Virtual reality training versus blended learning of laparoscopic cholecystectomy

Nickel, Felix; Brzoska, Julia Anja; Gondan, Matthias; Rangnick, Henriette Maria; Chu, Jackson; Kenngott, Hannes Götz; Linke, Georg Richard; Kadmon, Martina; Fischer, Lars; Müller-Stich, Beat Peter

Published in:
Medicine (Baltimore)

DOI:
10.1097/MD.0000000000000764

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Virtual Reality Training Versus Blended Learning of Laparoscopic Cholecystectomy

A Randomized Controlled Trial With Laparoscopic Novices

Felix Nickel, MD, Julia A. Brzoska, MD, Matthias Gondan, PhD, Henriette M. Rangnick, MD, Jackson Chu, MD, Hannes G. Kenngott, MD, MSc, Georg R. Linke, MD, Martina Kadmon, MD, MME, Lars Fischer, MD, and Beat P. Müller-Stich, MD

Abstract: This study compared virtual reality (VR) training with low cost-blended learning (BL) in a structured training program.

Training of laparoscopic skills outside the operating room is mandatory to reduce operative times and risks.

Laparoscopy-naïve medical students were randomized in 2 groups stratified for sex. The BL group (n = 42) used E-learning for laparoscopic cholecystectomy (LC) and practiced basic skills with box trainers. The VR group (n = 42) trained basic skills and LC on the LAP Mentor II (Simbionix, Cleveland, OH). Each group trained 3 × 4 hours followed by a knowledge test concerning LC. Blinded raters assessed the operative performance of cadaveric porcine LC using the Objective Structured Assessment of Technical Skills (OSATS). The LC was discontinued when it was not completed within 80 min. Students evaluated their training modality with questionnaires.

The VR group completed the LC significantly faster and more often within 80 min than BL (45% vs 21%, P = .02). The BL group scored higher than the VR group in the knowledge test (13.3 ± 1.3 vs 11.0 ± 1.7, P < 0.001). Both groups showed equal operative performance of LC in the OSATS score (49.4 ± 10.5 vs 49.7 ± 12.0, P = 0.90). Students generally liked training and felt well prepared for assisting in laparoscopic surgery. The efficiency of the training was judged higher by the VR group than by the BL group.

VR and BL can both be applied for training the basics of LC. Multimodality training programs should be developed that combine the advantages of both approaches.

(Medicine 94(20):e764)

Abbreviations: BL = blended learning, GTS = general technical skills, LC = laparoscopic cholecystectomy, MC = multiple choice, OSATS = Objective Structured Assessment of Technical Skills, POP = pulsating organ perfusion, STS = specific technical skills, VR = virtual reality training.

INTRODUCTION

Laparoscopic surgery provides advantages for the patients compared with open surgery.1 Learning to perform and manage laparoscopic surgical procedures, however, comes with additional difficulties for novice surgeons.2 The learning process is prolonged by the 2-dimensional view, challenging hand-eye coordination, difficult instrument handling, fulcrum effect, and restricted haptic feedback.3 Training of laparoscopic skills and operations outside the operating room is mandatory to reduce operative times and risks.4–8 Currently used training methods for laparoscopic surgery include live animals, cadavers, video and box trainers, and virtual reality (VR) trainers.8 Live animal training is the most realistic training option. But is expensive, ethically questionable, and not widely available.9 Training with cadaveric animal organs can provide excellent tissue feeling to simulate parts of operations but requires supervision by experienced teachers.10 Box trainers are known to be effective for the acquisition of laparoscopic basic skills but do not provide possibilities for training complete operations.11 VR offers training of basic skills, procedural skills, and complete operations in an animated setting with the possibility of monitoring training progress using automatic recording of training parameters. VR training has proven efficacy for not only acquisition of basic skills in laparoscopic surgery, but also for the improvement of clinical operative skills and patient outcome when compared with no training.12,13 VR training is independent of work hour restrictions and is ethically sound. VR training has become mandatory as preparation for surgical residents in some countries for example, Denmark.14 The disadvantages of currently available VR trainers include lack of realism, high costs, and, as a consequence, restricted availability.15

