

# Module Based Training Improves Performance in Laparoscopic Surgery: A Nationwide Randomized Controlled Trial

Charlotte Green Carlsen\*, Karen Lindorff-Larsen, Peter Funch-Jensen, Lars Lund, Peder Charles and Lars Konge

Aarhus University, Center for Medical Education, INCUBA Science Park, Brendstrupgårdsvæj 102, Building B, 8200 Aarhus C, Denmark

## Abstract

**Objectives:** New working time directives reduce the available training hours for futuresurgeons. Laparoscopic training must be more efficient, and deliberate clinical operative training within a short time-frame (module based training) might help achieve this goal. This study set out to measure the effect of such a module based training programme for surgical trainees applied to laparoscopic cholecystectomy.

**Method:** Randomized controlled trial in Danish surgical training hospitals. Thirty-five Danish surgical specialty trainees (registrars) were randomized to either module based training including performance of 20 procedures within 4-8 weeks or standard clinical training. The laparoscopic cholecystectomies were videotaped and blindly rated by three professors/consultants using the validated tool: Global Objective Assessment of Laparoscopic Skills (GOALS).

**Results:** Data show a significant effect of the module based training program on technical skills compared to standard training at one year follow-up despite equal number of performed procedures in the two groups ( $p=0.010$ ). Due to a large number of drop-outs only ten participants completed the study with the required number of videotaped operations. Inter-rater reliability showed Cronbach's alpha=0.37.

**Conclusions:** Only few trainees completed the study and inter-rater reliability was low, which highlight some of the challenges in educational randomized studies. Despite this, module based training showed promising effect as an efficient method in clinical skills training of surgical registrars.

**Keywords:** Competency-based training; Laparoscopic cholecystectomy; Surgical training; Practice-based learning

## Introduction

The need for efficient laparoscopic training is evident as surgical trainees spend less time in the operating room [1]. This may disperse training critically, leading to inefficient learning. In spite of advanced simulation courses, workplace learning is still the cornerstone of surgical training. An important aspect to be considered in this connection is patient safety, which must be ensured by supervision and regular assessment of the trainee. High quality assessment that is reliable and valid is of major importance.

Laparoscopic procedures such as cholecystectomy are common, and surgical trainees must be able to perform uncomplicated cases unsupervised after few years in the field. A well-defined training strategy in laparoscopic procedures including bench models, simulation and workplace learning is needed.

Built on Miller's framework for assessment [2] that includes the stages "knows", "knows how", "shows how" and "does", a new concept was described and explored: Module based training. The model aims to train and assess Miller's "does"-level, i.e. the training that takes place in the workplace. The model aligns with many other frameworks for competency-based training, deliberate training, situated learning, workplace-based training and content specificity [3-7]. It also fits the intuitive belief that training matters.

The module based training program for laparoscopic cholecystectomy consisted of a one day course comprising theory and simulator practice followed by workplace based training of a set number of surgical procedures within a given time period in order to reduce dispersion in training.

The introduction of new training models should be based on solid evidence. However, applying randomized controlled trials (RCT) to educational studies may be challenging. In conducting educational

studies a number of things have to be observed [8] and in a real life setting there are many pit-falls. Case-mix, work load and acute cases may change educational priorities. Intervention and control groups may not be blinded within identical departments, which may impact training. Time is a crucial component in educational studies as trainees progress over time. Finally, outcome must be measured in a reliable and valid way. Validity, however, is construct specific and depends on the circumstances of the measurement.

With these difficulties in mind we decided to explore the feasibility of a nationwide RCT using a previously "validated" assessment tool to measure the effect of a new training regime: Module based training.

### Objectives:

1. To explore the feasibility of including multiple departments in a nationwide educational RCT.
2. To explore the feasibility of using a previously validated assessment tool in a new context (i.e. rater group)
3. To measure the effect of a module based training model applied in laparoscopic cholecystectomy for surgical specialty trainees. The intervention was measured against the standard training program. The intervention was built on Miller's framework including a course to provide and test: 1. Knowledge (knows), 2.

