Chapter 10

Summary
Summary

The purpose of this thesis was to improve selection, training and assessment in laparoscopic surgery.

Assessment of candidates for medical specializations that require laparoscopic skills is subjective and lack scientific support. While some training institutions in dentistry, aviation and space exploration use aptitude test scores to obtain the optimal distribution of aptitude within their work force, studies on the predictive value of aptitude tests in laparoscopic surgery have been inconclusive about the value of aptitude assessment. Part I (Chapter 2) contains a review of the literature that describes the use of aptitude measurements to predict the acquisition and performance of laparoscopic skills. A meta-analysis was conducted to estimate the predictive power of 4 aptitude measurements: visual-spatial ability, psychomotor ability, perceptual ability and simulator-based assessment of aptitude. Although all aptitude tests showed a significant correlation (visual-spatial ability \( r = 0.32; p < 0.001 \), perceptual ability \( r = 0.31; p < 0.001 \) and psychomotor ability \( r = 0.26; p = 0.003 \)), the highest correlation was observed for simulator-based assessment \( r = 0.64; p < 0.001 \). Moreover, simulators are nowadays widely available in surgical departments involved in surgical training. The most straightforward option would therefore be to use these devices for assessment of aptitude for laparoscopic skills in a ‘laparoscopy aptitude test’. It is important to keep in mind that medical knowledge, communication skills, decision-making skills and clinical judgment are core clinical competencies that should always be considered in conjunction with technical abilities when surgical competence is assessed in candidates.

Part II focuses on training laparoscopic skills in the OR. Training of laparoscopic skills has not been standardized and a large part of the learning process is completed in the high risk environment of the OR.

Chapter 3 describes the identification of common pitfalls during a laparoscopic cholecystectomy. In the ideal scenario, the whole learning curve for procedural learning would be completed in a simulator. However, current simulator computed measurements of improvement and proficiency criteria are often solely based on psychomotor skills such as time taken, instrument path length and number of collisions of the instruments with objects in the simulated work environment. These metrics do not reflect the full spectrum of vital elements for skillful and safe laparoscopic surgery in patients. We used the Pareto analysis to perform a training needs analysis of in vivo laparoscopy. We identified 11 aspects of trainee behaviour that account for 80% of verbal corrections given by supervising surgeons. These included behaviours like exercising the right amount of traction in the right direction with the (non-dominant) left-hand, choosing the right dissection plane and insertion of trocars on the right place and in the right direction. By conducting this analysis we have demonstrated that the Pareto-analysis can be seen as a highly potential method for calibrating laparoscopy skills lab training to on-the-job challenges. However, we have only demonstrated this for the laparoscopic cholecystectomy. Other procedures will also have to be analysed in order to establish content criteria for the whole scope of surgical procedures in laparoscopic surgery. Furthermore, Pareto-analysis based education methods should be evaluated to see whether this kind of training-needs-analysis leads to a higher training efficiency.

Chapter 4 compares two commonly used operation positions for the laparoscopic cholecystectomy, the French and the American position. The operation position used for training is the operation position that a trainee will probably use for the rest of his/her career. There is discussion over whether the French position has an ergonomic advantage in respect to the American position. Because of the difference in orientation of the surgeon towards the work field one would expect that the French position should be preferred. In our study of the surgeon’s poster during the laparoscopic cholecystectomy no statistically significant difference was found between the French and American
position in terms of cervical spine flexion/extension ($p = 0.273$), thoracolumbar spine flexion/extension ($p = 0.273$), cervical spine torsion ($p = 0.715$), thoracolumbar spine torsion ($p = 0.465$), cervical spine lateroflexion ($p = 0.144$), or thoracolumbar spine lateroflexion ($p = 0.465$). No statistically significant difference was found in terms of the time spent within ergonomic acceptable angles in the sagittal plane for the cervical spine (French position, 71.5%; American position, 71.5%; $p = 0.273$) and the thoracolumbar spine (French position, 97.5%; American position, 95.1%; $p = 0.715$), the horizontal plane in the cervical spine (French position, 97.0%; American position, 82.8%; $p = 0.144$) and the thoracolumbar spine (French position, 94.7%; American position, 98.6%; $p = 0.144$) and the coronal plane in the cervical spine (French position, 98.4%; American position, 97.0%; $p = 0.715$) and the thoracolumbar spine (French position, 98.3%; American position, 97.4%; $p = 1.000$).

Our results therefore indicate that, in the MIS suite, it does not seem to matter for the posture of the vertebral column whether the French or American position is used for the laparoscopic cholecystectomy. This is most likely a consequence of the presence of movable monitors in the MIS suite. However, more research is necessary to identify the aspects of the upper extremities which may especially be at risk of being overburdened in the American position in the MIS suite considering the position of the surgeon in this operation setup.

Chapter 5 states the key steps for 2 standardized laparoscopic procedures. This study uses a validated method, the Delphi method, to reach consensus among a group of twenty-one experts in laparoscopic surgery about which procedural steps should be seen as key steps for the laparoscopic cholecystectomy and appendectomy. Consensus was observed after the first round of Delphi on the key steps for laparoscopic appendectomy (Crohnbach’s alpha 0.92) and laparoscopic cholecystectomy (Crohnbach’s alpha 0.90). After the second round, 15 proposed key steps for laparoscopic appendectomy and 30 proposed key steps for laparoscopic cholecystectomy were rated as important (importance score ≥4/5) by at least 80% of the expert panel. These key steps will be used in standardized training for trainees in the North-East surgical school and were used to create a procedure-based assessment for the laparoscopic cholecystectomy. The procedure-based assessment was evaluated in part III.