Box trainers offer high availability and low prices compared with current VR trainers.16 Box training usually depends on supervision by experienced trainers and does not offer...
procedural training without the use of cadaveric organs. E-learning has become increasingly popular over the last decades due to the possibilities offered by video recordings of laparoscopic operations in combination with the World Wide Web. It uses electronic media for the presentation of theoretical and practical knowledge. This includes detailed instructions and videos for operative procedures. E-learning is independent of teachers and is highly available at a low cost any time. 7,18 Blended learning (BL) combines the advantages of box trainers with E-learning tools to provide basic skills training as well as procedural training with high availability at a low price. 19–23 Studies comparing box trainers with VR training have not shown clear superiority of one method over the other. However, most of the existing studies had the trainees tested in one of the used training modalities.24–27 A recent meta-analysis by Larsen et al28 showed superior procedural performance for VR simulators offering training of complete operations over simulators offering only basic skills training. Box training offers basic skills training only and lacks procedural knowledge teaching, which can be compensated by E-learning in the BL concept.

The aim of this study was to compare VR training with low-cost BL for teaching the basic performance of laparoscopic cholecystectomy (LC) to laparoscopic novices in a standardized and structured training program within an adequately powered trial. Secondary questions included sex differences, influences of the participants’ personal characteristics on training outcome, and the participants’ opinions about the training modalities.

MATERIALS AND METHODS

Participants

Laparoscopy-naïve medical students in their clinical years of study were invited to participate in this study. Students with ≥2 hours of experience in laparoscopic surgery training were excluded. The participation was voluntary and the participants were allowed to leave the study at any time. The participants received information about the study and informed consent was obtained. The local ethics committee at Heidelberg University approved the study protocol before inclusion of the trainees (S-334/2011). The cadaveric porcine livers used for the simulation of operations were obtained as side products from the local food industry.

Setting and Study Design

The study was designed as a prospective monocentric, 2-arm, randomized trial with 2 active intervention groups (Flow diagram). The randomization to the VR group (n = 42) or the BL group (n = 42) was stratified for sex. The VR and BL groups were compared for the effectiveness of training a basic operation in laparoscopic surgery. Each group employed their training method (VR or BL) for gaining laparoscopic skills. The training consisted of 3 sessions of 4 hours, adding to a total training time of 12 hours per participant in each group. After completing the training, the students participated in a post-test consisting of a knowledge test and a simulated operation on a cadaver model. The study was carried out at the training center for laparoscopic surgery at the Department of General, Visceral and Transplantation Surgery at Heidelberg University.

Training Groups

Virtual Reality Training

The VR group received a total of 12 hours of laparoscopy training using the LAP Mentor II (Simbionix, Cleveland, OH).

The simulator software enables training within 9 laparoscopic basic skills scenarios as well as procedural skills training in the form of partial or complete laparoscopic operations. For this study, LC was the chosen procedure. Aggarwal et al29 provided a detailed description of the simulator tasks and investigated how many repetitions of each task were generally needed to reach expert level. This information was used to develop a structured VR curriculum for the present study. The VR curriculum started with 9 basic skills tasks (9 repetitions each), followed by 4 procedural skills tasks, for example, dissection of Calot’s triangle (3 repetitions each). Afterwards, the participants had to perform 6 virtual cases of LC with different patient histories and anatomical conditions (2 repetitions each). Additional repetitions of each task were performed in the remaining training time.