**\*Corresponding author:** Charlotte Green Carlsen, MD, PhD, Aarhus University, Center for Medical Education, INCUBA Science Park, Brendstrupgårdsvæj 102, Building B, 8200 Aarhus C, Denmark; Tel. +45-7841 0183, E-mail: [cgc@medu.au.dk](mailto:cgc@medu.au.dk)

**Received** August 09, 2014; **Accepted** November 28, 2014; **Published** December 05, 2014

**Citation:** Carlsen CG, Lindorff-Larsen K, Funch-Jensen P, Lund L, Charles P, et al. (2014) Module Based Training Improves Performance in Laparoscopic Surgery: A Nationwide Randomized Controlled Trial. *Surgery Curr Res* 5: 214. doi:[10.4172/2161-1076.1000214](https://doi.org/10.4172/2161-1076.1000214)

**Copyright:** © 2014 Carlsen CG, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Bench station training (knows how), 3. Animal model training (shows how) followed by supervised training: 4. Performance and assessment in the operating room (does).

## Material and Methods

### Design

The study was designed randomized in two parallel groups: intervention and control. ClinicalTrials.gov identifier: NCT01877824.

### Rating scale

A number of different tools are available in the laparoscopic domain, especially for laparoscopic cholecystectomy. A literature search in PubMed and Embase led our group to Global Objective Assessment of Laparoscopic Skills (GOALS), apparently an “off-the-shelf” solution that had formerly been validated for several purposes including rating on video recordings [9-11]. Laparoscopic cholecystectomy is a well-described procedure that we assumed was performed under quite similar circumstances throughout the world. The procedure comprises removal of one well-defined organ, though some of the cases may be more complicated due to e.g. cholecystitis, prior surgery etc. GOALS [10] was considered useful for this study, as the tool had been validated for both direct observation and video rating [9]. However, validation is specific of context and therefore the tool was tested in the rater group on pilot recordings. Following introduction to the tool, the raters watched the videos individually. Afterwards, the results were debated in the rater group. The inter-rater agreement was low and potential reasons were elucidated. Possible misinterpretations and other issues in the tool were discussed. Finally, all raters agreed on common principles for their future use of the tool.

### Randomization

Prior to randomization, hospitals were invited to participate. All new surgical trainees (registrars) in 2010 and 2011 in surgical departments of participating hospitals were asked for their individual consent. Exclusion criterion was prior operative experience to the extent of permission to perform laparoscopic cholecystectomies unsupervised. Randomisation was made with consecutively numbered envelopes shortly before the trainees started their first year of formal surgical training. If two trainees from the same department started simultaneously, they were block randomised due to considerations of the colleagues and avoidance of grudge.

### Setting

Surgery was performed in day-units and supervised by local senior staff. This setting was identical to any other clinical surgical training session in the hospitals. The aim was not to change the surgical volume obtained within the first year. Only the sequence and timing of the procedures was changed.

### Participates

Intervention group participants were allocated to a one-day skills lab course in laparoscopic cholecystectomy. The course comprised a theoretical lesson followed by skills training on LapSim (basic skills training). Through the course the participants followed Millers' steps 1. Knows (theory), 2. Basic skills (knows how), 3. Simulated cholecystectomy (shows how). The participants were tested at the beginning and the end of the course day. After the course the participants were assigned to perform 20 supervised laparoscopic cholecystectomies within 4-8 weeks (Millers' number 4. Does). They were allocated 4-6 days in their departments' day surgery facilities. On

the first day of the assigned procedures their performances were video recorded. Another video recording was made of a similar procedure performed at the end of the module based training program. Finally, a recording was obtained within the last months of the first training year.

In the control group only the video recordings at the beginning and end of the first year of specialty training differed from an ordinary training session. Besides the planned day-unit assignments, all participants followed the regular portfolio training program in their departments.

Both groups were asked to fill in a form containing operative time and task taken over by the supervisor. All participants were asked to count their personal number of laparoscopic cholecystectomies performed before study entry and during the study (from the first video recording in the study to the last video recording by the end of the first training year).

### Video recordings and ratings

The procedures were taped on a recorder connected to the laparoscope (different techniques depending on the department's equipment. This meant that only the procedure from camera insertion in the abdominal cavity to removal was recorded and rated. The recordings were taped directly in the operative theatre and recordings were mailed to the investigator. No investigator was present in the operating room. All video recordings were blinded regarding patient, trainee, supervisor and department. The ratings were performed by three independent consultants all specialists in surgery. The raters were blinded to trainees' training status. Each video was rated by three independent raters once, and in order to mix the recordings for the raters, no recordings were rated before the study was finished.

