Take-Home Training in Laparoscopy

Ebbe Thinggaard

This review has been accepted as a thesis together with five original papers by University of Copenhagen 10th of August 2016 and defended on 2nd of December 2016.

Tutors: Ismail Gögenur and Lars Konge

Official opponents: Jacob Rosenberg, Lars Lund and Teodor Grantcharov

Correspondence: Department of Surgery, Zealand University Hospital, Lykkebækvej 1, 4600 Køge, Denmark.

E-mail: ebbe.thinggaard@gmail.com

Dan Med J 2017;64(4):B5335

THE ORIGINAL FIVE PAPERS ARE

1. Thinggaard E, Kleif J, Bjerrum F, Strandbygaard J, Gögenur I, Ritter E M, Konge L. Off-site training of laparoscopic skills, a scoping review using a thematic analysis. Surgical Endoscopy 2016;11:1-9.

2. Thinggaard E, Bjerrum F, Strandbygaard J, Gögenur I, Konge L. Validity of a cross-specialty test in basic laparoscopic techniques (TABLT). British Journal of Surgery 2015;102:1106–1113.

3. Thinggaard E, Bjerrum F, Strandbygaard J, Gögenur I, Konge L. Ensuring competency of novice laparoscopic surgeons – Exploring standard setting methods and their consequences. Journal of Surgical Education 2016;73:986-991.

4. Thinggaard E, Bjerrum F, Strandbygaard J, Konge L, Gögenur I. Take-home training in laparoscopy facilitates distributed training, a randomized trial. Surgical Endoscopy 2016; Submitted.

5. Thinggaard E, Konge L, Bjerrum F, Strandbygaard J, Gögenur I, Spanager L. Take-home training in a simulation-based laparoscopy course. Surgical Endoscopy 2016;1-8.

INTRODUCTION

Minimally invasive surgery, and laparoscopy in particular, have revolutionised surgical care (1). The implementation of operative laparoscopy has reduced the duration of hospital stay and the convalescence period and has helped improve patient outcomes and enhance recovery after surgery (2, 3). Laparoscopy has gone through a series of developmental phases and has become the primary choice as a surgical technique (4). Laparoscopy is what surgical trainees across the different specialties are trained to do when performing surgery. Although the laparoscopic technique has many benefits, implementation was a challenge. Laparoscopic skills are very different to those used in open surgery and require specific training (5). The need for more training was recognised in the early phases of the development of the laparoscopic technique (6, 7). In the United States, training and passing a test is now a requirement in order to be certified as a general surgeon to be proficient in laparoscopy (8).

Originally, laparoscopic skills were primarily taught using the apprenticeship model. However, this method has certain limitations: it requires longer time to practice and more learning opportunities in clinical practice. The unique set of skills required in laparoscopy highlighted the need for new training methods, so simulation training was developed. During the last few decades, it has been firmly established that laparoscopic skills can be acquired outside the OR using simulators (9). Training can be done on either virtual reality simulators (VRSs) or boxtrainers (BTs); both methods have been shown to be effective methods for providing laparoscopic skills training (10).

Despite the evidence for the use of simulation training, implementation has been unsystematic (11). Barriers to simulation training – including the need for time to train, the high price of simulation equipment, and the lack of access – have halted the implementation of simulation training in laparoscopic training programmes (12). Low accessibility in particular has been a constraint. Access to training can be improved by using mobile BTs, as this method is affordable, accessible and mobile (12). BTs allow trainees to train according to their own schedule. Furthermore, BTs have the potential to improve laparoscopic training when implemented in a simulation-based laparoscopic training programme. Unfortunately, little is known about the use of BTs in off-site training. No review has been done to explore this, and little research has been conducted into the effects of off-site training.

BACKGROUND

DEVELOPMENT OF OPERATIVE LAPAROSCOPY

The move towards minimally invasive surgery can be traced back as far as Hippocrates in ancient Greece (13) and Abu Al Qasim Al-Zahrawi in medieval Spain (1). Both Hippocrates and Al-Zahrawi developed speculums to allow access to body cavities in order to alleviate symptoms and treat diseases. Minimally invasive surgery has gone through a serious of developmental phases since then. What we have come to know as operative laparoscopy today was primarily developed in the 1970s, when the technique was explored and many new devices were invented (13). The breakthrough in laparoscopy came when it was used for cholecystectomies (1). In 1987, Mouret was the first to perform a laparoscopic cholecystectomy, and soon others followed. There were initially challenges with the large-scale implementation. The initial increase in bile duct injuries illustrated the problems associated with implementing a new technique at such a quick pace (7). Despite the increase in bile duct injuries, laparoscopic cholecystectomy was implemented for more widespread use (1). The benefits of laparoscopy include the reduction of the duration of hospital stay and convalescence period (2, 3). Laparoscopy has since become the preferred technique for intra-abdominal surgery (4). Although the technique was initially used for benign surgery, it has also found a use in oncological procedures, where the same benefits have been demonstrated without compromising the oncological outcome (14).

TRAINING OF LAPAROSCOPIC TECHNIQUES

Laparoscopy requires a very different technique compared to open surgery. When operating using the laparoscopic technique, the surgeon uses a set of very specific skills (5). The surgeon will have to accommodate for the loss of depth perception and the limited range of movement and adjust to the use of long instruments fixed at skin level. However, the necessary skills can be acquired through training. Training is needed to become proficient in laparoscopy and a new surgical curriculum for laparoscopy was needed (6, 7). In the early days of laparoscopy, training was done in the same manner as for open surgical skills. Training was based on the apprenticeship model and primarily conducted during supervised surgery in the operating room (OR) (15). Nonetheless, the Hallstead approach, often referred to as 'see one, do one, teach one', was found to be insufficient for laparoscopic skills training . The Hallstead model required a long mentoring process, which was neither cost-effective nor compatible with the increased awareness of the potential risk to patients from having untrained surgeons performing laparoscopy (15). Simulation-based training was suggested as a solution and BTs were developed (16). Even VRSs were proposed early on as a potential training method (17-19). Although research has firmly established that laparoscopy can be taught outside the OR using simulation training (9), it has not been implemented systematically (11). A body of evidence supports the argument that VRSs and BTs are effective methods of acquiring laparoscopic skills (9). Laparoscopic simulation-based training shortens operating time, increases intra-operative skills, and reduces the risk of intra-operative and post-operative complications (20-24). Although many VRSs and BTs exist, systematic training programmes are lacking and have only been implemented in a few countries (11). An example of this is the FLS, which is a requirement for becoming a general surgeon in the USA (8). However, the FLS is not accessible to most surgeons or to those most in need of laparoscopic simulation training. The tasks included in the FLS are advanced and are therefore inappropriate for novices. Requirements in training should correspond to the trainee's level of experience (25). There is a need for a new test that is appropriate, affordable and accessible to novices.

Testing and the pass/fail level of a test provide the minimum requirement for training in a proficiency-based training programme. Proficiency-based training has been recommended for simulation-based laparoscopy training and has been shown to be effective (26, 27). Training programmes using proficiency-based training rely on the use of testing and pass/fail levels. Standard-setting methods are used when setting a pass/fail level for a test. However, the pass/fail level of the test may vary considerably depending on the standard setting method (28). Despite this, few studies have explored the consequences of the choice of standard setting method in a proficiency-based laparoscopic training programme.

TAKE-HOME TRAINING

Proficiency-based laparoscopic training programmes have been developed for BTs (29). Training laparoscopic skills on BTs is an effective method of acquiring laparoscopic skills (30), and a variety of BTs have been developed for this purpose. These range from BTs that are similar to the OR setting where laparoscopic cameras are used, to simple trainers (31) and even open trainers to practice moving instruments (32). Some BTs include an eye-patch to remove stereovision (33) and other trainers use mirrors (34, 35). BTs have even been developed to address ergonomy (36). The development of BT designs has followed that of digital cameras. The early BTs used large home video cameras (16) and web-cameras were also used (37, 38). Both do-it-yourself (DIY) trainers (16, 39, 40) and commercially available ones have been suggested (41). Recently, tablet computers such as the iPad (31, 42) and even telephones (43) have been used in BT design. The on-going development of BTs have made them accessible, affordable and mobile (12).

Mobility in training can help overcome some of the barriers to simulation training. Times to train, access and the high price for simulation equipment are barriers to simulation training (12). Mobile BTs make training accessible to novice laparoscopic trainees as they can plan their training according to their own schedule. Despite this apparent advantage, no review has yet explored the literature regarding off-site training. BTs may not only provide trainees with the flexibility they need but could also prove beneficial when implemented in existing simulation-based training programmes. Currently, simulation-based training programmes primarily offer training at a simulation centre or skills lab. Additionally training at home could improve current training programmes and increase the use of laparoscopic simulation training in general.

THEORETICAL CONSIDERATIONS

Laparoscopic training programmes should be based on sound educational theories. Research in medical education must be placed in a theoretical framework in order to support evidence of findings (44). Below we outline the most prominent educational theories used in this thesis.

PROFICIENCY-BASED LEARNING

In the early 1900s the Flexner Report made it clear that competency-based medical education (CBME) was necessary to ensure high-quality training in medicine (45, 46). CMBE has recently undergone a revival in post-graduate training and surgical skills training. Passing a competency-based test is now a requirement in the USA in order obtain specialty registration as a general surgeon (8). CBME is being introduced as proficiency-based training in laparoscopic skills training (15, 27). Proficiency-based training is a further development of CBME, where the minimum requirement is that of proficiency rather than competency. Research has shown the effect of proficiency-based training on OR performance (21, 29). Proficiency-based training is also recommended for simulation training in laparoscopy (27). Proficiency-based training relies on the use of assessment and testing. Testing has a positive effect on retention of learned skills (47). Proficiency-based standard setting has been implemented using performance levels of experienced laparoscopic surgeons (48).

However, when implementing tests it is important that they are supported by evidence of validity (49). Testing should be an integrated part of the training curriculum in which learning objectives, content, assessment and use of skills in the OR are aligned. Proficiency-based training and testing are discussed in the papers included in this thesis.

DISTRIBUTED PRACTICE

Distributed practice, or spaced repetition as it is sometimes referred to, is an effective educational approach. Distributed practice is superior to massed practice as learners can divide their learning into manageable parts (50). This is also referred to as the spacing effect. Distributing learning experiences across a number of days is effective and may also apply in motor skills training (51). It is recommend for gaining knowledge and motor skills (52, 53). The effect has also been demonstrated in laparoscopic skills training (54, 55). Nonetheless, the optimal distribution of practice remains to be established (56). Distributed practiced is also a part of deliberate practice, which is effective in procedural skills training (57). In this thesis, distributed practice is discussed in the first, fourth and fifth papers.