Blended Learning

The BL group received a total of 10 hours of box training and 2 hours of E-learning for LC training. The participants received practical basic skills training with a conventional box trainer (Karl Storz GmbH, Tuttlingen, Germany). A structured training curriculum with 9 basic skills exercises was used (shown in Table 1 of the supplemental material). For each task, characteristic errors were identified and described in the curriculum. The participants had to document the task completion time and error rate of each task. They repeated each task several times until a plateau phase in the learning curve was reached for task time and error rate. The procedural skills for LC were taught through E-learning using 2 web-based surgical education platforms with specific content concerning LC. The first 4 didactic chapters of the German website www.webop.de LC module included fundamental knowledge about the relevant anatomy, perioperative management, step-by-step approach to the operation, and management of operative complications.17,18,30 The English website www.WebSurg.com module, “Laparoscopic cholecystectomy for symptomatic cholelithiasis with or without cholangiography” provided additional information about the procedure. The final component of E-learning was a video recording of a narrated LC on www.WebSurg.com, “Laparoscopic cholecystectomy: a gold standard case for the dissection of Calot’s triangle.”30 A tutor was present during the E-learning session for assistance and to answer questions.

Skill Testing and Recording of Data

The comparison of the investigated training methods was made with a post-test at the end of the last training session. At the end of the training, every participant had to take a multiple choice (MC) test consisting of 16 questions about LC with 1 correct answer each. MC test results were compared between the BL and the VR groups to assess the effect of E-learning. The surgical post-test was carried out on a Pulsating Organ Perfusion (POP) trainer, a mechanical laparoscopic training device for procedural laparoscopic training with explanted cadaveric animal organs that enables lifelike pulsatile perfusion of the organs with artificial blood.31–33 Participants had to perform an LC on an explanted porcine liver with preserved gallbladder and associated structures. Three specially trained raters, who were blinded to the groups of the participants, evaluated the operative performance. To ensure rater blinding, the participants were instructed to not disclose their training status to the raters. Each rater evaluated the same number of participants out of each group to avoid confounding due to rater-specific differences in evaluation.
The participants’ performances were judged using the Objective Structured Assessment of Technical Skills (OSATS), which is a standardized and validated tool for assessing surgical skills.\(^ {34} \) OSATS consists of 2 subscores each including 7 different surgical criteria. The general technical skills (GTS) score gives an overall non-tas-specific evaluation of each participants’ performance, whereas the specific technical skills (STS) score evaluates the single steps of a specific procedure. Each GTS item is scored from 1 (worst) to 5 (best), whereas each STS item is scored in increments of 2 points from 2 (worst) to 10 (best). Consequently, a maximum of 35 points can be achieved in the GTS score and a maximum of 70 points in the STS score, yielding a maximum total score of up to 105 points. For the purpose of this study, the OSATS items used by Sarker et al\(^ {35} \) were slightly modified. The GTS assessment was supplemented by a further criterion concerning the quantity and quality of required assistance. For the STS assessment, the criteria “access & port insertion” and “extraction of gallbladder” were replaced by “knowledge of procedure” and “quality of product.” The operation time was recorded, and the operation was discontinued when it could not be completed within 80 min; the operation time was then treated as a right-censored observation.

The participants evaluated the laparoscopic simulators and their individual training method by answering a standardized anonymous questionnaire after the post-test. Each training modality was evaluated for several training criteria on 5-point Likert scales (fully agree to fully disagree). Based on this questionnaire, differences in the acceptance of the training modalities were explored.

**Outcome Measures**

**Primary Outcome**

The OSATS total score was defined as the primary outcome measure. It was compared between the VR and BL groups.

**Secondary Outcomes**

Secondary outcomes were the results of the GTS and STS subscales, operation time, and rate of operations completed within 80 min. Furthermore, the results of the MC test were compared between the VR and BL groups. Changes in performance were compared among the participants’ opinions concerning the training modalities. OSATS scores of male and female participants were compared to investigate potential sex differences in laparoscopic performance. Possible correlations between participants’ baseline characteristics, questionnaire evaluations, and surgical performances (MC and OSATS results) were explored to identify factors that potentially influence surgical education, and can thus be taken into account for future research.

