed. Churchill Livingstone; Edinburgh, 2009.

13. Harden RM. Ten questions to ask when planning a course or curriculum. *Med Educ.* 1986;20(4):356–365.

14. Grantcharov TP, Reznick RK. Teaching procedural skills. *BMJ.* 2008;336:1129–1131.

15. Palter VN, Orzech N, Reznick RK, Grantcharov TP. Validation of a Structured Training and Assessment Curriculum for Technical Skill Acquisition in Minimally Invasive Surgery: A Randomized Controlled Trial. *Ann. Surg.* 2012, Sep 25.

16. Rosenthal ME, Ritter EM, Goova MT, et al. Proficiency-based Fundamentals of Laparoscopic Surgery skills training results in durable performance improvement and a uniform certification pass rate. *Surg Endosc.* 2010;24:2453–2457.

17. Scott DJ, Ritter EM, Tesfay ST, Pimentel EA, Nagji A, Fried GM. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. *Surg Endosc.* 2008;22:1887–1893.

18. Swanstrom LL, Fried GM, Hoffman KI, Soper NJ. Beta test results of a new system assessing competence in laparoscopic surgery. *J. Am. Coll. Surg.* 2006;202:62–69.

19. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. *Am. J. Surg.* 2010;199:115–120.

20. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery.* 2004;135:21–27.

21. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov.* 2007;14:107–112.

22. Zheng B, Hur H-C, Johnson S, Swanstrom LL. Validity of using Fundamentals of Laparoscopic Surgery (FLS) program to assess laparoscopic competence for gynecologists. *Surg Endosc.* 2010;24:152–160.

23. Hur H-C, Arden D, Dodge LE, Zheng B, Ricciotti HA. Fundamentals of laparoscopic surgery: a surgical skills assessment tool in gynecology. *JSLS.* 2011;15:21–26.

24. Kern DE, Thomas PA, Hughes MT eds. *Curriculum Development for Medical Education: A Six-Step Approach*, second edition. The Johns Hopkins University Press; Baltimore, 2009.

25. Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BLH, Scherpbier AJJA. Validation and implementation of surgical simulators: a critical review of present, past, and future. *Surgical endoscopy.* 2010;24:536–546.

26. Morozov V, Nezhat C. Proposal of a formal gynecologic endoscopy curriculum. *J Minim Invasive Gynecol.* 2009;16:416–421.

27. Palter VN, Graafland M, Schijven MP, Grantcharov TP. Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: A Delphi approach. *Surgery.* 2012;151:391-7

28. Moulton C-AE, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect? a randomized, controlled trial. *Ann. Surg.* 2006;244:400–409.

29. Ericsson K, Krampe R, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. *Psychological review.* 1993:363–406.

30. Fitts PM & Posner M. Human performance, Belmont, CA, Brooks/Cole, 1967

31. Miller G. The assessment of clinical skills/competence/performance. *Acad Med.* 1990:63–67.

32. Larsen CR, Grantcharov T, Aggarwal R, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. *Surg Endosc.* 2006;20:1460–1466.

33. Oestergaard J, Maagaard M, Larsen CR, et al. Development and validation of a theoretical test in basic laparoscopy. *Surg Endosc*, DOI: 10.1007/s00464-012-2615-7

34. 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;115:908–916.

35. Panait L, Bell RL, Roberts KE, Duffy AJ. Designing and validating a customized virtual reality-based laparoscopic skills curriculum. *J Surg Educ.* 2008;65:413–417.

36. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann. Surg.* 2006;244:310–314.

37. van Dongen KW, Ahlberg G, Bonavina L, et al. European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. Surg Endosc. 2011 Jan;25:166-71

38. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, 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;325:1338–1341.

39. Condous G, Alhamdan D, Bignardi T, et al. The value of laparoscopic skills courses. *Aust N Z J Obstet Gynaecol.* 2009;49:312–315.

40. Fried GM. Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. *Gastrointest. Endosc. Clin. N. Am.* 2006;16:425–434.

41. Case SM. *Constructing written test questions for the basic and clinical sciences.* 2nd ed. National Board of Medical Examiners; Baltimore, 1998.

42. van Empel PJ, Verdam MGE, Strypet M, et al. Voluntary autonomous simulator based training in minimally invasive surgery, residents' compliance and reflection. *J Surg Educ.* 2012;69:564–570.

43. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ.* 2009;43:21–27.

44. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. *Med Educ.* 2008;42:959–966.

45. Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci.* 2006;17:249–255.

46. Palter VN, Grantcharov TP. Development and Validation of a Comprehensive Curriculum to Teach an Advanced Minimally Invasive Procedure: A Randomized Controlled Trial. *Ann. Surg.* 2012;256:25-32.

47. Korndorffer JR, Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J. Am. Coll. Surg.* 2005;201:23–29.

48. Oestergaard J, Bjerrum F, Maagaard M, et al. Instructor feedback versus no instructor feedback on performance on a virtual reality simulator: a randomized trial. *Ann. Surg.*, accepted September 2012.

49. Kneebone R, Bello F. Surgical Simulation. In: Riley RH, ed. *A Manual of Simulation in Healthcare.* Oxford: Oxford University Press, USA; 2008:435–448.

50. Oestergaard J, Larsen CR, maagaard M, Grantcharov T, Ottesen B, Sorensen JL. Can both residents and chief physicians assess surgical skills? *Surg Endosc.* 2012;26:2054-60

51. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg.* 2010;97:972–987.

52. Haluck RS, Satava RM, Fried G, et al. Establishing a simulation center for surgical skills: what to do and how to do it. *Surg Endosc.* 2007;21:1223–1232.

53. Kusurkar RA, Cate Ten TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: a review of the literature. *Med Teach.* 2011;33:e242–62.

54. Maruscak AA, Vanderbeek L, Ott MC, Kelly S, Forbes TL. Implications of current resident work-hour guidelines on the future practice of surgery in Canada. *J Surg Educ.* 2012;69:487–492.

55. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann. Intern. Med.* 2002;137:511–520.

56. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med.* 2008;15:988–994.

57. Maagaard M, Sorensen JL, Oestergaard J, et al. Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. *Surg Endosc.* 2011;25(3):722–727.

58. Stefanidis D, Arora S, Parrack DM, et al. Research priorities in surgical simulation for the 21st century. *Am. J. Surg.* 2012;203:49–53

Table 1
Based on 11 variables listed below the virtual reality simulator generated a performance score (%) on the examination module *upgraded salpingectomy*, which were available for all participants after each repetition. When all variables were within the passing range predefined proficiency was reached. The predefined proficiency level was set and validated in a previous study [32]

Variable	Passing range	Weight in calculating performance score
Total time	>280 (s)	15
Blood loss	>180 (ml)	15
Pool volume	>10 (ml)	0
Ovary diathermy damage	>3 (s)	5
Tube cut: uterus distance	>4 (mm)	5
Bleeding vessel cut	0	Fail if performed
Evacuation from body	>1	Fail if not performed
Left instrument path length	>2 (m)	15
Left instrument angular path	>350 (degrees)	15
Right instrument path length	>3 (m)	15
Right instrument angular path	>450 (degrees)	15

Figure 1
Flow chart for enrollment in the four-step curriculum in basic laparoscopy.