SELF-REGULATION IN LEARNING

Self-regulation in learning relies on theories from both pedagogy and psychology. Two terms that are often used to describe self-regulation in learning are Self-Regulated Learning (SRL) and Self-Directed Learning (SDL). SRL is sometimes confused or used interchangeably with SDL, and these two terms have been used in various ways in the literature (58). SRL is when one considers the students' independence in training, whereas SDL is used when training promotes autonomous learning. A new theory in medical education is Directed Self-Regulated Learning (DSRL), which is based on theories of SRL. DSRL is a new approach to learning in which trainees regulate their own training within a framework provided by educators. Trainees thereby control a part of the training and are active participants in their own learning process. Faculty act as facilitators to guide the self-regulated learning by providing a framework in which learners operate (59). SRL is the basis for DSRL, where the focus is on understanding autonomous learning to help provide guidance during learning experiences (60). SRL is recommended in a recent systematic review for simulation-based training (61). One study has proven DSRL to be effective for retention of skills in a simulation training for lumbar puncture (60). DSRL could be of value in training programmes where supervision is difficult to provide and feedback from faculty is unavailable. In the present thesis, the potential for DSRL in off-site training in laparoscopy is discussed in Paper 1.

METHODOLOGICAL CONSIDERATIONS

High-quality research is based on the use of appropriate methods. The research question, aim of the study and hypotheses should determine the choice of methodology. In the following section we will describe the setting and methods used in this thesis.

SETTING

<u>The CAMES Training Programme in Laparoscopy</u>

All of the studies included in this PhD dissertation were conducted at the Copenhagen Academy for Medical Education and Simulation (CAMES) (62). At CAMES, doctors in their first year of specialty training can participate in a basic simulation-based laparoscopic training programme. The training programme is a cross-specialty training programme for doctors working in the gynaecology, urology and surgery departments (63) and aims to prepare participants for their first supervised laparoscopic surgical procedure. The training programme is structured across two formalised training days separated by a period of self-directed training. Training takes place using VRSs and BTs. The first part of the training programme is an introductory course. This is followed by a period of self-directed training in which participants practice on VRSs and BTs. Participants train in the simulation centre, assisted by a simulator technician who can provide some feedback during training. There are two mandatory requirements in the training programme. The first requirement is to pass the TABLT (64) test on the BT and the second is to reach a predefined level of proficiency on the VRS. Gynaecologists are also required to take a theoretical test (65). When participants have completed the requirements of the training programme they are able to enrol in the final operative course. The operative course marks the end of the training programme.

<u>Training and Assessment of Basic Laparoscopic Techniques (TABLT)</u>

To address the need for a training and assessment tool for basic laparoscopic skills training, we developed TABLT. The tasks included in TABLT were developed for a cross-specialty training curriculum. The TABLT consists of laparoscopic tasks that test various domains of laparoscopic skills. The test includes the testing and training of ambidexterity, hand-eye coordination, accommodating for the fulcrum effect, guiding instruments via a screen, and economy of movement. Five tasks were developed covering appropriate handling of laparoscopic instruments, cutting, blunt dissection and sharp dissection. There are specific errors for each task, which are used when calculating the TABLT test score.

The tasks:

Task 1 is a coordination task. The goal of the task is to move four beads on a pegboard from one line of pegs on the right side, to another line of pegs on the left side, and back again. An error is counted if a bead is dropped. If the bead rolls outside the range of movements it is counted as two errors.

Task 2 is a cutting task, the goal of which is to cut out a circle. A circle is drawn on a soft sponge cloth and the task has been completed when the entire circle has been cut out. The circle is two millimetres wide. An error is counted when cutting outside the two-millimetre line.

Task 3 is a sharp dissection task, the goal of which is to dissect a vessel using sharp dissection technique. The vessel is made from a balloon that has two lines drawn on it. The lines are two millimetre wide and two centimetres apart. The balloon is wrapped in a soft sponge cloth. A cut into the vessel is counted as an error.

Task 4 is a blunt dissection task and its goal is to dissect a vessel using blunt dissection technique. As in Task 3, the vessel is made from a balloon that has two lines drawn on it. The lines are two millimetre wide and two centimetres apart. The balloon is wrapped in cotton wool. An error is counted if a piece of the cotton wool is ripped off, and completely removed from the rest of the task.

Task 5 is a cyst removal task, the goal of which is to remove a cyst. The cyst can also be made to simulate a gallbladder depending on the specialty. The cyst is made from two round balloons, one inside the other. The innermost balloon is filled with 60 millilitres of ultrasound gel.

The TABLT test scoring system

The scoring system is based on time and the number of errors, similar to the system used in the FLS system (48). The final score is between 0 and 708, where a higher score indicates a better result. The score is calculated by first determining a score for each task. Each task score is calculated by subtracting the time spent on a task in seconds from a maximum time of 600 seconds. The result is then divided by the average score of a group of experienced surgeons. Finally, all task scores are summed into a performance score. The score can be simply calculated using an Excel spreadsheet; the users only need to enter the number of errors and time spent completing each task.

A MIXED-METHODS APPROACH

The papers included in the thesis are based on several research questions. Depending on the research question, an appropriate choice of methodology has been chosen, as is recommended (66). Throughout the thesis, qualitative and quantitative methods are seen as equal and complementary (67). The first paper is a scoping review based on a systematic literature search and a qualitative thematic analysis. The second paper is a validation study. For this study, we used the unitary framework of validity, which relies on both qualitative and quantitative methods. The third was a methodological paper that explored the consequences of different standard setting methods. In this study we used a descriptive and a comparative quantitative analysis. The fourth paper was designed as a randomised controlled trial and we used quantitative statistical methods, described below. The fifth study is a mixed-methods study in which quantitative descriptive statistics were used and triangulated with findings from a qualitative content analysis of focus group interviews and individual interviews. The mixed methodology was used to explore training patterns and training methods as well as how trainees used BTs when training at home.

VALIDITY

Validity is the process of ensuring that what you intend to measure is what is actually measured (68). With a directly observable phenomenon, such as blood glucose concentration or haemoglobin levels, validation is done within the realm of natural sciences. However, some traits such as surgical technical competency cannot be observed or measured directly. As the trait is not directly measurable, a test or an assessment tool is necessary. To ensure validity for this type of measurement, a different approach to validation is used. Frameworks have been developed for exploring the validity of tests measuring competency. These frameworks include the unitary framework of validity, although other frameworks of validity theory do also exist (69). An early validity framework used types of validity and describes the use of content, construct and criterion validity. However, the unitary framework considers all validity as a process rather than dividing it into types of validity. Hereby, validity of construct is explored; therefore, validity is construct validity. The unitary framework of validity is recommended in the 'Standards for Educational and Psychological Testing' (70). A newer framework called use/argument based validity also exists (71).

In this thesis we use the contemporary framework of unitary validity, an approach that has been recommended in the assessment literature for more than 15 years (72, 73). In the unitary framework of validity, validity is described as a hypothesis or process and evidence is gathered from different sources to support or negate the hypothesis (74). The sources include content, response process, internal structure, relationship to other variables and consequences of testing. When using this framework we use the following methodological terminology. A test is made of content that is a representation of the underlying construct. The construct or trait is what the test intends to measure; for example, psychomotor skills in laparoscopic surgery. It is not the test itself that is said to be valid, but the interpretation of test scores. There is evidence to support validity when the interpretation of the test score corresponds to the construct being measured (49, 68). Evidence from *validity based on content* focuses on the content of the test and whether the content relates to the construct that the test is intended to assess (49, 68). Evidence from *validity based on response process* ensure that the intended response is elicited when administering the test. Furthermore, validity evidence from the response process includes data entry and maintaining data integrity as a means of eliminating bias that affects test scores (49, 68). Evidence from *validity based on internal structure* relates to evidence gathered from statistical analysis of the test scores to ensure reproducibility. Reliability of test scores is considered a source of validity evidence from the internal structure (49, 68). Evidence from *validity based on relationship to other variables* relates to how test score correlates to other measurements of the same construct (49, 68). Evidence from *validity based on consequences of the test* relates to the consequences of testing. This is an important source of validity evidence since potentially harmful consequences should be identified (49). Consequences as a source of validity are broad and standard setting is an important part of this. The first step in analysing the consequences of testing is to set a pass/fail level. Thereafter, the consequences of the test and the pass/fail level can be explored (75). A standard setting method is used to establish a pass/fail level.

STANDARD SETTING

Standard setting describes the methodology used when setting pass/fail levels on a test. Throughout the development of educational theory, many different types of standard setting methods have been used (76). Nonetheless, when setting standards, the decision on a score remains a policy decision (77). Standard setting methods can be either norm-based or criterion-based, also referred to as relative and absolute (28, 76, 78). Norm-based methods are used when a pass/fail level is set according to the percentage of the students that will pass or fail. Criterion-based methods use a set criteria of passing; that is, having performed to a certain level. A criterion can be set on both surgical assessment or according to the number of correct answers that is sufficient to pass a multiple choice test. Criterion-based tests are often used to assess competency (79). Criterion-based methods are traditionally divided into examinee-centred and test-centred methods (80). Examinee-centred methods look at the examinee, determine the ability of the students and use these observations to set a pass/fail level. Test-centred methods, on the other hand, look at test characteristics, such as difficulty and relevance, and set a pass/fail level according to these. Two common methods of standard setting are the contrasting groups method (an examinee-centred method) and the Angoff method (a test-centred method). A new method of criterion-based standard setting is the expert performance level, where the performance of experienced participants is used to set pass/fail levels. The pass/fail level is set at the mean performance level of a group of experienced participants. Although this methodology has been poorly described in educational research literature, it is used in simulation-based training (81).

Three standard-setting methods were used in this thesis. In Paper 2, the contrasting groups method is used to set a pass/fail level for the TABLT test. In Paper 3, three different methods were used to explore the consequences of the choice standard setting methodology. The three methods were the expert performance level, contrasting groups method and the Angoff method. In the expert performance level method, the pass/fail level is set at the mean performance level of a group of experienced laparoscopic surgeons. The Angoff method is a criterion-based test-centred method that, in its original form (82), consisted of asking judges to reach an agreement on the definition of a borderline student. The judges agree to define a borderline student as a student with a 50 per cent chance of passing the test. After reaching agreement on this definition, the judges would determine the performance level of a borderline student. The performance level would be described in passing percentages on each item in a test. The items scores were then averaged across different judges and a pass/fail level was set (79, 80). The Angoff method has since been modified in different ways, and sometimes includes performance data. These data are presented to the judges, who then determine passing levels through several iterations (28). There is considerable empirical evidence to support the use of Angoff method (28, 83). However, this method can provide a high pass/fail level as experts often expect too much of their students (77, 80). The Contrasting Groups method, on the other hand, is a criterion-based examinee-centred method (80). The pass/fail level is set based on a division of participants as either competent or not competent. Test scores from the participants are used and a pass/fail level is set at the intersection between the distribution of the two groups (84). The pass/fail level can be moved according to the purpose of the test, either to ensure that no competent student is failed or that no incompetent student will pass (79).