**Randomization**

The randomization process was conducted by using 2 computer-generated lists with variable block sizes, stratified for sex. The randomized assignment to the groups was implemented using closed envelopes. The envelopes were opened only after completion of all baseline assessments. The lists and envelopes were generated and handled by 2 persons who were not directly involved in the training, skills testing, and data collection.

**Sample Size Calculation**

Based on data reported by Sarker et al\(^ {35} \), a standard deviation of 7.86 for the mean OSATS total score among novices was anticipated (assuming a medium size positive correlation \( r = 0.5 \) between GTS and STS). A sample size of 80 participants (40 per group) was calculated to detect a standardized effect size of \( d = 0.63 \) at a 2-sided 5% significance level with a power of 80% using a \( t \) test for independent groups. This is equal to a mean difference of 5 points between the BL and VR groups, or slightly >1 criterion on the GTS subscale. Group differences of less than this margin of 5 points on the OSATS scale were considered to be irrelevant. A total of 84 participants were included in the study to allow for 5% dropout.

**Statistical Analysis**

To minimize the possibility of incorrect inputs, 2 independent persons entered the data in parallel. Data were analyzed using R version 3.1.0.\(^ {36} \) Descriptive data are given as absolute frequencies and as mean ± standard deviation. A normal distribution of OSATS mean scores was assumed,\(^ {37} \) so that a \( 2 \times 2 \times 3 \) analysis of variance (Type III sums of squares, alpha 5% two-tailed) was applied to compare the groups for the primary outcome measure (OSATS total score) as well as GTS and STS scores, controlling for the stratification factor Gender and the influence of the 3 raters. The effect of the intervention is reported as difference in marginal means (“least-squares means”) with the corresponding 95% confidence interval. The operation time was analyzed using Cox proportional hazards model, with hazard ratios >1 denoting shorter operation times than in the control condition. Group was again used as the main factor of interest and the analysis was stratified by Gender. Pearson correlation coefficient was applied for correlation analyses between participants’ baseline characteristics, questionnaire evaluations, and surgical performances. Differences were defined as statistically significant if the 2-tailed \( P \) value related to the group comparison was <0.05. Missing values were not imputed (available cases analysis). This decision was based on the consideration that in a real-life environment, the completion of the training and the assessment would be required for all candidates in surgery.

**RESULTS**

**Demographics**

In total, 84 medical students participated in the randomized study (42 per intervention group) and all of them completed the entire study protocol between September 15\(^ {th} \) 2012 and February 15\(^ {th} \) 2013. There were 21 female and 21 male participants in each group. All participants were novices with no laparoscopic experience and limited experience in open surgery (Table 1).

**Surgical Test Performance**

**Virtual Reality Versus Blended Learning**

Students in the VR training group and in the BL group showed equal performance of LC in the primary outcome OSATS (49.4 ± 10.5 vs 49.7 ± 12.0, mean difference = 0.3 OSATS units, 95% confidence interval −4.7 to 5.3, \( P = 0.92 \)). Figure 1 shows box plots with the OSATS total and component scores per group. The VR and BL groups were equal for GTS (16.7 ± 5.0 vs 16.2 ± 4.9, \( P = 0.63 \)), STS (32.7 ± 6.7 vs 33.5 ± 8.2, \( P = 0.89 \)), and for the OSATS items (Table 2). No significant differences between the 3 raters were observed for the total and component scores of the OSATS.
The average operation time was significantly shorter for the VR group with 75.8 ± 7.1 min compared with the BL group with 77.6 ± 7.0 min (hazard ratio = 2.39, \( P = 0.03 \)). The BL group completed 9 operations compared with 19 in the VR group (21% vs 45% completed operations, \( P = 0.02 \)). In contrast, the score of the knowledge test about LC was significantly better for the BL group than for the VR group (13.3 ± 1.3 vs 11.0 ± 1.7 out of 16 total points; \( P < 0.001 \)).