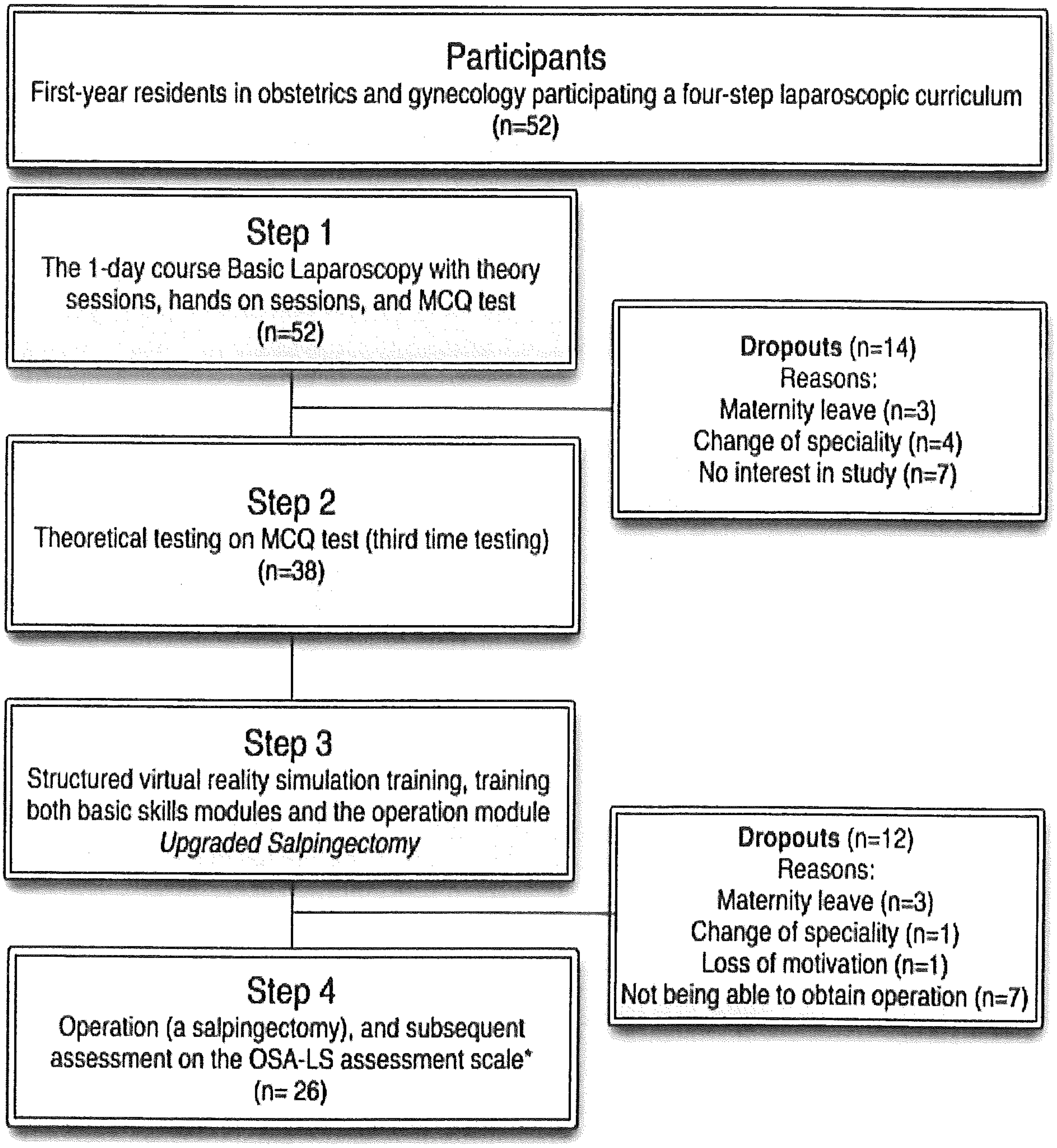


Figure 2

Box plot showing multiple-choice test scores at three different times. Significant differences were found between pre-course and post-course scores, p=0.001, and between post-course score and the third test, p=0.001. Minimum and maximum scores (whiskers) and medians are presented. The boxes represent 25th and 75th percentiles.

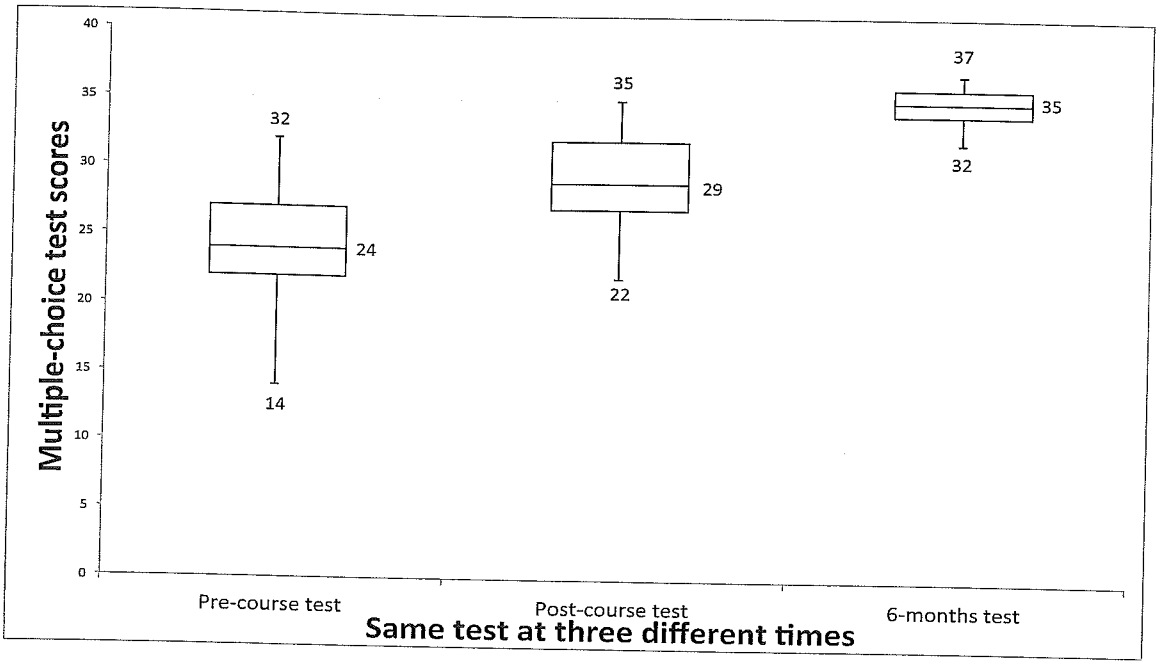


Figure 3

Time spent training basic skills modules and the procedural examination module *upgraded salpingectomy*

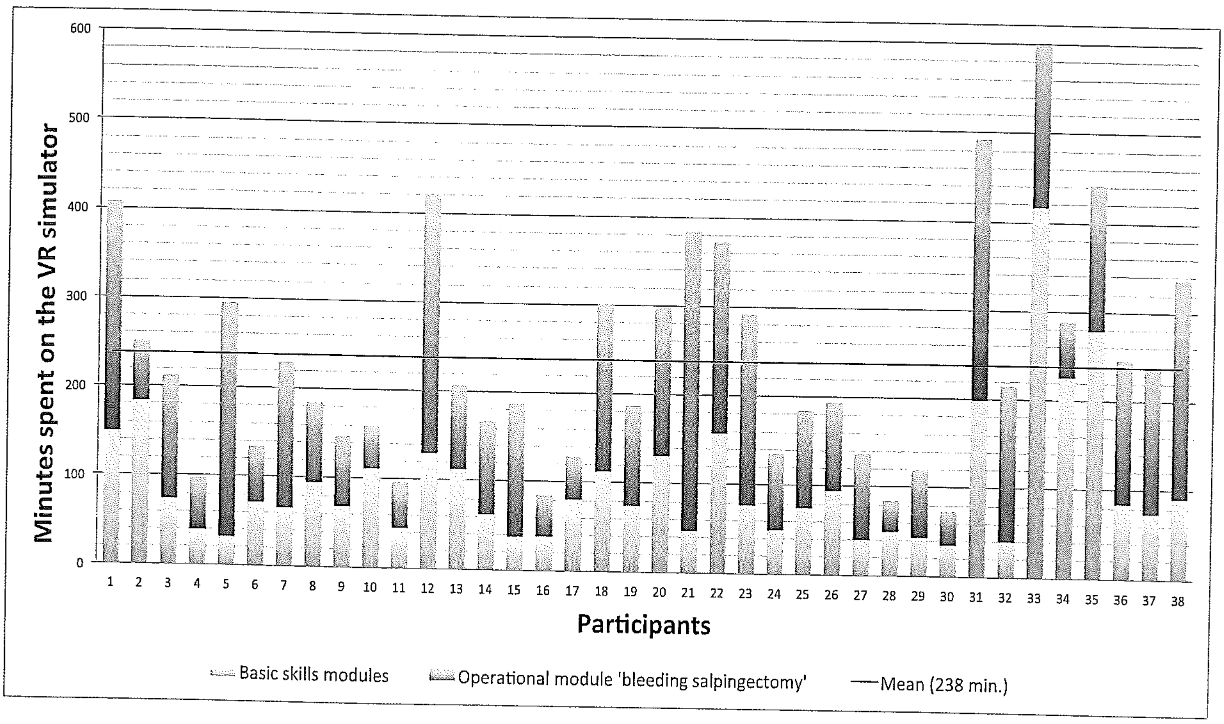


Figure 4
Answers from the post-study questionnaire.