FOCUS GROUPS AND INDIVIDUAL INTERVIEWS

Interviewing participants is a well-established method of investigation in qualitative research. Interviews can help shed light on complex problems (85). Interviews have been shown to be a particularly good methodology when exploring behaviours and experience (86). Using interviews can provide researchers with answers to 'how' and 'why' questions (87). Interviews can be analysed in different ways depending on the theoretical foundation. One method of analysis is content analysis (88), which we used in this thesis.

Focus groups are increasingly being used in healthcare research (89, 90). A focus group interview can help researchers gain insight into norms of behaviour through group interaction (90). In focus groups, relations between participants can be used to help obtain information from individuals who would not otherwise participate in an interview (91). Individual interviews, on the other hand, can help create a deeper understanding of each individual experience. Semi-structured individual interviews are often conducted based on a guide with open-ended questions (86). Individual interview are sometimes the only source of data in qualitative research projects (85). In the present thesis, we used both focus group interviews and individual interviews. Both methodologies were used in the fifth study to help create an understanding of the use of BTs in take-home training.

STATISTICAL CONSIDERATIONS

A variety of statistical methods have been used in the studies included in this thesis. Various descriptive statistics were used, as appropriate. Descriptive statistics were also presented in tables and illustrated using graphs. In the second study, reliability was measured by calculating the intraclass correlation coefficient (ICC). We used the ICC definition of single measures and absolute agreement. Pearson's correlation was used to analyse correlation. We calculated the Pearson's product moment correlation coefficient and a Pearson's r value of >0.7 was considered an acceptable degree of correlation (92). Analysis of variances (ANOVA) was used to analyse differences between groups, and a Bonferroni correction was used to allow for comparisons of more than two groups. Independent samples t-test was used to compare difference of means between groups in Papers 3 and 4. A p-value of less than 0.05 was considered statistically significant. Statistical software (SPSS, vs. 20.0 Chicago IL, USA) was used for analysis.

ETHICAL CONSIDERATIONS

All of the studies using data from participants were reported to the regional ethics committee. In accordance with Danish legislation, the regional ethics committee deemed that no approval was necessary for these studies. The studies were also reported to the Danish data protection agency and approval was given prior to commencing. The randomised controlled trial was registered at www.clinicaltrials.gov to ensure transparency. All participants were informed in writing and verbally, written consent was given before participating, and participants were informed of their rights, including their rights to withdraw their consent at any point during the studies. For the interviews, participants were anonymised on transcription of the interviews. Ebbe Thinggaard was the only researcher who had access to non-anonymised recordings.

OVERALL PURPOSE OF THE THESIS

The overall purposes of thesis were to review the current knowledge of training off-site, to develop and explore validity for a training and assessment system for off-site training, to investigate the effect of take-home training in simulation-based laparoscopy programmes, and to explore the use of take-home training.

AIMS

From the overall purpose of this thesis, the following aims for the studies were developed.

Study 1: Create an overview of the current knowledge of the use of BTs and the instructional designs used in off-site training programmes.

Study 2: Explore evidence of validity for a test of basic laparoscopy skills for training at home and establish a reasonable pass/fail standard.

Study 3: Determine the consequences of different standard setting methods and explore what level of competency is perceived to be adequate to begin performing supervised surgery.

Study 4: Investigate the added effects of training at home in a simulation-based laparoscopic training programme and explore the reliability of self-rating when training unsupervised.

Study 5: Describe how surgical trainees use mobile boxtrainers when training at home and explore the use of self-rating in unsupervised training.

HYPOTHESES

The hypotheses for the papers included in the thesis were as follows.

Study 1: Simulation-based laparoscopic training programmes for off-site training on BTs use instructional designs based on educational theory.

Study 2: Evidence of validity for a test in basic laparoscopy skills for training at home including a credible pass-fail score can be established.

Study 3: Different standard setting methods affect the consequences of a test in basic laparoscopic skills and there is a difference in the level of competency, which is perceived to be adequate to begin performing supervised surgery.

Study 4: Training laparoscopy at home has a positive effect on training and participants are able to rate their own performance using a structured self-rating system.

Study 5: Surgical trainees use mobile boxtrainers when training at home and are able to use self-rating to guide unsupervised training.

PRESENTATION OF THE INCLUDED PAPERS

Paper 1:
Thinggaard E, Kleif J, Bjerrum F, Strandbygaard J, Gögenur I, Ritter E M, Konge L. Off-site training of laparoscopic skills, a scoping review using a thematic analysis. Surgical Endoscopy 2016;11:1-9.

Background

Simulation training is becoming a valuable addition to laparoscopic skills training in the clinical setting. Laparoscopic simulation-based training is being implemented using VRSs and simple BTs. However, barriers to simulation training still pose challenges to implementation. These barriers include duty-hour restrictions, the high price of simulation equipment and the opening hours at skills labs and simulation centres. Mobile BTs may help overcome barriers to simulation training and provide trainees with the opportunity to train when they have the time. Although portable BTs may help remove the barriers associated with simulation-based laparoscopy, our knowledge of this area remains limited.

Objective

The objective of the study was to create an overview of the current knowledge on off-site training in laparoscopy.

Methods

Based on the research question and objective of the study, a search string was created. The search string was adapted and used in various online databases including: MEDLINE, ERIC, PsychINFO and Scopus. A snowballing search was also conducted and relevant websites and references from reviews on laparoscopic skills training were used to identify records. A consensus was reached regarding which records to include in the review. Two independent researchers working in collaboration did the screening. We analysed the records iteratively using a thematic analysis approach.

Results

We identified and included 22 records. Based on a thematic analysis, the following underlying themes were identified: access to training, protected training time, distribution of training, goal-setting and testing, test design, and unsupervised training. The underlying educational theories we identified included: proficiency-based learning, deliberate practice, and self-regulated learning.

Conclusions

A variety of instructional designs are used in laparoscopic skills training programmes for training on simple BTs. Instructional designs are based on different educational theories including proficiency-based learning, deliberate practice and self-regulated learning. Directed self-regulated learning could prove valuable when designing laparoscopic off-site training.

Strengths and limitations

The search identified 1978 studies, from which 22 records were included. The records varied in terms of where and how they were published. The records were published either as conference papers or articles in peer-reviewed journals. The quality and methodology of studies also varied. However, following the methodology described for scoping reviews, we did not assess the quality of the studies as each study was considered equally important (93). All records were assumed to provide findings that would help generate an understanding on the use of off-site training. The inclusion of only 22 records from a search identifying 1978 records demonstrates the broad scope of the search. We aimed to identify any source of literature that could be relevant to simulation training at home on BTs. The search was systematic in nature. A strength of our study was that we used a well-established reporting format: the STructured apprOach of the Reporting In health care evaluation of Evidence Synthesis (STORIES) (94). For the analysis we used a thematic analysis approach (95).

The analytical approach shows the difference between a scoping review methodology and traditional systematic reviews used for randomised controlled trials (96). Although our study included a wide array of records and sources, our search could have been even broader. We could have included websites from relevant manufacturers of equipment for laparoscopic skills training, as well as abstract books available online from relevant conferences, websites from educational institutes and other stakeholders in surgical laparoscopic training. When deciding how broadly to search, there is always the limit of feasibility, as a broader search would have taken more time and may not have provided a better result. The records we included from websites and conference abstract books did not impact our analysis to a very high degree as these sources provide very brief descriptions of studies.

We used a thematic analysis for analysing the records (95). A benefit of thematic analysis is that it uses an open approach to existing literature and thereby allows themes to emerge from the literature. However, limitations of this approach include researcher bias and its effect on findings. Our understanding of educational theory could have influenced our findings. All of the members of our research group were medical doctors working in surgical specialties. However, all but one of the researchers had conducted research within educational sciences and we were all familiar with a wide variety of educational theory, especially theory used in surgical training. Being familiar with the field of study helped ensure that the thematic analysis could be carried out appropriately. Familiarity with the reviewing process and educational theory created trustworthiness of our findings (97, 98).

From our analysis, we identified proficiency-based learning, deliberate practice and self regulated learning as underlying educational theories guiding laparoscopic training programmes for off-site training. Based on our findings, we recommended that training at home in laparoscopy should be guided by sound educational principles and that the use of directed self-regulated learning should be explored further.

Paper 2:

Thinggaard E, Bjerrum F, Strandbygaard J, Gögenur I, Konge L. Validity of a cross-specialty test in basic laparoscopic techniques (TABLT). British Journal of Surgery 2015;102:1106-1113.

Background
Acquiring laparoscopic techniques by training on simple BTs has shown to be an effective instructional method. Training on BTs has shown to affect both patient outcomes and improve performance on assessments. However, current training programmes have been developed for specific specialties and include advanced training such as suturing. For training to be relevant for novice laparoscopic surgeons, a new training and assessment tool is needed. The TABLT is a training tool developed for trainees to acquire the necessary laparoscopic techniques used during supervised surgery in clinical practice.

Objectives
The objectives of this study were to develop and explore evidence of validity for the TABLT test and to set a credible pass/fail level.

Methods
The TABLT test was developed in separate training programmes for basic laparoscopic skills training in surgery and gynaecology. From these training programmes, the need for a cross-specialty test emerged. The TABLT test was developed with a specific focus on construct alignment of tasks to facilitate transfer of skills into the clinical setting. To explore the validity of the TABLT test, we used the contemporary framework of validity known as the unitary framework of validity. The unitary framework of validity relies on five sources of validity evidence: content, response process, internal structure, relations to other variables, and consequences. We also established a credible pass/fail level using the contrasting groups' method.