### Ethics

Patients gave consent to the recordings prior to the procedure. The video recordings were anonymous, as no patient data were recorded or known to the investigators. Consequently, no permission from the Research Ethics Committee or the Danish Data Protection Agency was needed according to Danish legislation. The Ethics committee was consulted prior to study start.

### Main outcome measures

The average rating score of technical skills (GOALS) and the time to complete the operative procedure.

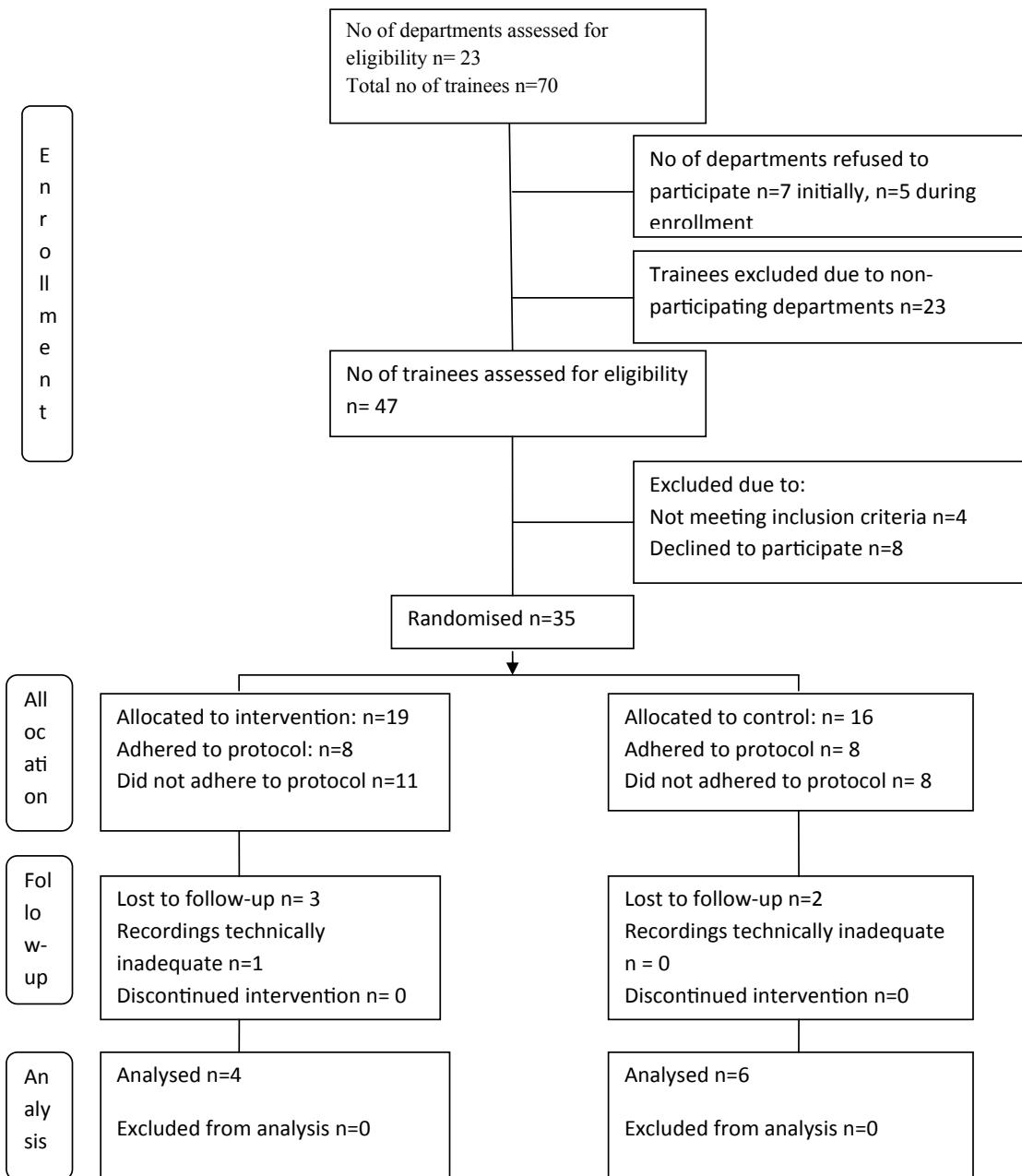
### Statistics

Rating scores are presented as average score of all observers. No calculation of strength was performed as all departments were asked. A number of participants >30 was considered acceptable compared to similar studies. The level of statistical significance was defined as  $P < 0.05$ . Cronbach's alpha was calculated to compare raters. A generalizability study and D-study was performed to estimate the needed number of raters. Rater scores are shown as median and compared using Mann-Whitney U test. Statistical analyses were performed using a statistical software package (PASW, version 18.0; SPSS Inc., Chicago, Illinois, USA).