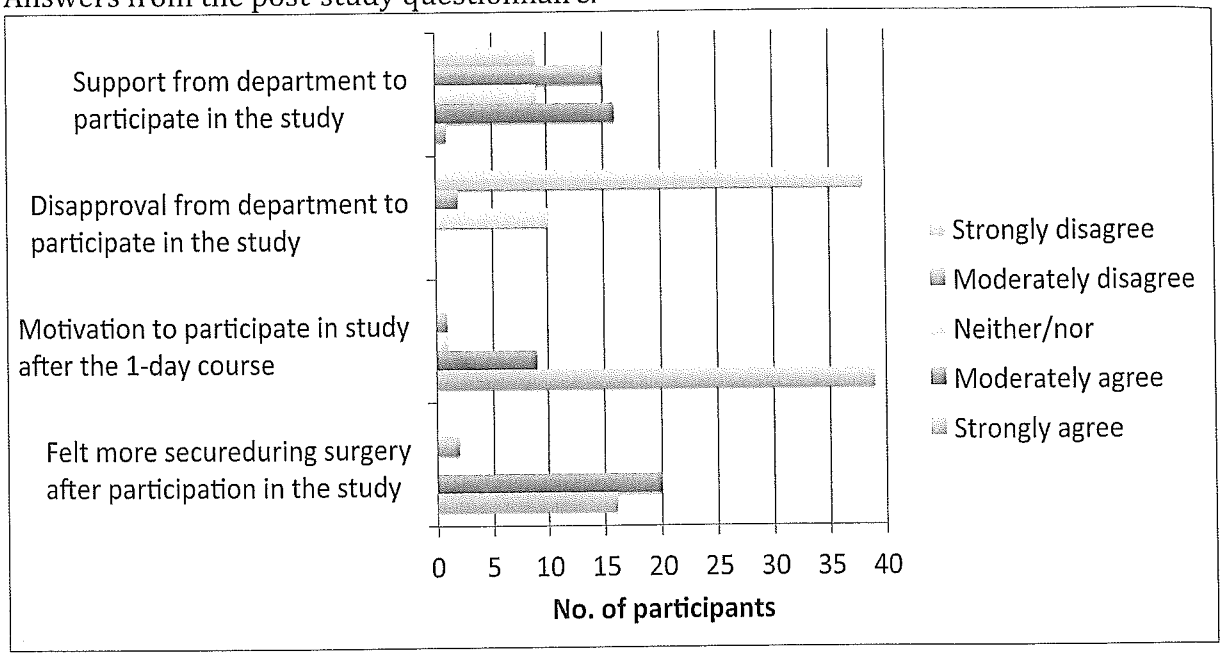
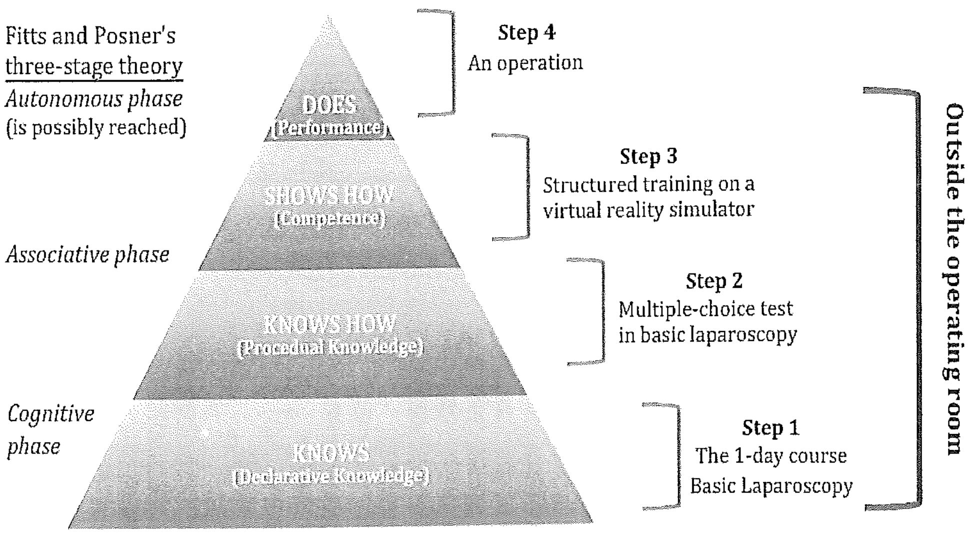


Figure 5
The structured four-step curriculum in relation to Miller's framework of competence and assessment in education. Furthermore, Fitts and Posners three-stage theory on motor skills learning is applied.



Discussion

Outline

On the basis of the current curriculum literature, we developed a four-step curriculum in basic laparoscopy with clear educational objectives that mirrored actual practice. This four-step curriculum was developed in Study IV, and Studies I, II, and III were developed jointly to substantiate and validate each step. Figure 4 shows the different steps of the curriculum and the interaction of the different studies. Each step of the four-step curriculum is discussed separately.

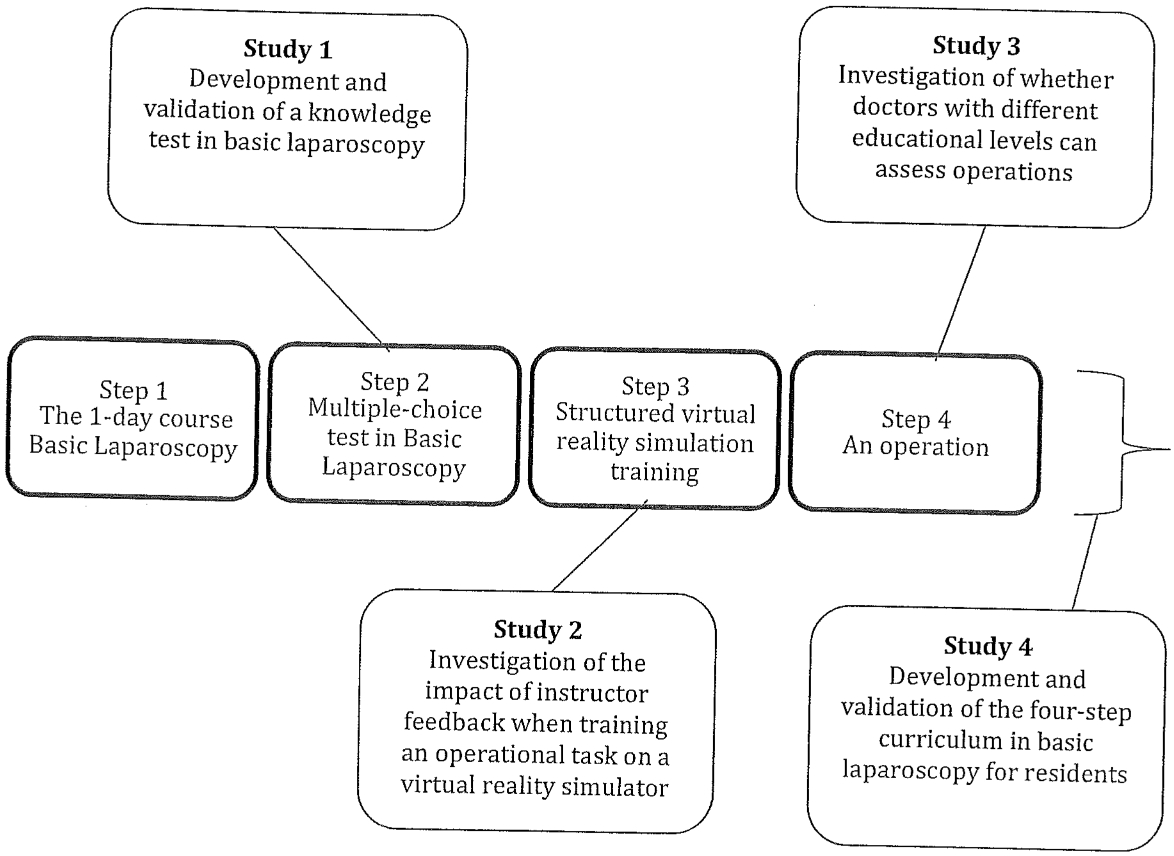


Figure 4. The four studies constituting this thesis linked together to validate the four steps in the four-step curriculum.

The first part of the Discussion chapter is a brief summary of each study. The second part discusses each step of the four-step curriculum individually in relation to Studies I, II, III and IV and existing literature. The third part reviews some limitations related to the studies, and the final part of the chapter discusses perspectives and future studies.

Summary of the four studies

Study I: *'Development and validation of a theoretical test in basic laparoscopy'.*

Through interviews with four experts in laparoscopy and subsequently through a Delphi audit involving regional laparoscopic surgeons, a test with 37 multiple-choice questions in basic laparoscopy was developed. The test could discriminate between medical students and doctors at different educational levels, i.e., it showed good construct validity; furthermore, there was no evidence of differential item functioning.

Study II: *'Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial'.*

This randomized trial with 99 participants demonstrated that instructor feedback significantly increased efficiency when training a complex operational task on a virtual reality simulator. The intervention group, who received standardized instructor feedback for 20–30 minutes, used half the amount of time and number of repetitions to reach a predefined proficiency level compared with the control group. Three participants in the control group dropped out due to the frustration of not being able to complete the task. The control group reached the predefined proficiency level with a higher performance score than the intervention group.

Study III: *'Can both residents and chief physicians assess surgical skills?'*

Through a cohort study, we demonstrated that senior residents and chief physicians in gynecology were equally able to assess laparoscopic operations on the basis of a previously validated assessment scale: Objective Structured Assessment of Laparoscopic Salpingectomy (OSA-LS). Both groups revealed no difficulties when utilizing the assessment scale.