Results
We included sixty participants in the study. Participants were doctors from surgery, gynaecology, and urology departments. Novice, intermediate and experienced laparoscopic surgeons were recruited. Novices were defined as doctors who had never performed a laparoscopic procedure. Intermediates were defined as doctors who had performed between one and 100 procedures and experienced laparoscopic surgeons were defined as doctors who had done more than 100 procedures. From the content source of validity we found that the test content was appropriate for cross-specialty training and included a sufficient range of relevant tasks for laparoscopic skills training. From the process of scoring we found that scoring was done easily using an Excel spreadsheet and that scoring was transparent and data integrity maintained. Test scores correlated well with procedural experience, Pearson's r value was 0.73 ($p < 0.001$). We found a significant difference between groups of different levels of experience ($p < 0.001$), and scores were reliable; ICC (0.99, $p < 0.001$). These measures provided evidence validity from the internal structure of the test. As part of exploring evidence of consequences, a defensible pass/fail level was set at 358 points. At this pass/fail level, 10 per cent of novices passed the test and 10 per cent of experienced laparoscopic surgeons failed the test.

Conclusions
We found evidence of validity from five sources: content, response process, internal structure, relations to other variables and consequences. We established a credible pass/fail level for surgical trainees to reach prior to performing supervised surgery in the OR.

Strengths and limitations
A major strength of our study was that we used the unitary framework of validity (99). This is a contemporary framework recommended in the literature to establish evidence of validity for testing in surgical training (73). Previous frameworks relied on types of validity, whether it was content, construct and criterion validity, or even face validity. Face validity contributes very little to the understanding of the validity of a test (100). Types of validity only offer a limited understanding of validity and provide insufficient sources of evidence for validity in testing. In our study, we could also have chosen to use a newer framework of validity, the framework of the use-argument validity proposed by Kane (77). The use argument framework of validity has been proposed for assessment in medical education (69). The choice of methodology was based on wishing to use both a contemporary approach and one that has been used in the literature to establish validity evidence for assessments in surgical training.

We included doctors from three specialties and with different levels of experience. We included 60 participants in total. The sample size could be considered small when considering the doctors were recruited from three departments and divided into three groups. However the TABLT training was developed for cross-specialty training, which made it necessary to include doctors from different specialties. The inclusion criteria were the same for all specialities and the trait we wished to measure (basic laparoscopic technical skills) was similar across the different specialties. Using stringent inclusion criteria helped reduce the source of bias from having a wide range of doctors participating in the study from different specialties. Having three groups of doctors also demonstrated the difference of performance in different groups. Our study included a very wide range of performance and experience levels, particularly in the group of experienced surgeons. Nonetheless, across three groups we were still able establish a firm correlation between clinical experience, measured as the number of procedures, and performance scores on the TABLT test.

When establishing validity evidence for testing, there are two sources of bias that should be considered: the risk of construct under-representation and the risk of construct irrelevant variance (101). Looking at evidence of validity from content helps explore the bias associated with construct underrepresentation. In our study we found that the content of the TABLT test was appropriate and represented the domain of laparoscopic skills. Nonetheless, our inclusion of tasks could have been even broader. Also, the tasks could have been more advanced and included content such as suturing or camera navigation; however, suturing is not considered a simple task and would not be relevant to include in a basic laparoscopic skills test. It would also have been unreasonable to require novice laparoscopic surgeons to be able to suture prior to supervised surgery in clinical practice. Camera navigation is a relatively simple task that can easily be learned when assisting in surgery.

Bias from construct irrelevant variance can be explored both from the process and the internal structure. From the process, as a source of evidence of validity, we looked at the process of rating. The TABLT rating system was developed as a simple rating system that is easy to comprehend and record. Rating can be done using a simple spreadsheet, which allows for transparency in record keeping. Transparent recording helps maintain data integrity. To minimise the impact of different settings, all tests were done in only two places, by the same researcher. The two settings were the simulation centre and a hospital at which participants were working. Using a setting that was familiar to participants helped minimise the risk of the setting affecting the performance of the surgeons. The tests were rated on-site by the researcher and afterwards by a blinded researcher using video recordings. The blinded rater was one of the researchers from the research group, which could have affected ratings. A surgeon outside the research group could have done the ratings, but the rater would have required further training. Rater training for the TABLT test was not explored in the study.

Participants were only asked to make two attempts at the test. The first one to allow participants to familiarise themselves with the system and the second attempt was used for rating and was video-recorded. Having participants do more attempts could have created a learning curve and informed us about the learning potential of TABLT test. However, the focus of this study was on the validity of the testing aspect of the TABLT. Two attempts were considered sufficient to provide us with the data we needed in order to establish validity for the TABL test. As a part of establishing validity from consequences, a pass/fail level was set. The pass/fail can provide an understanding of the immediate consequences. The immediate consequences of the pass/fail level are the fail rates for experienced surgeons and passing rates for novices. Validity from consequences was analysed in this way. However, pass and fail rates only explain some of the consequences of testing.

Paper 3:
Thinggaard E, Bjerrum F, Strandbygaard J, Gögenur I, Konge L. Ensuring competency of novice laparoscopic surgeons – Exploring standard setting methods and their consequences. Journal of Surgical Education 2016;73:986-991.

Background
Proficiency-based simulation training in laparoscopy is growing and testing is an important part of this approach to training. In proficiency-based training, testing is used to set the minimum requirement for competency. The pass/fail level determines the minimum requirement; however, the method used to set the pass/fail has not been explored in great detail. The effect of choice of methodology on pass/fail levels should be investigated more thoroughly, as should the perceived adequacy of the pass/fail levels.

Objectives
The objectives of this study were to explore the effect of the standard setting method on the pass/fail level and to investigate whether there was a difference in the level of competency, experienced and novice laparoscopic surgeons expected a novice to reach on the TABLT test during training.

Methods
Participants were included in a validation study of the TABLT test. Participants were novice and experienced laparoscopic surgeons from surgery, gynaecology and urology departments. Each participant was asked to make two attempts on the TABLT test; the second attempt was recorded and rated on-site. The second attempt was also video-recorded and rated by a blinded rater. Three standard setting methods were used to set a pass/fail level: the expert performance level, the contrasting groups method and the Angoff method. After testing, participants were asked how high a score a novice should reach during training, how many errors were acceptable and how much time they should be allowed to spend on each task.

Results
The study included 40 participants, 20 experienced laparoscopic surgeons and 20 novices. The three standard setting methods resulted in three different pass/fail levels. The expert performance levels set the pass/fail level at 474 points, the contrasting groups method at 358 points and the Angoff method amongst experienced at 311 points and amongst novices at 386 points. The consequences of the different pass feel levels varied. The fail rate for experienced surgeons was between 0 and 50 per cent and the pass rate for novices was between 0 and 25 per cent. There was a significant difference in the level of proficiency deemed adequate by experienced and novices ($p < 0.008$). Novice laparoscopic surgeons expected novices to reach a higher score on a test during training than experienced laparoscopic surgeons did.

Conclusions
The pass/fail level of a basic test in laparoscopic skills varies highly depending on which standard-setting method is used. Experienced and novice laparoscopic surgeons have different expectations . Novices expect that other novices will be able to reach a higher test score during training than experts do.

Strengths and limitations
The focus of this study was on the choice of standard-setting method and its consequences. We examined this by using three different methods to set a pass/fail level. We calculated the fail rates amongst experienced and pass rates among novice laparoscopic surgeons to explore the consequences of the three different standard-setting methods. The aim of the study was selected on the basis of the fact that few studies have been done within surgical training on the effects of standard setting on the pass/fail level, and even fewer on the consequences. Standard setting has been explored outside the field of surgical skills training. In the literature of medical education, studies describe that the pass/level depends on the choice of standard setting method (28). These findings are in accordance with findings from our study, with the fail rate amongst experienced surgeons varying from 0 to 50 per cent. Our result illustrates a high variance in the fail rate, which depended on the choice of standard-setting method.

Choosing the right pass/fail level depends on the purpose of the test. Tests used when there are a limited number of spaces, such as an entry-test for a surgical training programme, have a different purpose than tests used for assessment during training. Testing used for formative or summative feedback should use an examinee-centred standard-setting method (28). To explore what the pass/fail should be, one must consider the level at which performance is considered sufficient. In proficiency-based training, the focus is on a high pass/fail level as trainees are expected to be not only competent but also proficient. Setting new pass/fail levels for the TABLT would require the test to correspond to a different level of training in clinical practice. A new proficiency level could be used when trainees begin to do more advanced surgery, such as when they move from partial procedures to full procedures. In this way, different pass/fail levels on a test could be explored so that the progression in clinical training corresponds to the progression in simulation training. This would ensure a continuous use of simulation training and help trainees prepare for each step in their continued progression in the clinical setting.

Paper 4:

Thinggaard E, Bjerrum F, Strandbygaard J, Konge L, Gögenur I. Take-home training facilitates distributed training, a randomized trial. Surgical Endoscopy 2016; Submitted.

Background

Acquiring surgical skills during simulation training is becoming a valuable addition to clinical training in the OR. Laparoscopic techniques can be acquired in simulation-based training using simple box trainers, which enable training at home. Take-home training helps trainees overcome barriers and makes training accessible. Although the feasibility of take-home training has been investigated, the effect of training at home has not been explored sufficiently. It is necessary to explore the effects of take-home training. Training at home also presents the challenge of missing feedback when training unsupervised. Rating their own performance may provide participants with a type of feedback, but the reliability of self-rating in laparoscopic training has not been thoroughly investigated.

Objectives

The objectives of this study were to investigate the added benefit of training at home in a simulation-based training programme and to explore if trainees were able to rate their own performance on the TABLT test.

Methods

We designed and conducted a randomised controlled trial with a blinded rater. Participants were doctors in their first year of training. Participants were recruited during a basic laparoscopic skills programme in which VRSs and BTs are used. After inclusion, participants were randomised to either having access to take-home training or following the regular course with only on-site training. Participants used logbooks to record their training. Training patterns were measured, including the time to complete the training course, the time spent on training and the number of training sessions.

Results

Thirty-six participants were included in the study; 18 in each arm. We found a significant difference in the number of training sessions (5.8 versus 2.3, $p < 0.001$). We found no difference in the time to complete the training programme (86 vs. 89 days, $p = 0.89$), the time spent on training (302 vs. 218 minutes, $p = 0.26$) or the performance score (493 vs. 460, $p = 0.07$). Participants were able to rate their own performance using video recordings. Self-rating provided reliable ratings correlating well to those of a blinded rater; ICC 0.86, $p < 0.001$.

Conclusions

When free to choose, participants choose to train in a distributed manner. They divide their training into shorter sessions and they train more frequently. They do not take longer to complete a training course and do not spend more time on training. Participants in a laparoscopic skills course are able to rate their own performance using a simple scoring system. Ratings from self-ratings are reliable.

