Chapter 9

General discussion and future perspectives
General discussion and future perspectives

After the achievement of proficiency criteria in laparoscopic skills training outside the OR, trainees commence their learning experience in the OR supervised by a consultant surgeon who guides them through the procedure. This thesis is focused on the enhancement of these educational efforts. The goal of this thesis is to improve: 1) assessment of candidates for medical specialties that require laparoscopic surgery, 2) procedural laparoscopy training on the OR and 3) post-operative procedure specific performance feedback and assessment.

Part I: Aptitude

The question of whether aptitude tests that evaluate visual-spatial ability, perceptual ability and psychomotor ability can be used in the assessment of candidates for medical specialties that require laparoscopic skills is currently a topic of debate. We conducted a meta-analysis to: 1) evaluate whether aptitude assessments can be used to predict the ability to acquire and perform laparoscopic skills, 2) to quantify how much of the variability in skills can be predicted by aptitude assessment and 3) to obtain insight in the factors that influence the strength of this relationship.

Although no statements can be made about surgery in general, we can state that the acquisition and performance of laparoscopic skills can partly be predicted with aptitude measurements. Assessment of aptitude in the form of visual-spatial ability tests, perceptual ability tests, psychomotor ability tests and simulator-based assessment all showed a significant correlation with surgical training. A significant correlation was also found when only studies that used aptitude tests to predict performances during an OR training session were singled out.

The ergonomic challenges of the OR environment encountered during surgical procedures on human beings provide a theoretical support for the association of laparoscopic surgery with the content of these aptitude tests. As the results in the meta-analysis support this construct, program directors can feel legitimized to use a laparoscopy aptitude test (LAT) in the assessment of candidates that require laparoscopic skills, even without the extensive validation of these tests on the basis of job performances as a fully certified laparoscopic surgeon, which is the ultimate measure of predictive validity. The latter is a difficult task considering the fact that during in vivo laparoscopy there can always be unexpected visual-spatial, perceptual or psychomotor challenges that place high demands on the cognitive abilities of laparoscopic surgeons. It might be hard to estimate individual performance level on these instances, as operative demands that exceed the capacity of the surgeon often involve an inversion of the indirect and direct control dynamics described in chapters 3 and 8, manifesting as the cognitive or physical support by another, often more experienced, surgeon in the department. At other times, the increased demands might be only expressed in a longer operation time, a variable which is also determined by a multitude of other variables.

It is important to keep in mind that a LAT can be used to optimize the selection process of candidates, but can be just as beneficial for career coaching of medical students. A LAT can help make the right career decision and/or support surgical educators in the recommendation to opt for a specific area of medicine. Students tested with a high aptitude interested in a non-surgical career can obtain a stimulus to consider pursuing a surgical career and those with a low aptitude interested in a career involving laparoscopy have the opportunity to invest their valuable time and energy in a specialty or differentiation program that better matches their talent or to keep pursuing their dream knowing that they might have to work harder to attain the same level of competency as their peers.

If the evidence supports the theory that visual-spatial, perceptual and psychomotor ability are predictors of laparoscopic skills, which aptitude measurement should be used in a LAT? As simulators seem to measure all 3 forms of aptitude at once and are available in most academic hospitals with surgical training programs, the most straightforward option would be to use these devices for aptitude assessment with a norm-referenced scoring system. On the other hand, the increasing
availability of simulators in the form of serious games\textsuperscript{1} introduces the danger of measuring the amount of adaptation to a human-machine interface instead of aptitude. Other options which include the use of a test battery of visual-spatial ability, perceptual ability and psychomotor ability, with or without simulator based assessment, are organizational demanding and therefore financially more burdening. Thus, the question is whether we can find a way to prevent adaptability to the human-machine interface to become a problem in aptitude assessment. Perhaps, the most practical solution would be to introduce an instructional course and allow free practice on freely accessible simulators for medical students and subsequently use a set of inaccessible difficult simulator tasks, tasks that have a high level of unpredictability and have the ability to distinguish in aptitude even after motivated in vitro training, to assess these students during a LAT (Figure 1). This idea is supported by a recent publication that showed that the relationship of aptitude with laparoscopic skills remains present also after an instructional course in combination with voluntary practice.\textsuperscript{2}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Laparoscopy Aptitude Test (LAT) is preceded by a time period of supervised simulator training and free practice on simulators for applicants. The LAT itself contains unpredictable simulator tasks that are inaccessible to applicants and has the ability to distinguish different levels of cognitive ability in motivated simulator trained subjects. Red: applicant with previous experience on serious video game but low aptitude score on LAT, blue: applicant with no previous experience but high aptitude score on LAT, green: applicant with no previous experience and intermediate aptitude score on LAT.}
\end{figure}