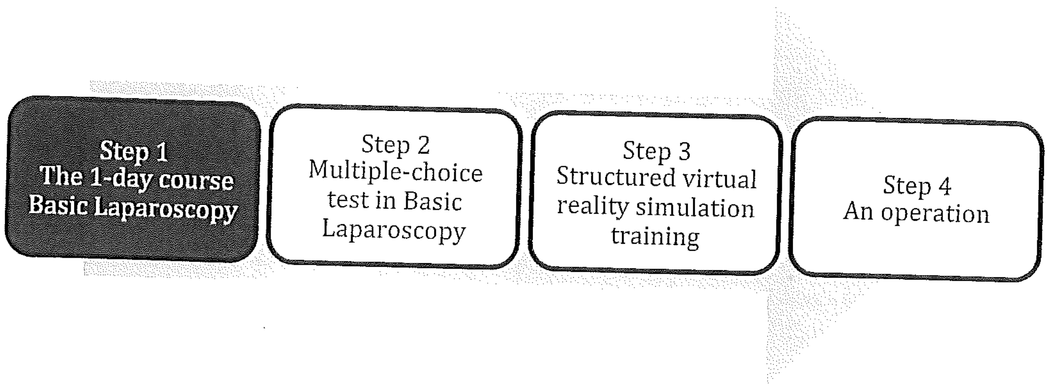
Study IV *'A four-step curriculum in basic laparoscopy: development and validation. A curriculum blueprint'*

A curriculum containing four steps was developed: Step 1) 1-day course in basic laparoscopy, Step 2) A multiple-choice test in basic laparoscopy, Step 3) Structured virtual reality simulation training, and Step 4) Operation on a patient with subsequent formative assessment. A cohort of 52 first-year residents signed up for the study. There was 100% attendance during the first step, and 55% completed all four steps. There were several reasons for dropping out; the main reason was the voluntary nature of the study. Additionally, some logistical problems were identified for Step 4.

Assessment of the four-step curriculum in basic laparoscopy

When focusing on assessment the two key concepts *validity* and *reliability* are essential elements. Traditionally, validity is divided into several facets, such as face validity, content validity, construct validity, and predictive validity. Validity is based on various aspects of clinical competence and cannot be expressed as a simple coefficient [65,66]. It all adds up to the question 'does the assessment instrument produce the desired educational outcome?' In the four-step curriculum in basic laparoscopy, each of the four steps was researched regarding validity and is described separately.
Reliability refers to the consistency of a measure; a measure is said to have a high reliability if it produces consistent results under consistent conditions [35].

Step 1. The 1-day course Basic Laparoscopy

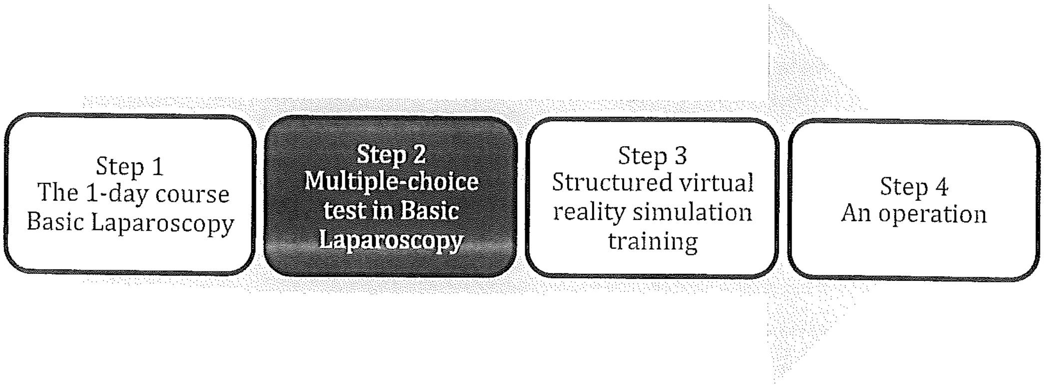


The 1-day course, Basic Laparoscopy, was developed on the basis of a regional needs analysis. The aim of the needs analysis was to design a course that structured knowledge of basic laparoscopy among first-year residents in obstetrics and gynecology. It has been demonstrated that attending short, intensive courses can improve both knowledge and skills in evidence-based medicine [67], and laparoscopic 'hands on' courses have also shown to improve short-term knowledge and motor skills [68]. The course content and set-up for the 1-day course was approved regionally, which is an indication of acceptable content validity.

In Kirkpatrick's four levels of evaluation, the *reaction evaluation* shows how the participants feel after the training and learning experience. In an anonymized questionnaire the participants gave the course a high rating, which displays good face validity—and good *reaction evaluation*. The *learning evaluation*, which is the measurement of the increase in knowledge from before to after was satisfactory; all participants significantly increased their test scores from pre-course to post-course. Three courses were conducted during the study and the good evaluation of all the courses is an indication of reliability.

Kirkpatrick mentions three reasons for evaluating a course. The first is to justify the existence and budget by showing how it contributes to the organizational objectives and goals; the second is to evaluate whether the current course should continue; and the third is to retrieve information on how to improve future training sessions [27]. All three reasons are met in the 1-day course, Basic Laparoscopy, since it fulfills demands on the organizational level as well as on the individual level.

Step 2. The multiple-choice test in basic laparoscopy



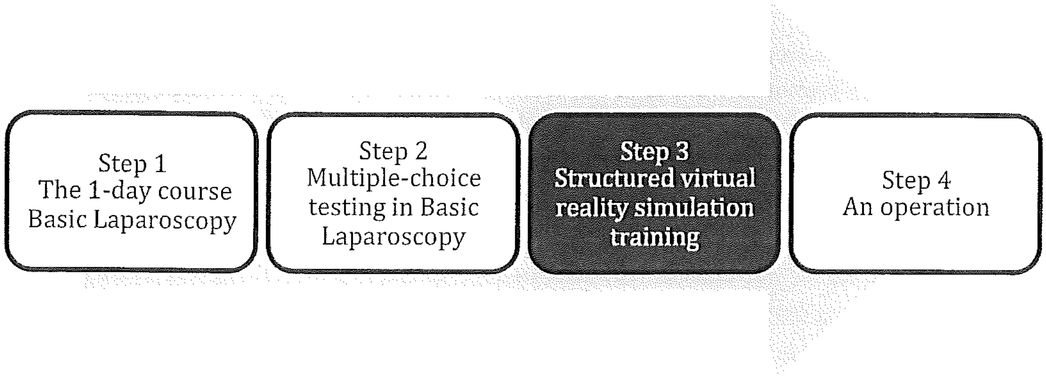
A cognitive component is an important part of any curriculum and is expected to provide basic understanding of the surgical principles and knowledge in order to handle unexpected events [7,46]. A supporting argument for this was demonstrated in a randomized study finding that cognitive skills training enhanced ability to correctly execute a surgical task [69]. Nonetheless, structured testing of basic procedure-relevant knowledge in the surgical domains is not standard practice.
Resident training may vary from hospital to hospital, and in order to facilitate homogeneous learning—regional or national—a unified test can facilitate knowledge and aid accomplishment of key curricular goals [70].

The knowledge test developed in Study I was based on inputs from laparoscopic experts and subsequently validated by a consensus panel consisting of chief physicians. It is therefore reasonable to state that the content sampled is within the domain of interest and thereby content valid. We cannot, however, rule out that there was a selection bias, i.e., different experts would have produced different results.
The test proved to be construct valid due to its ability to distinguish between different educational levels. Furthermore, good face validity was met; the residents rated testing high on a questionnaire, commenting positively on the challenge of test taking.

According to Kirkpatrick's evaluation model, *behavior evaluation* is the extent of applied learning back on the job—the implementation. Measurement of behavior change is more difficult to quantify and interpret than *reaction* and *learning* evaluation is. Evaluation of implementation and application is an extremely important assessment; there is little point in a good reaction and good increase in capability if nothing changes back on the job. Evaluation is therefore vital, albeit challenging. The significant increase in the test scores after the third test (after the residents had returned to their departments) could be viewed as positive behavior evaluation or implementation of knowledge. However, we tested the participants several times with the same test, which could be a bias: participants might improve their performance simply because they recognize the test. Further, being tested three times could induce some test-enhanced learning [71,72], which then increases test scores. This last-mentioned phenomenon, test-enhanced learning, is not, in our opinion, unwarranted since it has shown to increase knowledge [71-73].

Tests and assessment are powerful motivators for learning and can encourage residents to study on their own and to participate in available educational opportunities [70]. However, in Denmark there is no postgraduate testing or examinations to pass during specialty training, only in-training and work-based assessment. Due to the findings in Study I, it could be of interest to integrate validated tests in other curricular activities.

Step 3. Structured virtual reality simulation training



The laparoscopic salpingectomy was chosen as an index operation for several reasons. First, laparoscopic salpingectomy is a key operation in gynecology. It is used for the potentially life-threatening condition ectopic pregnancy and should be mastered by all gynecologists and obstetricians working on call duty. Second, the procedure possesses the necessary complexity for the assessment of laparoscopic skills and can be trained on the LapSim simulator, and third, a previous study has shown transfer of skills to the operation room after reaching proficiency level [1]. Basic skills tasks were also practiced since practicing a variety of cases and learning from errors is required for development of complex skills [74,75]. Gaudagnoli's framework indicates that when trainees reach their challenge point, no new learning will take place unless they are challenged with a task of increased difficulty [76]. The virtual reality simulator offers many tasks with different levels of complexity training, which is a continuous challenge for the trainees. It is uncertain whether some of the virtual reality tasks were too difficult, which could have hindered further learning due to cognitive overload.

The large variability in training time and repetitions to reach a proficiency level portrayed in Studies II and IV clearly support proficiency-based training, also referred to as goal-oriented training, in contrast to time-based training. The major advantages of proficiency-based training over a fixed number of repetitions include self-paced practice and goal-directed learning, which can lead to a maximal training benefit [49,77]. Additionally, proficiency-based training can be tailored to the individual participant [3,7,60], and there is consistency in the final result as all residents are expected to reach the performance standard.