Strengths and limitations

This study aimed to investigate the added benefit of training at home in a simulation-based laparoscopic training course. When designing the study we measured three equivalent outcomes: time to complete the training course, the time spent training, and the number of training sessions. For the sample size calculation we chose to use time to complete the training course. We anticipated that training at home would affect training patterns and make it possible for trainees to pass a test earlier. Training at home would reduce the length of the training course and allow trainees to start performing supervised surgery earlier. We also anticipated that participants would reach a higher proficiency level as they had more access to training. However, the length of the training course was determined by the mandatory elements of the course. The course started with an introduction course and ended with the operative course (63). Therefore, the length of the training remained unaffected. We also found no significant statistical difference in the level of proficiency. Although there was a tendency for the scores of participants training at home to be higher, we were unable to conclude that this was due to the access provided in take-home training. Our finding corresponds well with the fact that participants were able to rate their own performance. Self-rating helped participants reach the required pass/fail level, but also acted as a minimum requirement for trainees to reach.

One limitation of our study design is the sample size calculation. For a randomised controlled clinical trial, only one measurement can be chosen for the sample size calculation. At the end of the study inclusion period of one year, we did a new sample size calculation to evaluate the feasibility of continuing the trial. The new sample size calculation showed us that more than 11,000 participants would be required to measure the difference using time to complete the course. Based on this calculation it was not feasible to continue the trial. Furthermore, the trial had already yielded significant results about what had been chosen as secondary outcomes. These significant results in combination provided information on the distribution of training patterns amongst trainees. We designed the study as a randomised controlled trial set in an actual simulation-based laparoscopy course. As we conducted the trial in a real setting, we were unable to control all variables. Thus, we pushed the boundaries of what is possible to investigate using a randomised controlled trial design. Despite the challenges with variables such as the on-going clinical training and the use of both VRSs and BTs, we still found a significant difference when comparing the two groups. Using a real laparoscopic course setting has helped us gain information on the distribution and training patterns chosen by participants, and has also provided valuable information about how home training affects participant behaviour in a real laparoscopic skills course. Trainees divided their training into shorter and more frequent intervals. Participants opted for distributed training when training at home. However, these results do not address how and why the participants trained. To further explore these questions, a study with a different choice of methodology would be needed. 'How', 'why' and 'what' questions are in focus in medical education (94), but these questions were outside the scope of our study.

Paper 5:

Thinggaard E, Konge L, Bjerrum F, Strandbygaard J, Gögenur I, Spanager L. Take-home training in a simulation-based laparoscopy course. Surgical Endoscopy 2016;1-8.

Background
It has been established that simulation-based training is effective for acquiring laparoscopic skills. Nevertheless, implementation of simulation-based training, particularly implementation of training that is accessible to trainees, has been slow to spread. Training at home has been shown to be feasible, but there has been little research into how and why trainees train as they do when training at home. Even fewer studies have addressed the need of feedback in unsupervised training and the role of self-rating as a mean of providing guidance in unsupervised training.

Objectives
The objectives of this mixed-methods study were to describe the used of BTs when training at home and to explore the training patterns, and how participants trained and what they trained on.

Methods
The study was designed as a mixed-methods study in which we incorporated methodologies from both quantitative research traditions and qualitative research traditions. The study consisted of two equally important parts: a descriptive quantitative analysis followed by a qualitative analysis of data from interviews. The participants were recruited amongst doctors on a basic laparoscopic skills course. All participants were offered training at home on mobile BTs and training at a simulation centre on VRSs and BTs. Participants were given logbooks to record when and how they trained. Data from logbooks were analysed using descriptive statistics and graphs. Focus groups and individual interviews were held and analysed by researchers following a content analysis methodology.

Results
We included 18 participants in the study. From the quantitative analysis we found that participants used an individualised approach to training. This finding was supported in the qualitative data, where participants described that they trained the task they wanted to train in the sequence they deemed most appropriate. Participants also mixed their training methods by training on both BTs and VRSs. They also mixed their use of training settings by training both at home and at the simulation centre. Participants integrated the possibility of training in a specific location at the simulation centre with the flexibility of training at home. Findings from the quantitative data analysis were used to direct the qualitative analysis. The qualitative analysis was conducted on data from focus groups and individual interviews. We identified the following themes: training method, training pattern, feedback and self-regulation.

By way of conclusions, we found that training patterns varied and trainees used an individualised approach. Mandatory training strongly affected training patterns and testing provided this structure. We also found that self-rating helped guide participants during unsupervised training.

Strengths and limitations
This study was designed as a mixed-methods study. We chose to use methods from quantitative research traditions as well as qualitative research traditions. The mixed-methods methodology has been used in healthcare-related research as well as medical educational research (102). The quantitative part of our research was a descriptive statistical analysis, while the qualitative part of the study was an analysis of focus group and individual interviews. Selecting a mixed methodology enabled us to consider the weaknesses and strength of the study from both a quantitative approach and qualitative approach.

With regard to quantitative research, traditional sources of bias include internal validity, external validity, generalisability, reliability, and objectivity (97). The sample size in our study only consisted of 18 participants, which is a small sample size. The small sample size affects the external validity of the study and makes it more difficult to generalise the findings to another population. Despite the small sample size, we were able to include a high variety of participants. Participants were both female and male, and were doctors from three different specialties, working at hospitals in various distances to the simulation centre. Participants were primarily novices and had different levels of experience with laparoscopic surgery. Having a variety of participants helps generalise the findings to other settings. The study was a descriptive study and we used descriptive statistics. This type of study is explorative and cannot provide confirmation of observations. A blinded randomised controlled study would have been better suited in order to establish confirmatory findings. However, the aim of the study was to explore the use of take-home training and the methodology, so the sample was appropriately aligned with these aims. Bias related to objectivity included considering the role of the researcher in our project. The primary researcher had different roles, both as a faculty member present at the final test and also as the facilitator of the interviews. Therefore, the researchers' role was a source of bias that may have affected our findings. Within the qualitative research traditions, however, bias is treated in terms of trustworthiness (97, 98). To help provide transferability, it is recommended that researchers have an understanding of the context and that there is a sufficient length and number of data collection and collection sessions (98). The qualifications of the research group were appropriate. The research group behind the study consisted of researchers who were familiar with the use of qualitative research methodology, as well as researchers who were well versed in research in surgical skills training. This helped improve and ensure the transferability of knowledge generated from this study. Another source of evidence of trustworthiness is transferability, which includes reflective practice and having a transparent audit trail (98). Data collection was documented thoroughly to ensure an appropriate audit trail. To help improve the quality of the study we used different qualitative methodologies, both focus group interviews and individual interviews. Different sources of data helped ensure sufficient data for data triangulation and saturation, which is an important source of trustworthiness (97, 98). Iterative questioning was used during focus group and individual interviews, which improved the confirmability and credibility. An iterative questioning also ensured that meaning was explicit. Based on these control mechanisms, which are part of the methodological framework in quantitative research, there is ample evidence of trustworthiness to support the findings from our qualitative data analysis.

DISCUSSION

The overall purposes of the thesis were to review the current knowledge on training off-site in laparoscopy, develop and explore validity for a training and assessment system for off-site training, to investigate the effect of take-home training in a simulation-based laparoscopic training programme, and to explore the use of take-home training. We have found (1) that training programmes for off-site training are based on a variety of instructional designs and educational theory, (2) that the TABLT test is supported by evidence of validity, (3) that the choice of standard setting method affects the pass/fail level of a test, (4) that trainees use an individualised and distributed approach to training when training at home, (5) that training requirements and testing are determinants of training patterns, and (6) that self-rating is reliable and helps provide guidance when training unsupervised. These findings are discussed below.

Educational Theory and Take-Home Training

In the first paper in this thesis we conducted a review of the literature on off-site training and found that a variety of educational theories and instruction designs are used in off-site laparoscopic training programmes. However, the educational theories or conceptual frameworks were rarely stated. It has been recommended that the conceptual framework is stated clearly and described when conducting research in medical education, although this is often lacking in published literature (103). For research in medical education, the theoretical framework is an essential part of establishing evidence in support of findings (44). Furthermore, it is recommended that educational theory is used as a foundation for successful curricular designs (26, 104). In our review we found that educational theory used in training programmes for off-site training included deliberate practice, proficiency-based learning and self-regulated learning. Extant literature has recommended deliberate practice for acquisition of technical skills. Deliberate practice relies on the use of well-defined tasks, distributed learning and immediate feedback (57). In off-site training it is a challenge to implement immediate feedback, as faculty is not readily available. In Paper 4 and Paper 5 in this thesis we asked trainees to use self-rating as a method of getting feedback. In Paper 4, self-rating was shown to be reliable and in Paper 5 we found that self-rating was seen as a useful source of feedback that helped guide unsupervised training. We also found that trainees used a distributed training pattern and that there was a significant difference in the training pattern. Participants training at home trained more often and in shorter intervals than participants who trained only at the simulation centre. This finding was confirmed in Paper 5 and participants described how they used an individualised approach by distributing training according to their schedule. Distributed training has shown to be effective and to improve retention (52). Another way to improve retention is through testing. In Paper 5 we found that testing and mandatory training requirements determine training patterns. Furthermore, that self-rating was seen as a help in self-regulating in unsupervised training. In the review, we concluded that principles from the DSRL could help guide curricular design for simulation-based laparoscopic training off-site. DSRL could be used to help guide an independent learning approach in unsupervised training. We did this by using self-rating to support the structural framework in Papers 4 and 5 (59, 60). We found that self-rating provided the needed structure to guide unsupervised training. DSRL is an emerging educational framework and its use remains to be shown in laparoscopic skills training. Recently, DSRL has been used in combination with principles of mastery learning and has been shown to be cost-effective (105). Mastery learning has recently received renewed attention in medical education literature and has been recommended (106-108). In laparoscopy, one study showed that mastery learning was effective when implemented in a curriculum for laparoscopic inguinal hernia repair (20). Laparoscopic training programmes have often been based on proficiency-based training that requires a higher level of performance from trainees than competency-based training (27). In mastery learning, however, participants are expected to reach consecutive levels of mastery during training. Mastery learning could be used in take-home training in laparoscopy. However, mastery learning requires high-quality assessment supported by evidence of validity in order to be effective (106). To implement principles of mastery learning it is necessary to develop a series of assessments. Also, mastery learning requires that current standard setting methods for testing are modified (109). The need to modify standard setting methods shows the importance of using educational theory for curricular designs. For a off-site training curricula to be effective it should be based on sounds educational theory so that content, assessments and clinical practice are aligned appropriately.