### Results

#### Feasibility of a nationwide educational RCT

Seventy Danish surgical trainees were accepted for formal five-year surgical specialist training in 23 different hospitals during the years 2010-11. However, 12 departments declined to participate before



**Table 1:** Flow chart of participants

randomisation. Mainly the large university hospitals declined for a number of reasons. In total, 47 trainees were invited, 12 trainees were excluded or did not respond, leaving 35 trainees for randomisation. One control group participant did not contribute due to his department's late withdrawal from the study; another participant withdrew her consent after randomisation. Nine participants did not accomplish the protocol for a variety of reasons (illness, procedures not arranged or procedures cancelled, forgot to record the procedure), one recording was technically inadequate (a specific electronic feature in the recording made it impossible to view after a certain date), and three intervention group participants were lost to one year follow-up. In the control group eight did not adhere to protocol, two were lost to follow-up. Finally four participants in the intervention group and 6 in the control group completed the study. Table 1 shows participant flow.

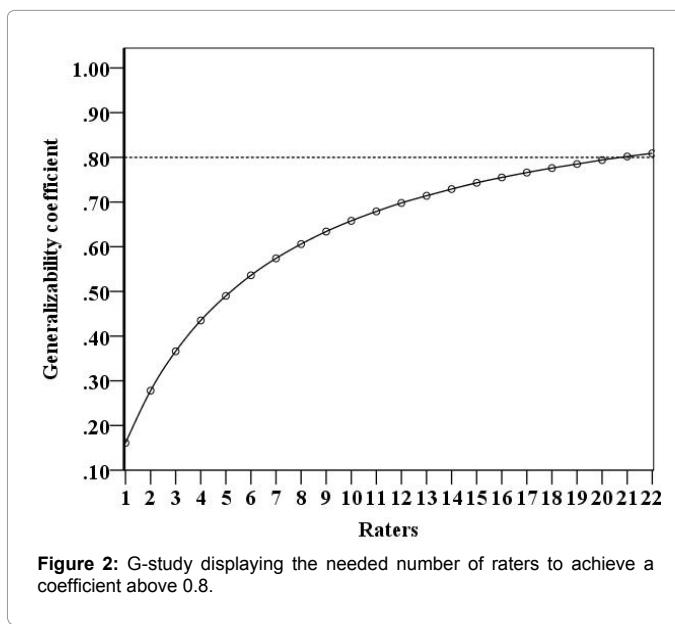
#### Feasibility of a the assessment tool in a new context

Participants were asked to fill in a form containing information of operative time and if part of the task had been taken over by the supervisor. However, only one participant in the intervention group turned in the form! Consequently the GOALS had to be modified, as the item "autonomy" was considered impossible to rate when the information concerning supervisor takeover was unavailable. All raters accepted the modifications (figure 1). In the modified form, the minimum score was 5 and maximum score was 25. Time to complete the procedure was also noted in another form and therefore available.

The inter-rater reliability results showed an Intraclass Correlation Coefficient (ICC) of 0.37 based on average measures. For a single rater it was 0.16. Based on absolute agreement the total was a bit lower, 0.35.