It is a controversial topic, without sound evidence, whether simulation accomplishment has predictive validity, i.e. whether some residents should be advised to change specialty due to poor performance on a simulator. Further research is required before the use of simulators for this purpose can be considered [53]. Previous work has portrayed four types of learning curve: individuals who demonstrate proficiency from the beginning, individuals who achieve proficiency between the second and ninth repetition, a third group that demonstrates the slow acquisition of skills but that cannot achieve proficiency by 10 repetitions, and a final group that underperforms from the beginning and shows no tendency towards skills improvement [78]. The study tested only 10 repetitions per task and included a small number of participants, which is currently insufficient to reject individual residents.

Questions have been asked as to whether there is a need for instructor feedback when training on a high-tech virtual reality simulator. The simulator generates detailed objective feedback, but is that sufficient? Within medical education it is well accepted that feedback is an essential component of the teaching and learning process [35,79-81]. Constructive feedback can, among other things, provide the learner with insight into own actions and clarify goals and expectations. When training basis tasks, such as coordination and instrument navigation, no advantages of instructor feedback have been found [82]. However, in Study II we found that instructor feedback significantly increased efficiency regarding time and repetitions to reach proficiency level for a complex virtual operation. This is important when planning a curriculum; instructor feedback provides information about the number of human resources that need to be allocated to skills training.

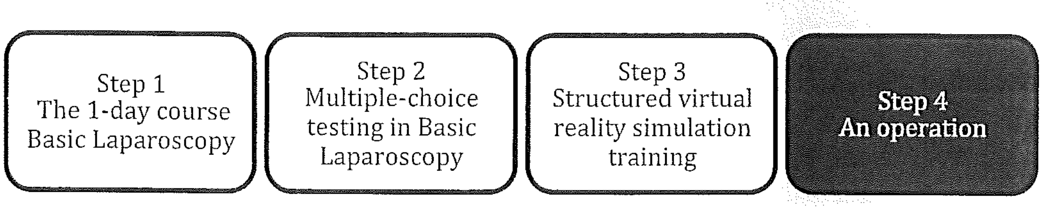
The standardized feedback provided in Study II was more of directional character than interaction between teacher and learner, which feedback normally is. We chose this method because we did not want any contamination of the intervention, which could be the case if feedback was individualized. It is feasible that the benefits of feedback during virtual reality training would be even greater if it went two ways, i.e., between teacher and learner, and could be tailored to the individual trainee.

The self-assessment regarding own surgical skills after participation in Study II revealed that the group who did not receive feedback assessed themselves lower compared with the group who received feedback. Although self-assessment is disputable, individuals who believe they are good at a certain task are more likely to engage in more practice and take on more challenging tasks [83-85]. It would be interesting to follow the participants and observe their future clinical career choices.

The major limitation of Study II is that the sample was senior medical students and hence generalizability of results to first-year trainees to whom a laparoscopic virtual reality curriculum would apply could be a problem. The participating senior medical students had completed all anatomy courses and the mandatory six-month surgical stay placed toward the end of medical school; however, they resembled first-year trainees in several ways because they often have no prior laparoscopic training. In comparison with a previous study by the same author group where first and second year residents performed the same operational virtual reality task (and also received feedback), the average number of repetitions used to reach the predefined proficiency level was almost identical to the average number used by the intervention group in Study II [1]. Whether the senior medical students are a true resemblance to first-year residents is unknown, thus we believe that motivation to learn this kind of complex technical skills was equally high among the students and the first-year postgraduate trainees.

Training laparoscopic skills, and especially complex operational tasks, on a virtual reality simulator allows the trainee to move from the cognitive phase to the associative phase—and perhaps also to the autonomous phase—of the three-stages presented by Fitts and Posner [29]. The learner starts at the cognitive phase trying to intellectualize the task, both for basic skills tasks and operational modules. This step will begin with irregular performance, but with continued improvement and relevant feedback, the learner moves to the next phase: the associative phase. Here the learner progresses to a more regular performance, and knowledge is translated into appropriate behavior. In the last phase, the autonomous stage, the learner does not have to reflect upon hand movements, and in this phase there should be mental capacity to concentrate on other aspects, for instance, anatomical variations and unexpected findings related to the procedure.
It is uncertain whether all participants in this thesis reached the autonomous phase since some might still have been on the ascending part of the learning curve. This is substantiated by the fact that the control group in Study II ended with a better performance score because of the extensive time training on the simulator. Nevertheless, it is possible to reach the autonomous phase in a simulated environment with continuous training.

Step 4. An operation



Simulation training must be complementary rather than supplementary to workplace learning [55], and although the operational virtual reality task salpingectomy has shown to improve operative performance, it is important not to overreach the capabilities of the virtual reality simulator [53]. This is one of the reasons for integrating an operational step linked to formative assessment. However, if the assessment of the operational step was of summative character with a pass or fail, it might be an attempt to measure clinical performance in areas that are beyond those trained in the virtual reality simulator.
This operational step of the four-step curriculum aimed to connect skills and knowledge obtained in the preceding three steps to clinical reality. It is highly relevant that learners engage in clinical activity after leaving the skills laboratory [55].

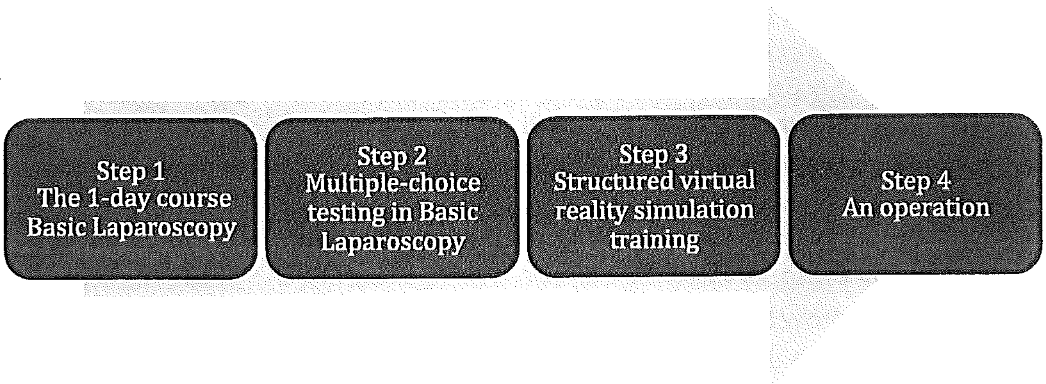
Another reason for integrating an operational step was the idea of a rewards system in which residents were offered the opportunity to perform in an operation after completing proper simulation training. All involved departments agreed to contribute to this reward by allowing the residents to perform an operation under supervision. However, several participants were not offered the operation due to logistical challenges when planning the operation schedule. This is one of the difficulties when conducting a multi-center study, and ongoing commitment from participating departments about the process and progress is essential for successful implementation of a curriculum [35].
The strategy during the four-step curriculum was to reflect social pressure towards collaboration rather than competition in education. None of the participants experienced disapproval from their respective departments about participation; however, only few reported support. Whether this reflects lack of interest in participation or missing information about the curriculum on all levels of the organization is unknown.

Several tools for assessment of technical skills have been investigated and, in particular, the well-established Objective Structured Assessment of Technical Skills (OSATS) [86,87] is currently accepted as the gold standard for objective skills assessment [88]. Larsen et al. developed and validated a laparoscopic procedure-specific assessment scale sequel to the OSATS scale; the Objective Structured Assessment of Laparoscopic Salpingectomy (OSA-LS) [89].
In Study III we investigated whether doctors with different educational levels were able to use the OSA-LS assessment scale with the purpose of integrating it into daily practice. The study demonstrated that senior residents and chief physicians were equally able to assess laparoscopic operations along with high face validity regarding use [90]. The two groups matched the experts in only one assessment out of three. This could be due to several reasons; one is whether our gold standard, i.e., our two expert assessors, is a true gold standard or whether there is a selection bias. The other reason could be the relatively small sample size. However, since the residents and chief physicians represent a variety of hospitals and could still produce similar ratings, we conclude that the assessment scale can be used for formative assessment. Further research is needed before using assessment scales such as the OSATS and OSA-LS for summative assessment [88].

It is important to not only focus on the outcome of one operation; the fundamental and structured element of laparoscopic ratiocination for future learning also needs to be considered.

Reflecting Kirkpatrick's model, the *results evaluation* is the effect on the environment by the trainee, and even though is it plausible to state that the residents have more skills and knowledge after participation in the curriculum, it is still difficult to measure on results in the clinical situation. Additionally, external factors greatly affect organizational performance, which can cloud the true cause of good or poor results. Several residents reported that 'the door opened to more surgery' after participation in the four-step curriculum, which allowed them to perform laparoscopic surgery that they would otherwise not have trained at their educational level. This effect on the environment could be viewed as a result of the curriculum.