Consequences of Testing and Validity

Testing and setting a pass/fail level can be used to help trainees prepare for clinical practice. Testing is not only important in mastery learning, but is also a part of proficiency-based training and also used in summative feedback (77). In Paper 3 we explored the immediate consequences of testing and the choice of standard setting method. We found that different standard setting methods resulted in different pass/fail levels. This finding was supported in the literature on standard setting in medical education research (28). When implementing tests it is important that they are supported by a wide range of validity evidence (49, 72). The analysis of consequences of tests and their pass/fail levels is a source of evidence for validity in the unitary framework of validity (99). In Paper 2 we found evidence in support of validity for the TABLT test including evidence from the source of consequences of testing. Nonetheless, the analysis of the immediate consequences of the pass/fail level is only part of the exploration of validity (75). Validity from the consequences of a test can include consequences to the trainees, faculty, patients and society in general. Consequences to the trainee are particularly important for a basic laparoscopic skills test used prior to supervised surgery of patients. Positive consequences of passing include a positive learning experience and increased confidence. However, if the pass/fail level is set too low, confidence may turn into overconfidence as participants have not acquired the necessary level of competency. Failing has both positive and negative consequences. Negative consequences can include a lack of progress for trainees and trainees having to face the fact that they are not yet proficient. However, these consequences are outweighed by the need for patient safety. However, failing can also help participants to reflect on their level of competency and identify areas that need improvement. To address these consequences of testing, a summative test could be accompanied by formative feedback. The consequences of testing are an important part of establishing evidence of validity and are the focus in the use/argument framework of validity proposed by Kane (71). The use/argument approach validity framework highlights the decisions and the inferences made from test scores (69). In this framework it is important to clarify what inferences and which decisions will be made from test scores (71). Inferences from test scores could include insight into the proficiency level in laparoscopic surgery and decisions could include an evaluation of a trainee's readiness to perform unsupervised surgery in a clinical setting. The use/argument approach relies on an exploration of scoring, generalising, exploration and implications (71). The use/argument has been used recently to explore validity for the Objective Structured Assessment of Technical Skills in current literature (110). In both the unitary framework of validity and the use/argument approach, validity is seen as a continuous process. Validity is seen as a process in which evidence is gathered to support validity (70). In the second paper in this thesis we used the unitary framework of validity. However, the TABLT test has since been implemented in a cross-specialty course. After the implementation, new consequences have emerged and the use/argument approach could be used to further explore validity evidence for the TABLT test. The focus on the analysis of consequences for testing demonstrates the importance of the effect of testing on training.

Training Patterns and Distributed Learning.

When participants trained without supervision, we found that they adopted an individualised and distributed training pattern. Distributed practice is an improvement on massed practice and has been recommended for laparoscopic skills training (27, 104). Distribution of training improves retention (52, 53), although the optimal frequency of training has not been determined (56). In Paper 4 we found that participants use a distributed training pattern. This was supported in Paper 5. The descriptive statistical analysis revealed a tri-phasic training pattern. The first phase was one of moderate training intensity, the second phase was a period of low-intensity training and the third phase was one of high-intensity training. We were able to triangulate these findings using quantitative and qualitative data analysis. Training patterns were discussed in detail during the focus group interviews and individual interviews. In the qualitative analysis we looked at why this training pattern emerged. In the first part of the tri-phasic training pattern, participants had to overcome the initial challenge of setting up the system. They overcame this challenge by being motivated by having access to training, which also helped them adopt a moderate level of training intensity. However, training intensity was reduced during the second phase of the training, which could be seen as a lack of internal motivation. Due to the requirements in clinical work placing and other time constraints, internal motivation was insufficient for maintaining a moderate training intensity. In the third phase of training, participants had a high level of training intensity. This high level of training intensity was described as being associated with the approaching test deadline. Participants were freely able to organise their testing sessions and could come for testing whenever they pleased. Nonetheless, 80 per cent of participants decided to go for testing within a week of the operative course. This finding suggests that deadlines posed by testing or other mandatory training elements provided external motivation to train. Training requirements ensure that participant train to reach a sufficient level of competency.

Testing and mandatory training requirements

Testing is said to be a driver of learning (111) and is an important part of competency-based medical education, particularly proficiency-based training. Although testing is a driver of learning, learning can vary from trainee to trainee and depends on several other factors, such as internal motivation (112). Despite this variation, testing has shown to improve learning outcomes (47). Testing in laparoscopy has been introduced as part of proficiency-based training (27) and the effects of proficiency-based training have been demonstrated (29). In Paper 1 we found that many instructional designs for off-site training included testing. Furthermore, it has been recommended to implement testing in off-site training programs (113). In Paper 5 we found that testing and mandatory training are determinants of training patterns. Participants trained according to their needs and external requirements such as mandatory testing. From the logbooks and interviews we found that participants trained in a tri-phasic training pattern, and the highest training intensity was seen when the deadline to pass a test approached. Although many other sources of motivation exist, our study showed that testing can help ensure that trainees reach a necessary level of technical competency. Implementing testing in a off-site simulation-based training program can help trainees prepare for supervised surgery on patients. Testing will hereby ensure that both trainees and patients benefit.

Unsupervised Training and Self-Rating

Take-home training is challenged by the fact that feedback from faculty is not readily available in unsupervised training. As an alternative, we asked the participants to rate their performance. In Paper 5 we explored the role of self-rating. Feedback and self-rating was discussed in the interviews and self-rating was mentioned as a source of feedback. Self-rating was described as being useful especially when preparing for testing. We believe that self-rating is perceived in this way because the testing system is relatively simple but still relevant for novice laparoscopic surgeons. It is easy for participants to interpret the scores as they are based on the number of errors and time. While some may criticise this as a simplification of the skills needed in laparoscopic surgery, the TABLT test was developed for basic technical skills. Simple numeric feedback based can provide useful feedback when training to pass a test (114). However, narrative feedback has shown to be effective for complex skills training (115, 116) and we suggest that narrative feedback from supervising surgeons could be implemented when simulation training is used for more advanced training. Narrative feedback will also help provide participants with information on other skills that are relevant to laparoscopic surgery, such as how to ensure proper ergonomy when working. In our study, self-rating allowed trainees the freedom to choose when to get feedback and this has shown to be effective as participants are in control of when they receive feedback (117). In Paper 4 we found that participants saw self-rating as a useful method with which to guide their training and was useful for directing self-regulation. The use of DSRL for laparoscopic training programmes off-site was discussed in Paper 1 and we concluded that this could be a potentially useful educational theory for off-site training programmes. DSRL is based on SRL which is important in unsupervised training and is recommended in a recent systematic review on simulation-based training (61). DSRL is when the focus is on autonomous learning combined with the need to guide trainees in their learning experiences (60). Faculty can provide facilitation and a framework in SRL by using principles of DSRL (59). In Paper 5 we could see that the principles DSRL were in use. Self-rating was used to provide guidance in unsupervised training. Participants were able to self-regulate within a structured framework. This was described as an individualised approach that was structured by mandatory training requirements.

CONCLUSIONS

Laparoscopic surgery requires very specific skills. Basic skills can be acquired through simulation-based training outside the OR. Simulation-based training is an effective method of instruction, but barriers to simulation training exist and implementation remains to be explored. Take-home training has been suggested as a solution to overcome barriers and we have explored this topic in the present thesis. We have created an overview of the literature on off-site training, developed and explored validity for a test for off-site training, looked at the consequences of the pass/fail levels for this test, and – through a randomised controlled trial and mixed-methods study – found that off-site training allows for distribution training. Furthermore, we established that self-rating is reliable, that self-rating help guide unsupervised training, and that mandatory training requirements and testing are determinants of training patterns. From these findings we conclude that training off-site is effective when there is access to training, guidance during training and mandatory training requirements.

IMPLICATIONS AND FUTURE RESEARCH

The findings presented in this thesis offer several practical implications. The TABLT test has been implemented in a simulation-based cross-specialty laparoscopic training programme (63), the TABLT test has been used in research, to explore the use of laser visual guidance (118) and is being used in an on-going research project to investigate the feasibility of self-certification. Furthermore, the TABLT test and findings from the thesis have some potential future implications. The TABLT test could be used in current cross-specialty laparoscopic training courses and passing the test could be implemented as a prerequisite for supervised surgery on patients. Our results regarding distributed training patterns indicate that laparoscopic training courses should move away from one-day courses and boot camps, and instead use an approach that encourages distributed practice.

Our findings and the discussion thereof have shown some of the gaps in current literature and can help guide new research in this area. Off-site training still remains to be explored in more depth, especially research that aims to explore the use in settings where simulation training is not accessible to trainees. Research should look at the role off-site training as long distance courses using internet-based simulation courses. Feedback should be a focus for this type of research especially for advanced laparoscopic skills. How best to implement feedback in off-site training should be examined. Investigating off-site training is of particular importance in countries where laparoscopy is still being implemented.

A limitation in our studies is that we did not explore the transfer of skills from training on TABLT into the clinical setting. Future research should investigate this more thoroughly. Research should aim to explore the use of mastery learning and different standard-setting levels and how this affects performance. Different pass/fail levels could be established for corresponding levels of clinical advancement and the effect of continued simulation training throughout the clinical progression of trainees should be investigated. Simulation training has a role not only in preparing trainees for clinical practice but also as a supplement for technical skills training in clinical practice. Furthermore, laparoscopic training also has a potential in maintaining technical skills and this aspect of simulation-based training remains to be explored in more depth.

Simulation-based training in laparoscopy has become an indispensable part of how trainees acquire laparoscopic skills. However, for simulation-based training to be effective we need to explore its role continuously. As the implementation of laparoscopic training spreads we need to move outside single institutions and simulation centres and conduct multiple centre research on a national and international level. National and international collaboration will help gain further insight in how to optimize simulation-based training, how best to equip our trainees and most importantly how to ensure that our patients receive the best surgical care we can deliver.

SUMMARY

When laparoscopy was first introduced, skills were primarily taught using the apprenticeship model. A limitation of this method when compared to open surgery, was that it requires more time to practise and more frequent learning opportunities in clinical practice. The unique set of skills required in laparoscopy highlighted the need for new training methods that reduce the need for supervision and do not put the patient at risk. Simulation training was developed to meet this need. The overall purpose of this thesis was to explore simulation-based laparoscopic training at home. The thesis consists of five papers: a review, a validation study, a study of methodology, a randomised controlled trial and a mixed-methods study. Our aims were to review the current knowledge on training off-site, to develop and explore validity for a training and assessment system, to investigate the effect of take-home training in a simulation-based laparoscopic training programme, and to explore the use of take-home training.