	1	2	3	4	5	Point
Depth perception	Constantly overshoots target, wide swings, slow to correct		Some overshooting or missing of target, but quick to correct		Accurately directs instruments in the correct plane to target	
Bimanual dexterity	Uses only one hand, ignores nondominant hand, poor coordination between hands		Uses both hands, but does not optimize interaction between hands		Expertly uses both hands in a complimentary manner to provide optimal exposure	
Efficiency	Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress		Slow, but planned movements are reasonably organized		Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach	
Tissue handling	Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips		Handles tissues reasonably well, minor trauma to adjacent tissue (ie, occasional unnecessary bleeding or slipping of the grasper)		Handles tissues well, applies appropriate traction, negligible injury to adjacent structures	
Overall competency	Unable to complete entire task, inefficient effort, even in a straightforward case.		Able to complete task safely even if the case is slightly challenging		Able to complete task safely in spite of challenging case (adherences, inflammation, etc)	

**Figure 1:** Rating scale modified from GOALS.



**Figure 2:** G-study displaying the needed number of raters to achieve a coefficient above 0.8.

The Index of Dependability in a G-study and the subsequent D-study was explored. The needed number of raters to obtain a score of >0.8 was 21 (Figure 2). Figure 3 illustrates the rater's individual scores.

### Effect of module based training compared to traditional training

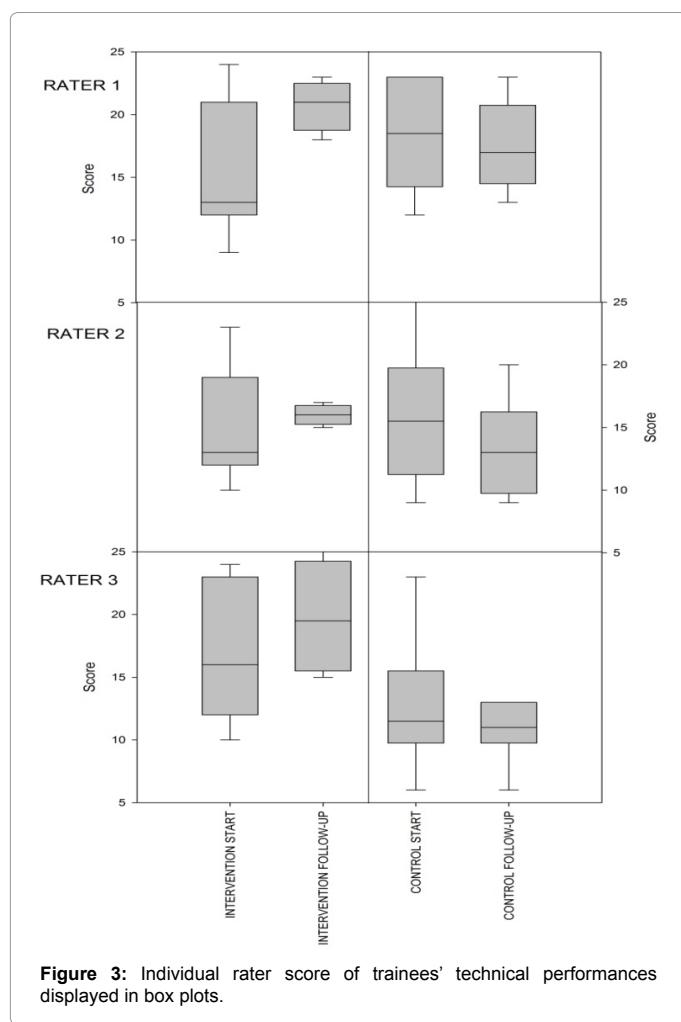
Four participants completed the intervention group. Their median

rating score entering the study was 14.3 and at end of study 19.2. In the control group six participants from three different hospitals were analyzed and they scored at start 14.7 and at end 13.2. Comparing the two groups in Mann-Whitney U test at study start no difference was detected ( $p=0.762$ ) but at end of study skills rating was much in favor of the intervention ( $p=0.010$ ) (figure 5). Participants in the two groups had averagely performed an equal number of laparoscopic cholecystectomies before entry and by the end of the first training year, 18 vs 19 ( $p=1.00$ ) before and 53 vs 54 after ( $p=0.76$ ). The time between the two observations in the interventions group was 201 days and in the control group 229 days ( $p=0.20$ ).