In Part III, we focus on the current subjectivity in the assessment of surgical skills and the absence of standardized methods for the assessment and certification for procedural skills.

Chapter 6 is focused on important aspects of the psychometrics behind inter-rater reliability. Twenty fold differences have been reported between the results of the 6 different mathematical models that can be used to calculate the intra-class correlation coefficient (ICC), a commonly used reliability coefficient for inter-rater reliability. This is a problem in current research about surgical education, as the majority of studies addressing the reliability of surgical skills assessment do not report which mathematical model was used. Second, there are some important issues pertinent in the evaluation of study quality in reliability research. Some of these have similarities with drug research, such as blinding of observers and random sequence generation, but manifest differently in the assessment of surgical skills. Third, the correct way of interpretation of the ICC is dependent on the purpose of the measurement instrument. Cut-off values, confidence intervals and probability distributions are all options that can be considered. Furthermore, it is important to take into account the constructivist social-psychological approach to assessment when interpreting surgical skills assessment. This means that different interpretation can sometimes be seen as equally valid as they are based on the individual professional experience, knowledge and socialization of the surgeon.

Chapter 7 describes the validity and reliability with a global rating scale (GRS) especially designed to assess laparoscopic skills, the Global Operative Assessment of Laparoscopic Skills (GOALS). GOALS was used to assess blinded randomized video fragments of a laparoscopic cholecystectomy. We used recordings of 6 consecutive cholecystectomies performed by 10 trainees. Out of the video recordings 3 fragments were edited to produce a total of 160 video fragments which were randomized and
assessed by two blinded laparoscopic surgeons previously unexposed to GOALS. Our study supports the existing evidence that GOALS has construct and concurrent validity for assessment of novice trainees performing a laparoscopic cholecystectomy. However, the reliability observed in this study was low (ICC=0.37) compared to the reliability found in other studies. There are a number of causes that could have been responsible for the low reliability: a lack of intrinsic/extrinsic motivation of the raters, fatigue of the raters, the lack of training in assessment and the characteristics of the Likert scale used in GRSs. The findings have lead to a more rigorous methodological approach in the study of surgical assessment described in chapter 8.

Chapter 8 describes a new method for assessing procedural learning: independence-scaled procedural assessment. A procedural assessment system for a basic laparoscopic procedure has been developed by linking the key steps of the laparoscopic cholecystectomy to a scale of independency. The scale consists of 5 different levels for every step: 0) Did not perform the step, 1) Able to perform a part of the task, 2) Performs the task with much guidance and instructions, 3) Performs the task with minimal guidance and instructions, 4) Can perform the whole task independently, safely and skilfully. The procedural assessment was compared with 2 GRSs, the Objective Structured Assessment of Technical Skills (OSATS) and GOALS, in terms of validity, reliability and support among 10 surgeons and 6 scrub nurses. The participants rated blinded and subtitled full procedural videos of: 1) a novice trainee 2) an intermediate trainee and 3) a subcompetent trainee. Because of our findings in chapter 7, we attempted to calibrate the participants by showing them a short video of a laparoscopic performance of the low- and high-end of the scales of the GRSs. In contrast to the procedural assessment, the GRSs were not able to differentiate the intermediate from the subcompetent trainee. Thus, the discriminative validity of the procedural assessment was higher than for the GRSs. Furthermore, the surgeons showed a good reliability for the GRSs (OSATS 0.78; p < 0.05 and GOALS 0.74; p < 0.05), but an almost perfect reliability for the procedural assessment (0.84; p < 0.05). A survey that was distributed together with the surgical assessment forms showed that most surgeons were of opinion that the independence-scaled procedural assessment: 1) gives a more detailed picture of procedural skills than the GRSs, 2) is more objective than the GRSs and 3) should be reproduced for other laparoscopic procedures than the laparoscopic cholecystectomy. These findings indicate that the independence-scaled procedural assessment is a candidate that meets up to the requirement of an assessment tool for post-operative formative feedback, but perhaps also for high-stakes examinations such as certification. Furthermore, the reliability coefficients increased when the ratings of scrub nurses were added to those of the surgeons, indicating that scrub nurses can reliably assess the procedural laparoscopic skills of surgical trainees.

Interestingly, the step ‘the dissection of Calot’s triangle’, displayed a moderate reliability (0.50). This key step has been identified as the most difficult subtask of the laparoscopic cholecystectomy and therefore demands a set of complex technical behaviours. In the light of constructivist social-psychological approach, raters will therefore focus on different aspects based on their knowledge, experience and previous socialization, causing a lower reliability than for the less difficult key steps. This also means that, although the ratings do not agree in the assessment of the dissection of Calot’s triangle, they might all be equally valid, because they are based on the unique professional experience and understanding of the individual assessors. However, further quantitative, but definitively also qualitative research, is necessary to investigate whether this decrease in reliability can be thwarted by (practical) adjustments in the method of procedural assessment.