The first paper in this thesis is a scoping review. The aim of the review was to explore the current knowledge on off-site laparoscopic skills training. We found that off-site training was feasible but that changes were required in order for it to become an effective method of training. Furthermore, the selected instructional design varied and training programmes were designed using a variety of educational theories. Based on our findings, we recommended that courses and training curricula should follow established education theories such as proficiency-based learning and deliberate practice. Principles of directed self-regulated learning could be used to improve off-site laparoscopic training programmes.

In the second study, we set out to develop and explore validity evidence of the TABLT test. The TABLT test was developed for basic laparoscopic skills training in a cross-specialty curriculum. We found validity evidence to support the TABLT test as a summative test in a basic laparoscopic training programme. We also established a credible pass/fail level using the contrasting groups method. We concluded that the TABTL test could be used to assess novice laparoscopic trainees across different specialties and help trainees acquire basic laparoscopic competencies prior to supervised surgery.

In the third study, we aimed to explore the consequences of the choice of standard setting method and whether there is a difference in terms of how high a score experienced and novice laparoscopic surgeons expect that novices should achieve during training. We used three different standard setting methods and found that pass/fail levels vary depending on the choice of standard setting method. We also asked experienced and novice laparoscopic surgeons how high a score they expected a novice laparoscopic surgeon should achieve on a test during training. We found a significant difference, with experienced surgeons setting a lower pass/fail level. We concluded that an

established standard setting method supported by evidence should be used when setting a pass/fail level.

In the first and second papers of this thesis, we found that off-site training is feasible and explored validity for the TABLT test. We used this knowledge in the fourth study to design a randomised controlled trial. The aim of the trial was to investigate the effect of take-home training in a simulation-based laparoscopic course. We hypothesised that training at home could help trainees plan their training according to their own schedule and thereby increase the effect of training. We found that participants had a distributed training pattern; they trained more frequently and in shorter sessions. We also found that participants were able to rate their own performance during unsupervised training and that self-rating was reliable.

The fifth and final study of the thesis was a mixed-methods study that aimed to explore the use of take-home training. To meet this aim, we recruited participants from the intervention arm in our randomised controlled trial. All participants had access to the simulation centre and were given a portable trainer to train on at home. Participants were asked to use a logbook during training. At the end of the course, they were invited to take part in focus group interviews and individual interviews. Based on data from logbooks, a descriptive statistical analysis was conducted and data from interviews were analysed using a content analysis. We found that participants took an individualised approach to training when training at home. They structured their training according to their needs and external requirements. We concluded that mandatory training requirements and testing help determine when and how much participants train. We also found that self-rating can guide unsupervised training by giving clear goals to be reached during training.
From the papers included in the thesis, we found that the literature describes training at home as a feasible method of acquiring laparoscopic skills. Nonetheless, changes to current training programmes are needed in order to make this method effective. We then developed and explored validity evidence for the TABLT test. We also established a reasonable pass/fail level and went on to explore the immediate consequences of the pass/fail level. Using our knowledge from the review, we conducted a randomised controlled trial and a mixed-method study. Based on these studies we found that training at home allows for distributed learning, that self-rating guides unsupervised training, and that mandatory training requirements and testing strongly influence training patterns. Access to training, guidance during training, and mandatory training requirements will make take-home training not just feasible but also effective.

REFERENCES

1. Spaner SJ, Warnock GL. A brief history of endoscopy, laparoscopy, and laparoscopic surgery. Journal of Laparoendoscopic & Advanced Surgical Techniques 1997;7:369-73.
2. Keus F, de Jong JA, Gooszen HG, van Laarhoven CJ. Laparoscopic versus open cholecystectomy for patients with symptomatic cholecystolithiasis. The Cochrane database of systematic reviews. 2006:CD006231.
3. Sauerland S, Jaschinski T, Neugebauer EA. Laparoscopic versus open surgery for suspected appendicitis. The Cochrane database of systematic reviews. 2010:CD001546.
4. Darzi SA, Munz Y. The impact of minimally invasive surgical techniques. Annual review of medicine. 2004;55:223-37.
5. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. American journal of surgery. 1998;175:482-7.
6. Dent TL. Training, credentialling, and granting of clinical privileges for laparoscopic general surgery. American journal of surgery. 1991;161:399-403.
7. Club. TSS. A Prospective Analysis of 1518 Laparoscopic Cholecystectomies. New England Journal of Medicine. 1991;324:1073-8.
8. Soper NJ, Fried GM. The fundamentals of laparoscopic surgery: its time has come. Bulletin of the American College of Surgeons. 2008;93:30-2.
9. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. Annals of surgery. 2013;257:586-93.
10. Zendejas B, Cook DA, Brydges R, Hamstra SJ, Bingener J. Comparative effectiveness of virtual reality versus box trainer simulators for laparoscopic surgery training: A systematic review & meta-analysis. Journal of surgical research. 2013;2(179):188.
11. Partridge R, Hughes M, Brennan P, Hennessey I. There is a worldwide shortfall of simulation platforms for minimally invasive surgery. Journal of Surgical Simulation. 2015;2:12-7.
12. Caban AM, Guido C, Silver M, Rossidis G, Sarosi G, Ben-David K. Use of collapsible box trainer as a module for Resident Education. J Soc Laparoendo. 2013;17:440-4.
13. Lau W, Leow C, Li AK. History of endoscopic and laparoscopic surgery. World journal of surgery. 1997;21(4):444-53.
14. Tanis PJ, Buskens CJ, Bemelman WA. Laparoscopy for colorectal cancer. Best practice & research Clinical gastroenterology. 2014;28(1):29-39.
15. Feldman LS, Sherman V, Fried GM. Using simulators to assess laparoscopic competence: Ready for widespread use? Surgery. 2004;135(1):28-42.
16. Gue S. Home-made videoscopic trainer for operative laparoscopic surgery. The Australian and New Zealand journal of surgery. 1995;65:820-1.
17. Coleman J, Nduka CC, Darzi A. Virtual reality and laparoscopic surgery. British Journal of Surgery. 1994;81(12):1709-11.
18. Brown W, Satava R, Rosen J. Virtual reality and surgical training: simulating the future. Minimally Invasive Therapy. 1994;3(2):81-6.
19. Satava RM. Virtual reality surgical simulator. The first steps. Surgical endoscopy. 1993;7(3):203-5.
20. Zendejas B, Cook DA, Hernndez-Irizarry R, Huebner M, Farley DR. Mastery learning simulation-based curriculum for laparoscopic TEP inguinal hernia repair. Journal of surgical education. 2012;69:208-14.
21. 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. Annals of surgery. 2002;236(4):458-63; discussion 63-4.
22. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, 3rd, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. American journal of surgery. 2007;193(6):797-804.
23. 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. British medical journal. 2009;338:1253.
24. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. The British journal of surgery. 2004;91:146-50.
25. Zevin B, Aggarwal R, Grantcharov TP. Surgical simulation in 2013: why is it still not the standard in surgical training? Journal of the American College of Surgeons. 2014;218(2):294.
26. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. The New England journal of medicine. 2006;355(25):2664-9.
27. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Annals of surgery. 2005;241(2):364-72.
28. Cusimano MD. Standard setting in medical education. Academic medicine : journal of the Association of American Medical Colleges. 1996;71(10 Suppl):S112-20.
29. 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. American journal of surgery. 2010;199(1):115-20.
30. Zendejas B, Ruparel RK, Cook DA. Validity evidence for the Fundamentals of Laparoscopic Surgery (FLS) program as an assessment tool: a systematic review. Surgical endoscopy. 2015:1-9.
31. Bahsoun AN, Malik MM, Ahmed K, El-Hage O, Jaye P, Dasgupta P. Tablet based simulation provides a new solution to accessing laparoscopic skills training. Journal of surgical education. 2013;70:161-3.
32. Arden D, Hacker MR, Jones DB, Awtrey CS. Description and Validation of the Pelv-Sim: A Training Model Designed to Improve Gynecologic Minimally Invasive Suturing Skills. Journal of minimally invasive gynecology. 2008;15:707-11.
33. Chandrasekera SK, Donohue JF, Orley D, Barber NJ, Shah N, Bishai PM, et al. Basic Laparoscopic Surgical Training: Examination of a Low-Cost Alternative Netherlands: Elsevier; 2006;6:1285-91.
34. Robinson JK, Kushner DM. Development and validation of a home-based, mirrored, gynecologic laparoscopy trainer. Journal of minimally invasive gynecology. 2006;13(2):102-7.
35. Bruynzeel H, De Bruin AFJ, Bonjer HJ, Lange JF, Hop WCJ, Ayodeji ID, et al. Desktop simulator: Key to universal training? Surgical endoscopy. 2007;21:1637-40.
36. Xiao D, Jakimowicz JJ, Albayrak A, Buzink SN, Botden SM, Goossens RH. Face, content, and construct validity of a novel portable ergonomic simulator for basic laparoscopic skills. Journal of surgical education. 2014;71(1):65-72.
37. Chung SY, Landsittel D, Chon CH, Ng CS, Fuchs GJ. Laparoscopic skills training using a webcam trainer. The Journal of urology. 2005;173:180-3.
38. Moreira-Pinto J, Silva JG, de Castro JLR, Correia-Pinto J. Five really easy steps to build a homemade low-cost simulator. Surgical innovation. 2013;20(1):95-9.
39. Blirup-Jensen D, Gogenur I, Rosenberg J. New model for training laparoscopic skills. Ugeskrift for laeger. 2005;167:3780-1.
40. Pokorny MR, McLaren SL. Inexpensive home-made laparoscopic trainer and camera. ANZ journal of surgery. 2004;74:691-3.
41. Johnston TJ, Tang B, Alijani A, Tait I, Steele RJ, Ker J, et al. Laparoscopic surgical skills are significantly improved by the use of a portable laparoscopic simulator: results of a randomized controlled trial. World journal of surgery. 2013;37:957-64.
42. Ruparel RK, Brahmbhatt RD, Dove JC, Hutchinson RC, Stauffer JA, Bowers SP, et al. "iTrainers"--novel and inexpensive alternatives to traditional laparoscopic box trainers. Urology. 2014;83(1):116-20.
43. Alfa-Wali M, Antoniou A. Eco-friendly laparoscopic home trainer. Simulation in Healthcare. 2011;6:176-9.
44. Prideaux D, Bligh J. Research in medical education: asking the right questions. Medical education. 2002;36(12):1114-5.
45. Flexner A. Medical education in the United States and Canada. Bulletin of the World Health Organization. 2002;80(7):594-602.
46. Cox M, Irby DM, Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. New England journal of medicine. 2006;355(13):1339-44.
47. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. Medical education. 2009;43(1):21-7.
48. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. Surgical endoscopy. 2003;17(6):964-7.
49. Downing SM. Validity: on meaningful interpretation of assessment data. Medical education. 2003;37(9):830-7.
50. Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. Journal of Applied Psychology. 1999;84(5):795.
51. Shea CH, Lai Q, Black C, Park J-H. Spacing practice sessions across days benefits the learning of motor skills. Human movement science. 2000;19(5):737-60.
52. Benjamin AS, Tullis J. What makes distributed practice effective? Cognitive Psychology. 2010;61(3):228-47.
53. Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT. Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychol Sci Public Interest. 2013;14(1):4-58.
54. Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training: a randomized controlled trial. Surgical endoscopy. 2002;16(6):957-61.
55. Spruit EN, Band GP, Hamming JF. Increasing efficiency of surgical training: effects of spacing practice on skill acquisition and retention in laparoscopy training. Surgical endoscopy. 2014;29(8):2235-2243
56. De Win G, Van Bruwaene S, De Ridder D, Miserez M. The optimal frequency of endoscopic skill labs for training and skill retention on suturing: A randomized controlled trial. Journal of surgical education. 2013;70(3):384-93.
57. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic medicine : journal of the Association of American Medical Colleges. 2004;79(10 Suppl):S70-81.
58. Saks K, Leijen Ä. Distinguishing Self-directed and Self-regulated Learning and Measuring them in the E-learning