\textbf{Part II: Training}

In 1887, Vilfredo Pareto published his observation of an exponential law between the amount of wealth an inhabitant owned and the rank-order of the inhabitant.\textsuperscript{3} He concluded that 80\% of property is owned by merely 20\% of the inhabitants. This concept, that a relative few account for a large proportion of a common effect, also has become known as the ‘80-20 rule’. We used the Pareto-analysis to state content criteria for surgical training. On the basis of the results, it seems that the Pareto-analysis is a tool with high potential for identifying on-the-job challenges as 35 of 253 (13.8\%) of the different verbal corrections were responsible for 80\% of the total number of verbal corrections counted. We have suggested training methods with a high content validity on the basis of the results. These methods vary from simple instructional courses to technically complex gadgets to enhance training efficiency. Further research is necessary to evaluate the effectiveness of surgical education methods developed on the basis of a Pareto-analysis.

To investigate the different positions that can be used to perform a laparoscopic cholecystectomy on patients, we conducted a cross-over study of the French versus the American position. No statistically significant difference was found between the French and American position in the...
posture of the vertebral column among surgeons. On the basis of the place of trocar insertion however, it can be hypothesized that the left arm is at risk for being overburdened in the American position, the position most often used in the Netherlands. Research focused on the shoulders, arms and hand movements instead of the vertebral column might be able to provide more information about the degree of strain of the upper extremities in the American and the French position, hence, making the preference for one of the two operation setups in surgical education more justifiable.

Besides the attempt to reach consensus on a preferable operation position for the laparoscopic cholecystectomy, a Delphi survey was performed to reach consensus on the key steps of 2 basic laparoscopic procedures, the laparoscopic cholecystectomy and the laparoscopic appendectomy. The Delphi method has recently been used by our research group to establish the key steps of more sophisticated laparoscopic procedures: the laparoscopic right hemicolecctomy and the laparoscopic sigmoid colectomy. We consider reaching consensus on key steps of laparoscopic procedures as the first steps towards a standardized curriculum in laparoscopic surgery training. Delphi survey based key steps can facilitate deliberate practice on the following ways:

1. Pre-operative preparation for OR training.
2. A roadmap for the stepwise teaching in surgical training for laparoscopic procedures.
3. Post-operative assessment of procedural skills.

In part III we evaluate whether the key steps of the laparoscopic cholecystectomy can be used for two of these goals, the post-operative assessment of procedural skills and the examination for independent surgical treatment.

**Part III: Assessment**

Assessment of procedural learning is an important aspect of assessment that has yet to mature in many aspects. We have made the first step in creating a procedural assessment of the laparoscopic cholecystectomy by connecting a list of procedural key steps of the laparoscopic cholecystectomy established with the Delphi method to a scale of independency. The independence-scaled assessment has a scale that connects to the control management used to guard patient safety by the supervising surgeon in the OR (Figure 2).

![Figure 2: Interaction between intra-operative safety control dynamics, introduction of stepwise autonomy and procedural assessment. 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 independent, safe and skilful. Red: within key step control management by supervising surgeon, blue: between key step control management by supervising surgeon.](image)

The study results demonstrate a higher discriminative validity and inter-rater reliability of independence-scaled procedural assessment compared with the assessment with global rating scales. Our research group is currently conducting a study to define a cut-off score with an
acceptable sensitivity and specificity in identifying trainees who can be labelled as competent and receive a certification for the independent treatment of uncomplicated disease. This would complete the structured stepwise training and assessment system of laparoscopic procedures depicted in figure 3. Further research will also be necessary to evaluate whether the higher discriminative validity and inter-rater reliability of independence-scaled procedural assessment are also present in the assessment of advanced laparoscopic procedures.