The average operation time was significantly shorter for male students with 75.0 ± 1.3 min compared with the female students with 78.4 ± 0.1 min (hazard ratio = 2.46, \( P = 0.03 \)). The female students completed 9 operations within the allowed 80 min compared with 19 completed operations for the male students (21% vs 45% proportion of completed operations, \( P = 0.02 \)). Male and female participants did not show significant differences in mean OSATS scores (50.3 ± 11.3 vs 48.8 ± 11.2; \( P = 0.59 \)). There was no significant difference between males and females for the STS score (32.8 ± 7.1 vs 33.4 ± 7.3; \( P = 0.79 \)). A trend toward better performance of males in the GTS score was observed (17.5 ± 4.7 vs 15.5 ± 5.0; \( P = 0.09 \)). This effect was < 1 GTS item (4 units) and was due to slightly better scores in the 3 STS items “time and motion,” “flow of operation and forward planning,” and “extent of support needed.”

### Evaluation of Training Methods by the Participants

#### Laparoscopic Simulators

The POP trainer used for the test received the best overall evaluation compared with the VR trainer and box trainer (Supplemental content Figure 1). Both the VR and box trainer were seen as generally helpful for laparoscopic training and helpful for training of hand–eye coordination. Concerning training for instrument coordination, the box trainer was evaluated better than the VR trainer (\( P = 0.04 \)). However, the VR trainer was found to be more helpful than the box trainer for the more procedural issues: “Exact simulation of intraoperative situation” (\( P < 0.001 \)), “Helpful for training tissue preparation” (\( P = 0.005 \)), and “Helpful for training complete operations” (\( P < 0.001 \)).

#### Individual Training Method

Each group evaluated their training method subjectively regarding training benefit, effects on training motivation, and changes in surgical interest, whereby no significant difference between the VR and BL groups was found. Both groups rated their individual training method as fun (Supplemental content Figure 2). Male participants felt significantly better prepared for the assistance (\( P = 0.008 \)) and performance (\( P < 0.001 \)) of basic laparoscopic operations than female participants. Further items concerning evaluation of training methods did not show any sex differences.

### Correlation Analysis

There was a significant correlation between good OSATS results and seeing the course as good preparation for assistance and performance of laparoscopic surgery, respectively, as relevant for future choice of profession (Table 3). A positive relation between surgical interest and performance was observed for the GTS score but not for total OSATS and STS scores. There was a significant correlation between prior experience in open surgery and the GTS score, but not with the STS and total OSATS scores. Concerning the other questionnaire evaluation issues, no correlations with OSATS results were found. Concerning the MC test results, there was a significant correlation (\( r = 0.24; \ P = 0.03 \)) between the STS item “gall bladder retraction and exposure of Calot triangle” and the MC question on the safe exposure and use of cautery near critical structures within Calot triangle.

### DISCUSSION

In the present study, the students in the VR group performed the operation faster and were more often able to complete the operation within the maximum allowed time of 80 minutes in comparison to the BL group. However, the BL...
The VR and BL groups showed equal operative performance in the OSATS score, which was the primary outcome of this study. Male students were more often able to complete the operation within the maximum allowed time of 80 min, whereas the operative performance of LC in the OSATS score was similar for male and female trainees. The time limit of 80 min was set for feasibility reasons of the study. In preliminary tests, the time limit had proven sufficient for the students to succeed in performing the main steps of Calot triangle exposure, clipping and dividing of cystic duct and artery. The students who did not finish the LC in the time limit were all in the phase of dissecting the gall bladder from the liver bed. Due to the anatomy of the porcine livers that differs slightly from humans in that the gall bladder can be very encased in the liver bed, it would often have taken more time to finish the LC. However, the time was always sufficient to adequately rate the performance. The time differences between the groups and sexes were of statistical significance despite rather small differences. However, the rates of finishing the operation in the time limit were of relevant differences between the groups and sexes.