Context. Procedia-Social and Behavioral Sciences. 2014;112:190-8.
59. Brydges R, Dubrowski A, Regehr G. A new concept of unsupervised learning: directed self-guided learning in the health professions. Academic medicine : journal of the Association of American Medical Colleges. 2010;85(10 Suppl):S49-55.
60. Brydges R, Nair P, Ma I, Shanks D, Hatala R. Directed self-regulated learning versus instructor-regulated learning in simulation training. Medical education. 2012;46(7):648-56.
61. Brydges R, Manzone J, Shanks D, Hatala R, Hamstra SJ, Zendejas B, et al. Self-regulated learning in simulation-based training: A systematic review and meta-analysis. Medical education. 2015;49(4):368-78.
62. Konge L, Ringsted C, Bjerrum F, Tolsgaard MG, Bitsch M, Sorensen JL, et al. The Simulation Centre at Rigshospitalet, Copenhagen, Denmark. Journal of surgical education. 2015;72(2):362-5.
63. Bjerrum F, Sorensen JL, Thinggaard E, Strandbygaard J, Konge L. Implementation of a Cross-specialty Training Program in Basic Laparoscopy. JSLS : Journal of the Society of Laparoendoscopic Surgeons / Society of Laparoendoscopic Surgeons. 2015;19(4).
64. Thinggaard E, Bjerrum F, Strandbygaard J, Gogenur I, Konge L. Validity of a cross-specialty test in basic laparoscopic techniques (TABLT). The British journal of surgery. 2015;102(9):1106-13.
65. Strandbygaard J, Maagaard M, Larsen CR, Schouenborg L, Ottosen C, Ringsted C, et al. Development and validation of a theoretical test in basic laparoscopy. Surgical endoscopy. 2013;27(4):1353-9.
66. Ringsted C, Hodges B, Scherpbier A. 'The research compass': an introduction to research in medical education: AMEE Guide no. 56. Medical teacher. 2011;33(9):695-709.
67. Malterud K. Qualitative research: standards, challenges, and guidelines. Lancet. 2001;358(9280):483-8.
68. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. The American journal of medicine. 2006;119(2):166 e7-16.
69. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Medical education. 2015;49(6):560-75.
70. American Educational Research A, American Psychological A, National Council on Measurement in E, Joint Committee on Standards for E, Psychological T. Standards for educational and psychological testing. 2014.
71. Kane M, Brennan R. Educational measurement. Validation, 4th ed. Westport: American Council on Education/Praeger. 2006; 17-64
72. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. Annals of surgery. 2015;261(2):251-62.
73. Korndorffer JR, Jr., Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. American journal of surgery. 2010;199(1):99-104.
74. Cook DA. When I say… validity. Medical education. 2014;48(10):948-9.
75. Cook DA, Lineberry M. Consequences Validity Evidence: Evaluating the Impact of Educational Assessments. Academic medicine. 2016; 91;6:785-795.
76. Jaeger RM. Certification of Student Competence. In: Linn RL, editor. Educational Measurement. 3rd edn. ed. New York: American Council on Education and Macmillian; 1989. 485-514.
77. Kane MT, Crooks TJ, Cohen AS. Designing and Evaluating Standard-Setting Procedures for Licensure and Certification Tests. Advances in health sciences education : theory and practice. 1999;4(3):195-207.
78. McKinley DW, Norcini JJ. How to set standards on performance-based examinations: AMEE Guide No. 85. Med Teach. 2014;36(2):97-110.
79. Norcini JJ. Setting standards on educational tests. Medical education. 2003;37(5):464-9.
80. Ben-David MF. AMEE Guide No. 18: Standard setting in student assessment. Medical teacher. 2000;22(2):120-30.
81. Kolkman W, Van De Put MAJ, Wolterbeek R, Trimbos JBMZ, Jansen FW. Laparoscopic skills simulator: Construct validity and establishment of performance standards for residency training. Gynecological Surgery. 2008;5(2):109-14.
82. Angoff WH. Scales, norms and equivalent scores. In: Thorndike RL, editor. Educational Measurement. 2nd edn. ed. Washington: American Council on Education; 1971. 508-600.
83. Norcini JJ, Shea JA. The credibility and comparability of standards. Appl Meas Educ. 1997;10(1):39-59.
84. Clauser BE, Clyman SG. A contrasting-groups approach to standard setting for performance assessments of clinical skills. Academic medicine : journal of the Association of American Medical Colleges. 1994;69(10 Suppl):S42-4.
85. DiCicco-Bloom B, Crabtree BF. The qualitative research interview. Medical education. 2006;40(4):314-21.
86. Britten N. Qualitative research: qualitative interviews in medical research. British medical journal. 1995;311(6999):251-3.
87. Kuper A, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. Bmj. 2008;337(7666):404-7.
88. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qualitative health research. 2005;15(9):1277-88.
89. Danielsen AK, Spanager L. Anvendelse af fokusgruppeinterview inden for sundhedsvidenskab. Ugeskr Læger. 2012;174:19.
90. Kitzinger J. Qualitative research. Introducing focus groups. British medical journal. 1995;311:299.
91. Malterud K. The art and science of clinical knowledge: evidence beyond measures and numbers. The Lancet. 2001;358(9279):397-400.
92. Downing SM. Reliability: on the reproducibility of assessment data. Medical education. 2004;38(9):1006-12.
93. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice. 2005;8(1):19-32.
94. Gordon M, Gibbs T. STORIES statement: Publication standards for healthcare education evidence synthesis. BMC Medicine. 2014;12(1).
95. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: A review of possible methods. Journal of Health Services Research and Policy. 2005;10(1):45-53.
96. Rumrill PD, Fitzgerald SM, Merchant WR. Using scoping literature reviews as a means of understanding and

interpreting existing literature. Work (Reading, Mass). 2009;35(3):399-404.
97. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. ECTJ. 1981;29(2):75-91.
98. Shenton AK, Hayter S. Strategies for gaining access to organisations and informants in qualitative studies. Education for Information. 2004;22(3-4):223-31.
99. Messick S. Validity. In RL Linn (Ed.), Educational measurement (pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education. 1989.
100. Mosier CI. A critical examination of the concepts of face validity. Educational and Psychological Measurement. 1947.
101. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. Medical education. 2004;38(3):327-33.
102. Schifferdecker KE, Reed VA. Using mixed methods research in medical education: basic guidelines for researchers. Medical education. 2009;43(7):637-44.
103. Cook DA, Beckman TJ, Bordage G. Quality of reporting of experimental studies in medical education: a systematic review. Medical education. 2007;41(8):737-45.
104. Stefanidis D, Heniford BT. The formula for a successful laparoscopic skills curriculum. Archives of surgery 2009;144:77-82.
105. Devine LA, Donkers J, Brydges R, Perelman V, Cavalcanti RB, Issenberg SB. An Equivalence Trial Comparing Instructor-Regulated With Directed Self-Regulated Mastery Learning of Advanced Cardiac Life Support Skills. Simuation in Healthcare. 2015;10(4):202-9.
106. Lineberry M, Park YS, Cook DA, Yudkowsky R. Making the case for mastery learning assessments: Key issues in validation and justification. Academic Medicine. 2015;90(11):1445-50.
107. McGaghie WC, Issenberg SB, Barsuk JH, Wayne DB. A critical review of simulation-based mastery learning with translational outcomes. Medical education. 2014;48(4):375-85.
108. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Medical education featuring mastery learning with deliberate practice can lead to better health for individuals and populations. Academic Medicine. 2011;86(11):e8-e9.
109. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM. Setting Mastery Learning Standards. Academic medicine : journal of the Association of American Medical Colleges. 2015;90(11):1495-500.
110. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): A systematic review of validity evidence. Advances in Health Sciences Education. 2015;20(5):1149-75.
111. Muijtjens A, Hoogenboom R, Verwijnen G, Van der Vleuten C. Relative or absolute standards in assessing medical knowledge using progress tests. Advances in Health Sciences Education. 1998;3(2):81-7.
112. McLachlan JC. The relationship between assessment and learning. 2006.
113. van Empel PJ, Verdam MG, Strypet M, van Rijssen LB, Huirne JA, Scheele F, et al. Voluntary autonomous simulator based training in minimally invasive surgery, residents' compliance and reflection. Journal of surgical education. 2012;69(4):564-70.
114. 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. Surgical endoscopy. 2008;22(8):1887-93.
115. Bjerrum F, Maagaard M, Led Sorensen J, Rifbjerg Larsen C, Ringsted C, Winkel P, et al. Effect of Instructor feedback on skills retention after laparoscopic simulator training: Follow-up of a randomized trial. Journal of surgical education. 2014;72:53-60.
116. Strandbygaard J, Bjerrum F, Maagaard M, Winkel P, Larsen CR, Ringsted C, et al. Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: A randomized trial. 2013;257:839-44.
117. Chiviacowsky S, Wulf G. Self-controlled feedback is effective if it is based on the learner's performance. Research quarterly for exercise and sport. 2005;76(1):42-8.
118. Sørensen SMD, Mahmood O, Konge L, Thinggaard E, Bjerrum F. Laser visual guidance versus two-dimensional vision in laparoscopy – A randomized trial. Surgical endoscopy. 2016;1-7.