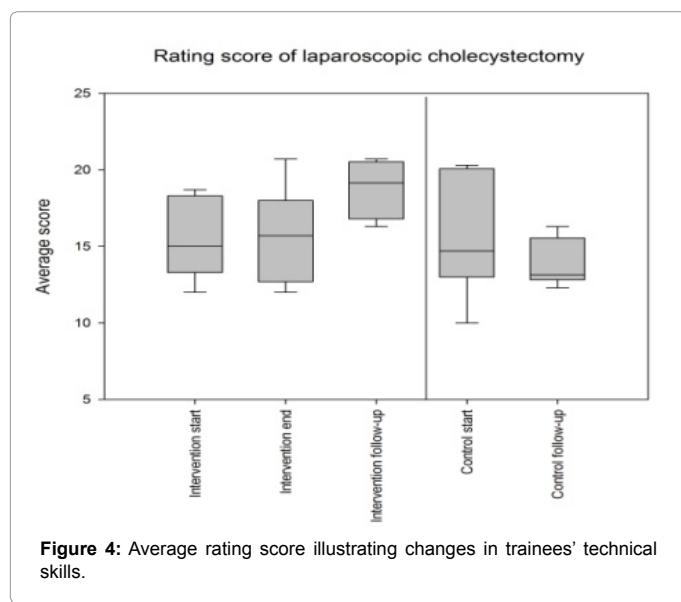
### Discussion

#### Feasibility of a nationwide educational RCT

The intention here was a nationwide study to compare educational outcome in terms of a work-place based assessment of surgical technical skills. To achieve a sufficient number of participants (>30), the study period was two years. An RCT in gynecology in one educational region in Denmark was formerly carried out [12] and had experienced drop-outs. Consequently, this large study was carefully planned and program directors were involved before study start. A positive spirit existed. A large number of the departments, however, did not attend the study due to different reasons. The trainees were very excited about this study as they hoped to achieve more real surgical procedures. In reality, it became hard to retain interest in the study from the program directors' work-planning performing personnel. Some of the participants did not achieve the planned operations within the fixed time due to sudden changes of work-schedules (emergency work, sick-leave, program changes). Or they never got the opportunity due to lack of planning.



**Figure 3:** Individual rater score of trainees' technical performances displayed in box plots.



**Figure 4:** Average rating score illustrating changes in trainees' technical skills.

Others forgot to record the videos or did not send the videos. Some of the videos were blank or otherwise unattainable. Only one participant returned the form concerning difficulty of the case and supervisor take over. This made it difficult for the raters, but may have enhanced

the authenticity that trainees did not worry about assessment while performing in the operative theatre. This massive drop-out from university hospitals and other hospitals with many trainees complicated the generalizability of the study to larger hospital settings. All together, this study highlights the difficulties in implementing new habits in education. Laparoscopic surgery is in need of efficient training, including workplace-based training models. But it seems difficult to introduce this new model.

#### Feasibility of a the assessment tool in a new context

The recordings were to be assessed with the formerly validated tool: Global Objective Assessment of Laparoscopic Skills (GOALS) [10,11]. As similar tools, GOALS was tested and used by the faculty involved in the development [11;13;14] which may impact generalizability. The use an “of-the-shelf” solution was decided. Rater training is needed but scarcely described in literature [15]. The raters discussed pro and cons of the selected tool and reached agreement on where to pay attention on techniques in operative performance. Consequently, the modified GOALS version for video [9,10] was chosen. Due to low inter-rater agreement in the pilot study, another meeting was held and results discussed as described above. Especially, tissue handling was emphasized. The autonomy item also had to be excluded due to lack of information as the form concerning supervisor takeover was not returned properly. It was decided to use three raters to assess the videos because of the initial problems with inter-rater agreement. However, the inter-rater agreement remained low in the main study as well.

Rater training is emphasized by Vassilio and colleagues especially in regard to video assessments [15] where four raters were used. However, some of their raters were inexperienced. Consequently, raters were debating their incongruent results and reached common conclusions on principles of rating. Three experienced raters were used. Differences in personal approach to surgical techniques were debated at the meeting and the raters subjectively felt well-prepared for the job. However, the importance and amount of the rater training must have been underestimated. Probably the rater training must be more specifically designed to be efficient and assessor training in the operative room may be needed to attain comfort. Such training, however, will be more costly for use in a clinical setting. In this study design rater training in the operating room was not applicable either, but must be recommended in future studies.

In the results of our study, each rater was individually consistent in the ratings. We made effort to emphasize the different approaches in the rater training and the issue of personal preferences was specifically addressed. How to perform efficient rater training remains an open question of great importance if research developed tools are used in clinical settings outside the developing institution.