All four steps



Due to mounting evidence favoring simulation training, patient considerations, and operating room efficiency, there is no longer any alternative: initial surgical skills training must take place in a simulated environment.
We have proposed a four-step curriculum in basic laparoscopy, which primarily takes place outside the operation room, but also links trained skills to clinical practice.
The key question must always be whether the desired goal produces the intended outcome; did the participants actually gain skills and knowledge from attending the four-step curriculum? It is not an easily answered question since training and practice do not equal learning. Nevertheless, the literature supports the hypothesis that practice is an important determinant of outcome [91,92], and a randomized study has shown improvement of both knowledge and skills after completing a surgical curriculum [52]. Further, it is fair to state that if proficient skills and knowledge in laparoscopy are present, time in the operating room can be allocated to learning more complicated operative techniques [31,93].

Formative assessment can be expressed as 'what ways can training be modified to increase its potential for effectiveness?' and summative assessment as 'was training effective?' [28]. In the four-step curriculum formative assessment is predominant because the assessment occurs in manageable, repeatable and increasing steps. However, summative assessment is also present since there is a set endpoint in terms of a passing level for both the test and the virtual reality training where there is an assessment of effectiveness. To date, the curriculum is not mandatory, but, potentially, it could become a mandatory part of residency training, making it a high stakes summative assessment. However, it is important not to focus only on certification, given that there is a risk that essential learning will be lost. The challenge, therefore, is to connect formative and summative assessment [94].

Several participants dropped out of the curriculum, the majority due to lacking motivation. Whether this was because some had already chosen the obstetrics side of the specialty or that training had to occur in their free time is not entirely known. In an anonymized questionnaire, the dropouts informed that lack of support from their departments, i.e., protected training time, was the main reason. This could be categorized as a system-related influence [31], whereas individual-related influences encounter, e.g., internal motivation. Even though learners must take responsibility for their own skills development, it is important that they have support from their departments as this can affect the encouragement they receive and their motivation [31].
In a similar study the attendance rate leapt from 6% to 71% when time was dedicated specifically to skills training and supervision staff were hired [6]. In the current study, completion of all four steps in the training program was correlated with predetermined training hours, known to both the resident and the faculty. This is consistent with other studies that also find that protected time is the only effective strategy [49,95,96], and conscientious efforts must be made to secure protected laparoscopic training time.

How certification of skills within medical specialties is managed varies between countries. In some countries it is the specialty associations or boards that determine the standard and in others it is the board of health or universities [97]. This can affect decision-making for development and the following implementation. Nevertheless, irrespective of whether a curriculum, and its certification, is regional, national or international, the most important part is the validation, thereby establishing a valid process for assessment and reporting, and ensuring uniformity [98].

Following Miller's theoretical framework, we believe the four-step curriculum can be matched to the four levels of the pyramid, Figure 5. In each step the trainee progresses through the cognitive and practical steps that underlie the following step. Miller's *does* level does not reveal competence, just performance. This additional step is added in Rethan et al.'s work, highlighting the need to assess true clinical performance in addition to true competence at performing a practical skill [99], which is much like the fourth step in the four-step curriculum.

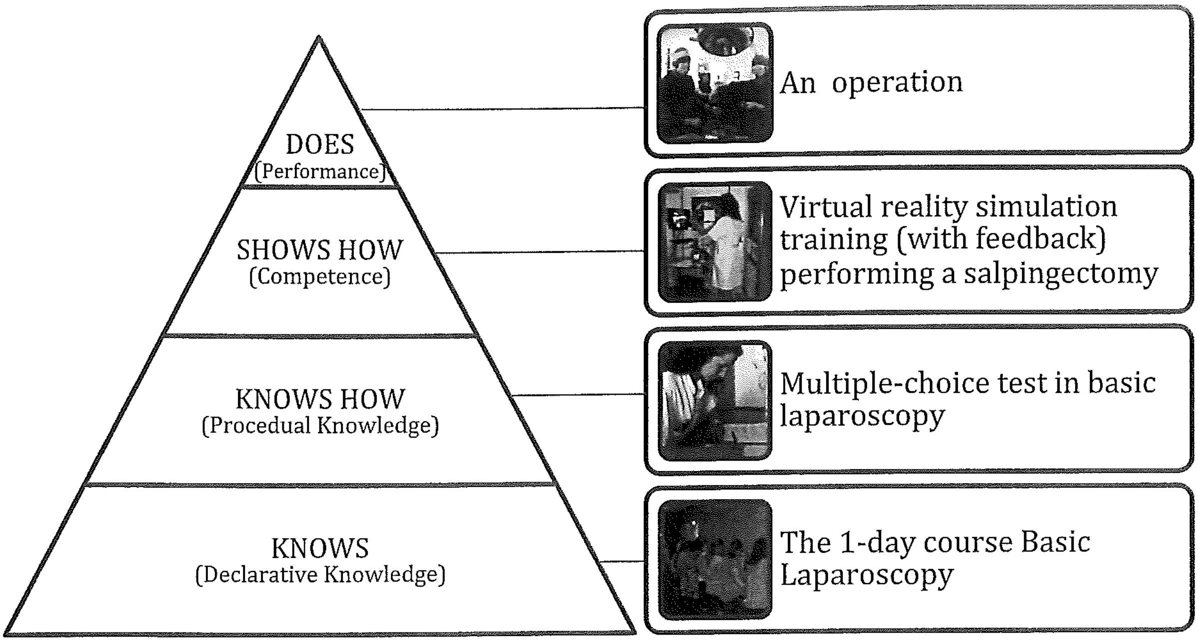


Figure 5. Linkage between Miller's competence pyramid and the four-step curriculum

In the light of the current thesis, we have found that a skills curriculum is not a steady state: it is a process that should be tailored to individual needs, Figure 6. The completion rate demonstrated that there are still several considerations to be made about, for example, in which order the four steps occur. There is no doubt that the fourth step, the operation, must be placed last, but, due to logistic considerations, the other three steps could be in random order to ensure more flexibility, both on an individual level and on an organizational level.

As portrayed in Figure 6, education is a process where a curriculum is a co-player. It is difficult to determine what step comes first in education; it is more rational to view it as a dynamic process between assessment and learning, teaching and a curriculum.

This dynamic process can also be considered in the light of Miller's pyramid. Some levels might need repetitions, e.g. step down one level and repeat if necessary.

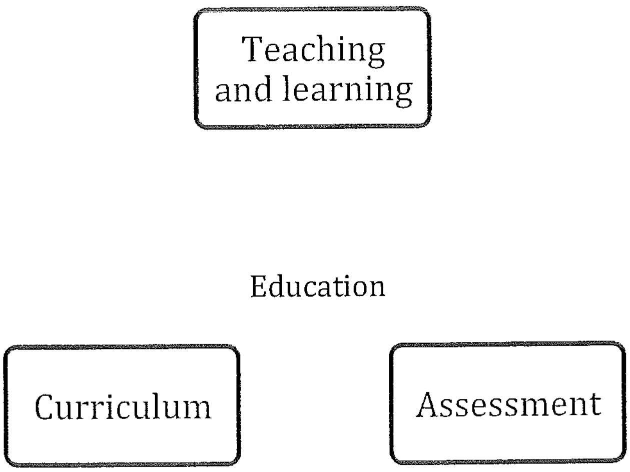


Figure 6. The dynamic process of education

The main reason for the various improvements in skills training is undoubtedly to improve patient safety, and the simulated setting can be the start of a growing pattern of personal development that interweaves practical skills with other elements of professionalism. Nonetheless, this cannot be done in a single curriculum: it is an ongoing process throughout a medical career.

Limitations

Some limitations are outlined under the individual steps.
One could argue whether it is reasonable to use the term validate about the four-step curriculum since the results are not compared with the performance of a control group. As mentioned in the background chapter, we decided not to conduct a randomized trial with a traditionally trained control group due to ethical considerations. Nevertheless, that would have been the ultimate justification of a curriculum.
We firmly believe that all steps in the four-step curriculum have been validated individually, although the fourth step, the operational step, was not measured on summative outcome. We deliberately did not choose a cut-off value for the assessment of the operation, because it is not supported in the present literature [88] and because the nature of the curriculum is a skills training intervention, not a screening examination.

During the development of the four-step curriculum, we involved doctors mainly from a regional setting and therefore the curriculum is not necessarily nationally representative. If the four-step curriculum were to be implemented nationally, a validation process involving the other regions would have to take place. However, we tried to accommodate a broad range of views on laparoscopy by involving all hospitals in a region, small as well as large.

Perspectives and future studies

The findings in the present thesis have contributed to implementation of obligatory training of basic skills and operational tasks on a virtual reality simulator in the national statement of aims. However, the whole four-step curriculum is still voluntary; we believe it should be mandatory. Furthermore, the 1-day course, Basic Laparoscopy, is now a permanent part of residency training in two of three regions in Denmark.

To optimize surgical performance and reduce errors, not only among novices, but also among laparoscopic inexperienced older doctors, simulation training could be of value. Simulation training needs to become an inherent part of training for all doctors involved in surgery. All doctors can learn from practicing.

It is important to continually develop the four-step curriculum to keep it contemporary. Furthermore, a future study should develop a curriculum for more senior surgeons with advanced laparoscopic procedures. The virtual reality simulators improve almost from day to day, and new virtual procedures, such as laparoscopic hysterectomy, are now available. Additionally, it would be interesting to elaborate in the field of expert performance and deliberate practice, which is the highest level of technical skills acquisition, in relation to the domains of surgical simulation.