BL had not been compared with VR training in a randomized trial prior to the present study. Quite a number of studies have been performed comparing VR training and box training without showing clear superiority of one method over the other.

<table>
<thead>
<tr>
<th>General technical skills</th>
<th>VR Group</th>
<th>BL Group</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>2.4 ± 0.8</td>
<td>2.5 ± 0.9</td>
<td>0.61</td>
</tr>
<tr>
<td>Time and motion</td>
<td>2.0 ± 0.9</td>
<td>1.9 ± 1.0</td>
<td>0.52</td>
</tr>
<tr>
<td>Instrument handling</td>
<td>2.4 ± 1.0</td>
<td>2.1 ± 1.0</td>
<td>0.13</td>
</tr>
<tr>
<td>Knowledge of instruments</td>
<td>2.6 ± 0.8</td>
<td>2.7 ± 0.9</td>
<td>0.60</td>
</tr>
<tr>
<td>Use of assistants</td>
<td>2.5 ± 1.0</td>
<td>2.4 ± 1.0</td>
<td>0.62</td>
</tr>
<tr>
<td>Flow of operation and forward planning</td>
<td>2.5 ± 1.1</td>
<td>2.3 ± 0.9</td>
<td>0.28</td>
</tr>
<tr>
<td>Extent of support needed</td>
<td>2.3 ± 1.0</td>
<td>2.2 ± 0.8</td>
<td>0.76</td>
</tr>
<tr>
<td>GTS score</td>
<td>16.7 ± 5.0</td>
<td>16.2 ± 4.9</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Specific technical skills

| Gall bladder retraction and exposure of Calot’s triangle | 4.2 ± 1.7 | 4.5 ± 1.5 | 0.43 |
| Cystic duct dissection | 5.6 ± 1.4 | 5.2 ± 1.8 | 0.19 |
| Cystic duct clipping and transaction | 5.0 ± 1.9 | 5.0 ± 2.0 | 0.94 |
| Cystic artery dissection | 5.1 ± 1.9 | 5.2 ± 1.9 | 0.69 |
| Gall bladder fossa dissection | 3.9 ± 2.1 | 4.5 ± 2.1 | 0.21 |
| Knowledge of procedure | 4.4 ± 1.4 | 4.6 ± 1.5 | 0.52 |
| Quality of product | 4.5 ± 1.6 | 4.5 ± 1.6 | 0.82 |
| STS score | 32.7 ± 6.7 | 33.5 ± 8.2 | 0.63 |
| Total OSATS score | 49.4 ± 10.5 | 49.7 ± 12.0 | 0.90 |

Note Objective Structured Assessment of Technical Skills (OSATS), general technical skills (GTS, 5-point Likert scales, 1 = worst, 5 = best), Specific technical skills (STS, 10-point Likert scale, 2 = worst, 10 = best). VR = virtual reality training, BL = blended learning.

### TABLE 3. Correlation of Objective Structured Assessment of Technical Skills Scores, Prior Experience, and Questionnaire Evaluation of the Training

<table>
<thead>
<tr>
<th>R = Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Score</td>
</tr>
</tbody>
</table>

| Prior experience in open surgery | 0.14 | 0.23* | 0.05 |
| The course has no relevance for my future profession | −0.29† | −0.19 | −0.33† |
| I am interested in the field of surgery | 0.17 | 0.22* | 0.12 |
| By attending this course, I feel well prepared for being first/second assistant in basic laparoscopic operations | 0.38† | 0.43* | 0.26* |
| By attending this course, I feel well prepared for performing basic laparoscopic operations | 0.33† | 0.43† | 0.21 |

GTS = general technical skills, OSATS = Objective Structured Assessment of Technical Skills, STS = specific technical skills.

* Significant correlation at 0.05 level (2-tailed).
† Significant correlation at 0.01 level (2-tailed).
The authors thank Jonathan D. Hendrie for proofreading of the manuscript.

ACKNOWLEDGMENTS

REFERENCES