The raters were selected from a broad group. One rater was an educational responsible consultant, another rater was specialized in urologic laparoscopy and the last rater specialized in gastrointestinal laparoscopy. This broad rater setting was believed to reflect daily clinical work thereby improving generalizability. The results also challenge the value of this “broad” assumption. A much more specific training and knowledge may be needed. Ratings may have been performed more uniform by sub-specialized surgeons who were experts in laparoscopic cholecystectomy. Though, the use of such “specialized” may be beneficial, the “general” rater is a common feature in the work-place based training. Furthermore, raters are rarely specifically trained to new assessment tools. A change in attitude, however, may be necessary. This study concludes that a more than brief training program is needed for raters. Results showed that performances were rated consistent from

the individual rater, but the raters displayed too different approaches to rating affecting reliability. This issue must be carefully considered using formative assessment tools in the surgical training.

## Module based training

In spite of the above described limitations, conclusions of this study's third aim are clear. The strength of the study was the approximation to the real-life setting. The recordings of the performance could be considered an almost genuine "does" assessment, as no investigator would be present at the operating room. Structured training displayed an advantage by the end of the first training year despite the same number of performed laparoscopic cholecystectomies within the two study groups. Structured training is beneficial according to literature [6,16] and results of this study may underpin this assumption. Efficient work-place learning is needed and module based training may be an answer. Training specific performances in a block of 20 planned cholecystectomies within a set time of 4-8 weeks leads to sustained skills compared to a dispersed training. Therefore, this module based approach may be considered in future training programs. Though, difficulties regarding continued engagement in these models must be considered. It is hard to determine if an educational program is good or less good, due to many interpersonal differences in trainees and trainers. Furthermore, the study was carried out in the workplaces among different habits and traditions. It did show promising results in favor of the module based training model in spite of low numbers.

## Conclusion

The conduction of a nationwide RCT showed many difficulties. However, the study added to our knowledge and experience in educational RTC of workplace based learning and could hopefully help other research groups avoid similar problems.

## Disclosures

Charlotte Carlsen, Karen Lindorff-Larsen, Peter Funch-Jensen, Lars Lund, Peder Charles, and Lars Konge have no conflicts of interest or financial ties to disclose.

## Acknowledgement

The research was founded by Central Denmark Region, Aalborg University Hospital and TRYG foundation.

## References

1. Blencowe NS, Parsons BA, Hollowood AD (2011) Effects of changing work patterns on general surgical training over the last decade. *Postgrad Med J* 87: 795-799.
2. Miller GE (1990) The assessment of clinical skills/competence/performance. *Acad Med* 65: S63-S67.
3. Harden RM (2007) Outcome-based education--the ostrich, the peacock and the beaver. *Med Teach* 29: 666-671.
4. Lave J, Wenger E (1991) *Situated Learning. Legitimate Peripheral Participation.* Cambridge University Press.
5. van der Vleuten CMP (1996) The Assessment of Professional Competence: Developments, Research and Practical Implications. *Adv Health Sci Edu* 1: 41-67.
6. Ericsson KA (2004) Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 79: S70-S81.
7. Norcini JJ (2005) Current perspectives in assessment: the assessment of performance at work. *Med Educ* 39: 880-889.
8. Ringsted C, Hodges B, Scherpier A (2011) The research compass': an introduction to research in medical education: AMEE Guide no. 56. *Med Teach* 33: 695-709.
9. Chang L, Hogle NJ, Moore BB, Graham MJ, Sinanan MN, et al. (2007) Reliable assessment of laparoscopic performance in the operating room using videotape analysis. *Surg Innov* 14:122-126.
10. Gumbs AA, Hogle NJ, Fowler DL (2007) Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. *J Am Coll Surg* 204: 308-313.
11. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, et al. (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190: 107-113.
12. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, et al. (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 338: b1802.
13. Konge L, Lehnert P, Hansen HJ, Petersen RH, Ringsted C (2012) Reliable and valid assessment of performance in thoracoscopy. *Surg Endosc* 26: 1624-1628.
14. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, et al (2008) Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG* 115: 908-916.
15. Vassiliou MC, Feldman LS, Fraser SA, Charlebois P, Chaudhury P, et al. (2007) Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. *Surg Innov* 14: 211-216.
16. Palter VN, Orzech N, Reznick RK, Grantcharov TP (2013) Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomized controlled trial. *Ann Surg* 257: 224-230.