Conclusion

We propose a flexible four-step curriculum in basic laparoscopy that is patient-safe, efficient and accommodates ethical considerations.
The four-step curriculum should contain:

- A 1-day hands-on course in basic laparoscopy
- A validated multiple-choice test revolving around basic laparoscopic principles and procedures.
- Structured virtual reality simulation training involving procedure modules and with instructor feedback.
- An operation of trained virtual reality procedure with subsequent formative feedback.

Acknowledgements

I would like to express my gratitude to all my co-workers, supervisors, participants, family and friends. In particular I would like to mention the following people:

Mathilde Maagaard, MD, PhD fellow, Department of Obstetrics and Gynecology, The Juliane Marie Center, Rigshospitalet, Copenhagen University Hospital, for her continuous, sharp insight into research along with detailed feedback and constructive criticism. She has done an invaluable job in supporting the project and me.

Sofie Leisby Antonson, MD, PhD fellow, Department of Obstetrics and Gynecology, The Juliane Marie Center, Rigshospitalet, Copenhagen University Hospital, has been a great support in discussing all aspects of research. Without her daily - and never failing - good spirits, at times it would have been hard to see the end of projects.

Jette Led Sørensen, MD, Senior Consultant, MMEd, Department of Obstetrics and Gynecology, The Juliane Marie Center, Rigshospitalet, Copenhagen University Hospital, for her inexhaustible source of ideas and drive to improve medical education. She is a skilled and inspiring supervisor along with being an excellent travelling buddy.

Christian Rifbjerg Larsen, MD, PhD, Consultant, Department of Obstetrics and Gynecology, Hillerød Hospital, Denmark, for drawing me into simulation research and entrusting me with the successor to his own eminent research. Together, I hope we can foster new PhD students and further develop simulation research.

Teodor Grantcharov, MD, PhD, Senior Consultant, Head of Water's Family Simulation Center, St. Michael's Hospital, Toronto, Canada, for his inspiring, pioneering— and continuous— work within surgical simulation. His unpretentious way of being a star in the surgical world is admirable.

Flemming Bjerrum, MD, PhD fellow, Department of Obstetrics and Gynecology, The Juliane Marie Center, Rigshospitalet, Copenhagen University Hospital, has, with his uncontrollable lust for work and great knowledge of the simulation world, contributed significantly to completion of several of the studies in this thesis. I am delighted that I get to be your supervisor in the ongoing projects.

Charlotte Ringsted, MD, PhD, MHPE, Professor, Head of Centre for Clinical Education (CEKU), Rigshospitalet, Copenhagen University Hospital, has, through her detailed feedback and constructive criticism, been an indispensable resource. Her knowledge within the educational field is incomparable. Also, many thanks to the staff at CEKU.

Christian Gluud, MD, DMSc and **Per Winkel**, MD, DMSc at Copenhagen Trial Unit, Center for Clinical Intervention Research, University of Copenhagen, have provided outstanding help to improve the protocol and the following statistics for the Feedback Trial. Both have given me essential insight in the nature of randomized controlled trials.

Andreas Strandbygaard, Master of Political Science, Management Consultant at Accenture Denmark for his limitless support to the thesis and me. Throughout the past three years I have benefitted from his brilliant mind, which has translated into some interesting research ideas.

Bent Ottesen, MD, DMSc, Professor, Head of the Juliane Marie Center at Rigshospitalet, Copenhagen University Hospital has an enviable future vision that is beneficial for all involved in research. He is a strategic genius and not many people can with such calmness, kindness and inspiration head a hospital center. I thank him very much for never losing faith in projects.

I would also like to thank my fellow PhD students for happy times and good discussions: Sune Werner Räder, Mette Petri, Lea Langhoff Thuesen, Janne Bentzen, Tanja Roien Jakobsen, Line Rode, Charlotte Ekelund, Martin Tolstrup, Carsten Fagö-Olsen, Louise Kelstrup, Tine Greve, Helene Hvidmand and lastly master student and midwife AnnLouise Westergaard.
Furthermore, a warm thanks to the following staff at the Juliane Marie Centre: Jonna Hemmingsen, Anette Kajander, Franciska Holm Hansen and Linda Svenstrup Munk.

Finally, I would like to thank my friends and family for their never-ending support and good advice.

Funding and disclosures

It would not have been possible to complete the research in this PhD thesis without the funding from the following foundations, departments and hospitals. I am very grateful for their support and hope they appreciate the result of a better and safer surgical education for residents and the potential benefits for patients.

- Tryg Foundation
- Righsospitalet, Copenhagen University Hospital
- Juliane Marie Center, Rigshospitalet, Copenhagen University Hospital
- Toyota-Foundation
- Beckett-Foundation
- Aase and Ejnar Danielsens Foundation
- Alternative Foundation
- King Christian X Foundation

There have been no conflicts of interest in any of the four studies.
The content of this thesis is solely the responsibility of the authors and does not necessarily represent the official views of the aforementioned foundations.

Reference list

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. 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 Oct.;236(4):458–63; discussion463–4.

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

4. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg. 2004 Feb.;91(2):146–50.

5. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev. 2009;(1):CD006575.

6. Stefanidis D, Acker CE, Swiderski D, Heniford BT, Greene FL. Challenges during the implementation of a laparoscopic skills curriculum in a busy general surgery residency program. J Surg Educ. 2008 Jan.;65(1):4–7.

7. Stefanidis D, Heniford BT. The formula for a successful laparoscopic skills curriculum. Arch Surg. 2009 Jan.;144(1):77–82; discussion82.

8. Korndorffer JR, Stefanidis D, Scott DJ. Laparoscopic skills laboratories: current assessment and a call for resident training standards. Am. J. Surg. 2006 Jan.;191(1):17–22.

9. Karvonen J, Gullichsen R, Laine S, Salminen P, Grönroos JM. Bile duct injuries during laparoscopic cholecystectomy: primary and long-term results from a single institution. Surg Endosc. 2007 Jul.;21(7):1069–73.

10. Avital S, Hermon H, Greenberg R, Karin E, Skornick Y. Learning curve in laparoscopic colorectal surgery: our first 100 patients. Isr. Med. Assoc. J. 2006 Oct.;8(10):683–6.

11. Lin YY, Shabbir A, So JBY. Laparoscopic appendectomy by residents: evaluating outcomes and learning curve. Surg Endosc. 2010 Jan.;24(1):125–30.

12. Kumar U, Gill IS. Learning curve in human laparoscopic surgery. Curr Urol Rep. 2006 Mar.;7(2):120–4.

13. Eto M, Harano M, Koga H, Tanaka M, Naito S. Clinical outcomes and learning curve of a laparoscopic adrenalectomy in 103 consecutive cases at a single institute. Int. J. Urol. 2006 Jun.;13(6):671–6.

14. Chin EH, Hazzan D, Edye M, Wisnivesky JP, Herron DM, Ames SA, et al. The first decade of a laparoscopic donor nephrectomy program: effect of surgeon and institution experience with 512 cases from 1996 to 2006. J. Am. Coll. Surg. 2009 Jul.;209(1):106–13.

15. Adibe OO, Nichol PF, Flake AW, Mattei P. Comparison of outcomes after laparoscopic and open pyloromyotomy at a high-volume pediatric teaching hospital. J. Pediatr. Surg. 2006 Oct.;41(10):1676–8.

16. Fleisch MC, Newton J, Steinmetz I, Whitehair J, Hallum A, Hatch KD. Learning and teaching advanced laparoscopic procedures: do alternating trainees impair a laparoscopic surgeon's learning curve? J Minim Invasive Gynecol. 2007 May;14(3):293–9.

17. Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, et al. Surgical simulation: a systematic review. Ann. Surg. 2006 Mar.;243(3):291–300.

18. Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand. 2012 Jun. 13.

19. Coleman RL, Muller CY. Effects of a laboratory-based skills curriculum on laparoscopic proficiency: a randomized trial. Am. J. Obstet. Gynecol. 2002 Apr.;186(4):836–42.

20. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. Br J Surg. 2008 Sep.;95(9):1088–97.

21. Halsted WS. The Training of the surgeon. 1904. p. 25.

22. Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, et al. Beyond "see one, do one, teach one": toward a different training paradigm. Postgrad Med J. 2009 May;85(1003):244–9.

23. Schout B. Training in Urology. From Virtual to Reality, PhD thesis, Eindhoven, the Betherlands, 2010

24. Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. Acad Med. 2005 Jun.;80(6):549–53.

25. Miller G. The assessment of clinical skills/competence/performance. Acad Med. 1990 65:63-7

26. Kirkpatrick DL. Evaluating Training Programs. Tata McGraw-Hill Education; 1975.

27. Kirkpatrick DL, Basarab D, Freitag E. Evaluating Training Programs: The Four Levels. 2nd ed. Berrett-Koehler Publishers; 1998.

28. Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Eval Program Plann. 2004 Jul. 31;27(3):341–7.