Figure 3: Structured curriculum for laparoscopic surgery training. Input of regional experts is used in a Delphi survey for the establishment of a list of key steps for procedural training and assessment for laparoscopic procedures. Delphi key steps are used for: 1) Pre-operative statement of learning goals, 2) Intra-operative teaching and a stepwise increase in autonomy 3) Post-operative assessment and 4) Certification for surgical treatment of uncomplicated disease.

After the publication of chapter 7 a sub-analyses was performed according to the guidelines in chapter 6, a chapter largely written after the publication of chapter 7 and 8. A correlation matrix was calculated to investigate whether the low reliability was somehow related to the (unrandomized) chronological order in which the video fragments were rated. The absolute agreement intra-class correlation coefficient (AA-ICC(2,1)) was calculated for video fragment 1 to 3 split into three groups: 1) early assessments, 2) middle assessments and 3) late assessments. In contrast to the earlier found low ICCs, the inter-rater reliability in the early groups were reasonable to good; 0.61 for fragment 1 (p = 0.004), 0.63 for fragment 2 (p = 0.002) and 0.42 for fragment 3 (p = 0.016) (Table 1). Also, the middle and late group ICCs were all non-significant and a consistent decrease was observed in the inter-rater reliability towards the late assessments.

Table 1: The AA-ICC’s (2,1) of the early assessments, middle assessments and late assessments of the three video fragments F1-F3.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Early (N=54)</th>
<th>Middle (N=54)</th>
<th>Late (N=52)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 (N=51)</td>
<td>0.61*</td>
<td>0.27</td>
<td>0.19</td>
</tr>
<tr>
<td>F2 (N=53)</td>
<td>0.63*</td>
<td>0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>F3 (N=56)</td>
<td>0.42*</td>
<td>0.24</td>
<td>0.04</td>
</tr>
</tbody>
</table>

* Statistical significant ICC-value (p < 0.05)

Raters involved in surgical education probably have a high drive to invest energy and time in the assessment of trainees. However, during the assessment, intrinsic motivation can be jeopardized by factors such as mental fatigue or time pressure. The results shown in table 1 indicate that these factors indeed may have decreased the accuracy in the assessment of surgical skills. In the subsequent evaluation of inter-rater reliability among assessors of surgical skills in chapter 8 we
therefore limited the assessments to an acceptable time frame and rewarded raters for the assessment of trainees to counteract any loss of motivation due to fatigue as much as possible. The above results should furthermore be seen as a warning to researchers and program directors who (unconsciously) overburden assessors during an attempt to gain surgical skills assessments.

Effect of training of assessors has been addressed in this thesis in chapters 6, 7 and 8. Teaching and assessing surgical raters is a field wherein much remains to be discovered. Questions like, what kind and how much training is necessary to make an assessment an accurate measurement of surgical skills, are still unanswered. An option would be to randomize a group of raters into two groups, a trained and untrained group, which uses a global rating scale and procedure-specific assessment to rate a series of performances.

Help from an experienced supervisor is crucial in the completion of a high risk complex task. It is the common perception that training with intensive training support leads to a higher performance level and faster attainment of proficiency during learning. Interestingly, psychologists have emphasized that learning should not be considered without taking in account the amount of retention of the learned skills. The degree of retention can differ significantly between learning methods. It has been shown in psychology that intensive guidance by trainers leads to higher performance level during training, but also to a higher decay of the acquired skills when the subject is to perform the same task in a later moment of time without the help of the trainer. This phenomenon has become known as the ‘guidance hypothesis’ and seems to be caused by the continuous provision of instructions. The constant verbalisation of the mind of the supervisor leads to insufficient free work capacity in the trainee to transform the work memory into chunks of information and store the chunks in long term memory. In our interaction with surgeons in our institution we have noticed that the prospect of an assessment that takes into account the amount of supervision a trainee actually needs to complete the operation induces a reticence in the supervising surgeon that would otherwise be absent. Restraining on the provision of cognitive support during a laparoscopic procedure might have the same diminishing effect on the decay of laparoscopic skills as has been observed in the retention of acquired complex motoric tasks in the field of educational psychology. Further research could provide information about whether this phenomenon also exists in surgical training.
References