29. Posner M. Human performance. Brocks/Cole Pub. Co, Belmont 1967

30. Reznick RK, Macrae H. Teaching surgical skills--changes in the wind. N. Engl. J. Med. 2006 Dec. 21;355(25):2664–9.

31. Sadideen H, Kneebone R. Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? Am. J. Surg. 2012 Jun. 9.

32. Kopta JA. The development of motor skills in orthopaedic education. Clin. Orthop. Relat. Res. 1971 Mar.;75:80–5.

33. Nguyen PH, Acker CE, Heniford BT, Stefanidis D. What is the cost associated with the implementation of the FLS program into a general surgery residency? Surg Endosc. 2010 Dec.;24(12):3216–20.

34. Kapadia MR, DaRosa DA, Macrae HM, Dunnington GL. Current assessment and future directions of surgical skills laboratories. J Surg Educ. 2007 Sep.;64(5):260–5.

35. Dent J, Harden RM. A Practical Guide for Medical Teachers. 3rd ed. Churchill Livingstone; 2009.

36. Harden RM. Ten questions to ask when planning a course or curriculum. Med Educ. 1986 Jul.;20(4):356–65.

37. Kneebone RL. Practice, rehearsal, and performance: an approach for simulation-based surgical and procedure training. JAMA. 2009 Sep. 23;302(12):1336–8.

38. Harden RM. Ten questions to ask when planning a course or curriculum. Med Educ. 1986 20:356-65

39. Aggarwal R, Grantcharov TP, Darzi A. Framework for systematic training and assessment of technical skills. J. Am. Coll. Surg. 2007 Apr.;204(4):697–705.

40. Tang B, Hanna GB, Cuschieri A. Analysis of errors enacted by surgical trainees during skills training courses. Surgery. 2005 Jul.;138(1):14–20.

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

42. Rosenthal ME, Ritter EM, Goova MT, Castellvi AO, Tesfay ST, Pimentel EA, et al. Proficiency-based Fundamentals of Laparoscopic Surgery skills training results in durable performance improvement and a uniform certification pass rate. Surg Endosc. 2010 Oct.;24(10):2453–7.

43. Scott DJ, Ritter EM, Tesfay ST, Pimentel EA, Nagji A, Fried GM. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. Surg Endosc. 2008 Aug.;22(8):1887–93.

44. Swanstrom LL, Fried GM, Hoffman KI, Soper NJ. Beta test results of a new system assessing competence in laparoscopic surgery. J. Am. Coll. Surg. 2006 Jan.;202(1):62–9.

45. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am. J. Surg. 2010 Jan.;199(1):115–20.

46. Zheng B, Hur H-C, Johnson S, Swanstrom LL. Validity of using Fundamentals of Laparoscopic Surgery (FLS) program to assess laparoscopic competence for gynecologists. Surg Endosc. 2010 Jan.;24(1):152–60.

47. Hur H-C, Arden D, Dodge LE, Zheng B, Ricciotti HA. Fundamentals of laparoscopic surgery: a surgical skills assessment tool in gynecology. JSLS. 2011 Jan.;15(1):21–6.

48. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. Surg Innov. 2007 Jun.;14(2):107–12.

49. Panait L, Bell RL, Roberts KE, Duffy AJ. Designing and validating a customized virtual reality-based laparoscopic skills curriculum. J Surg Educ. 2008 Nov.;65(6):413–7.

50. Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. Am. J. Surg. 2006 Jan.;191(1):128–33.

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

52. Palter VN, Grantcharov TP. Development and Validation of a Comprehensive Curriculum to Teach an Advanced Minimally Invasive Procedure: A Randomized Controlled Trial. Ann. Surg. 2012 Jun. 1.

53. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg. 2008 Feb.;32(2):182–8.

54. Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, et al. Proving the value of simulation in laparoscopic surgery. Ann. Surg. 2004 Sep.;240(3):518–25–discussion525–8.

55. Kneebone R, Bello F. Surgical Simulation. Riley RH, editor. A Manual of Simulation in Healthcare. Oxford: Oxford University Press, USA; 2008. p. 435–48.

56. Kern DE, Thomas PA, Hughes MT, editors. Curriculum Development for Medical Education: A Six-Step Approach. second edition. The Johns Hopkins University Press; 2009. p. 272.

57. Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BLH, Scherpbier AJJA. Validation and implementation of surgical simulators: a critical review of present, past, and future. Surg Endosc. 2010 Mar.;24(3):536–46.

58. Grantcharov TP, Reznick RK. Teaching procedural skills. BMJ. 2008 May 17;336(7653):1129–31.

59. Morozov V, Nezhat C. Proposal of a formal gynecologic endoscopy curriculum. J Minim Invasive Gynecol. 2009 Jul.;16(4):416–21.

60. Palter VN, Graafland M, Schijven MP, Grantcharov TP. Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: A Delphi approach. Surgery. 2011 Oct. 21.

61. Moulton C-AE, 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 Sep.;244(3):400–9.

62. Ericsson K, Krampe R, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. Psychological review. 1993;(100):363–406.

63. Kolb DA. Experiential learning. Prentice Hall; 1984. p. 256.

64. University of California, Berkeley. Operations Research Center, Dreyfus SE. A five-stage model of the mental activities involved in directed skill acquisition. 1980.

65. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003 Sep.;37(9):830–7.

66. Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ. 1996;1(1):41–67.

67. Fritsche L, Greenhalgh T, Falck-Ytter Y, Neumayer H-H, 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 Dec. 7;325(7376):1338–41.

68. Condous G, Alhamdan D, Bignardi T, van Calster B, van Huffel S, Timmerman D, et al. The value of laparoscopic skills courses. Aust N Z J Obstet Gynaecol. 2009 Jun.;49(3):312–5.

69. 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 Aug.;47(4):277–83.

70. Case SM. Constructing written test questions for the basic and clinical sciences. 2nd ed. National Board of Medical Examiners; 1998.

71. Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. Psychol Sci. 2006 Mar.;17(3):249–55.

72. Larsen DP, Butler AC, Roediger HL. Test-enhanced learning in medical education. Med Educ. 2008 Oct.;42(10):959–66.

73. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. Med Educ. 2009 Jan.;43(1):21–7.

74. Magill RA. Motor Learning and Control. McGraw-Hill Humanities/Social Sciences/Languages; 2010. p. 480.

75. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010 Jan.;44(1):85–93.

76. Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. J Mot Behav. 2004 Jun.;36(2):212–24.

77. Korndorffer JR, Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. J. Am. Coll. Surg. 2005 Jul.;201(1):23–9.

78. Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. Am. J. Surg. 2009 Apr.;197(4):447–9.

79. Ende J. Feedback in clinical medical education. JAMA. 1983 Aug. 12;250(6):777–81.

80. Bienstock JL, Katz NT, Cox SM, Hueppchen N, Erickson S, Puscheck EE, et al. To the point: medical education reviews--providing feedback. Am. J. Obstet. Gynecol. 2007 Jun.;196(6):508–13.

81. Krackov SK. Giving feedback. A Practical Guide for Medical Teachers. 3rd ed. Churchill Livingstone; 2009. p. 357–67.

82. Snyder CW, Vandromme MJ, Tyra SL, Hawn MT. Proficiency-based laparoscopic and endoscopic training with virtual reality simulators: a

comparison of proctored and independent approaches. J Surg Educ. 2009 Jun.;66(4):201–7.

83. Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. Adv Health Sci Educ Theory Pract. 2002;7(1):63–80.

84. Duffy FD, Holmboe ES. Self-assessment in lifelong learning and improving performance in practice: physician know thyself. JAMA. 2006 Sep. 6;296(9):1137–9.

85. Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. Acad Med. 2005 Oct.;80(10 Suppl):S46–54.

86. 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 Feb.;84(2):273–8.

87. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Acad Med. 1996 Dec.;71(12):1363–5.

88. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010 Jul.;97(7):972–87.

89. 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 Jun.;115(7):908–16.

90. Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL. Can both residents and chief physicians assess surgical skills? Surg Endosc. 2012 Jan. 20.

91. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann. Intern. Med. 2002 Sep. 17;137(6):511–20.

92. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med. 2008 Nov.;15(11):988–94.

93. Korndorffer JR, Scott DJ, Sierra R, Brunner WC, Dunne JB, Slakey DP, et al. Developing and testing competency levels for laparoscopic skills training. Arch Surg. 2005 Jan.;140(1):80–4.

94. Carter Y, Jackson N. Medical education and training. Oxford University Press, USA; 2008. p. 350.

95. Chang L, Petros J, Hess DT, Rotondi C, Babineau TJ. Integrating simulation into a surgical residency program: is voluntary participation effective?

Surg Endosc. 2007 Mar.;21(3):418–21.

96. Haluck RS, Satava RM, Fried G, Lake C, Ritter EM, Sachdeva AK, et al. Establishing a simulation center for surgical skills: what to do and how to do it. Surg Endosc. 2007 Jul.;21(7):1223–32.

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

98. Sutherland K, Leatherman S. Does certification improve medical standards? BMJ. 2006 Aug. 26;333(7565):439–41.

99. Rethans J-J, Norcini JJ, Barón-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: implications for assessing practice performance. Med Educ. 2002 Oct.;36(10):901–9.