Association for Surgical Education

Does laparoscopic simulation predict intraoperative performance? A comparison between the Fundamentals of Laparoscopic Surgery and LapVR evaluation metrics


Department of Surgery, St. Boniface General Hospital, University of Manitoba, Z3039-409 Tache Avenue, Winnipeg, Manitoba, Canada R2H 2A6

KEYWORDS: Surgery; Laparoscopic; Education; Simulation; Assessment

Abstract

BACKGROUND: Considerable resources have been invested in low- and high-fidelity simulators in surgical training. To our knowledge, no investigation has compared the 2 head to head for operative assessment purposes. The purpose of this study was to assess the Fundamentals of Laparoscopic Surgery (FLS) low-fidelity video trainer and LapVR (high-fidelity virtual-reality simulator) for (1) construct and (2) predictive validity using a human cholecystectomy model.

METHODS: Twenty-six participants performed tasks from the FLS program and the LapVR simulator as well as a human laparoscopic cholecystectomy. Performance was evaluated using FLS and LapVR metrics and the Objective Structured Assessment of Technical Skills previously validated rating scale.

RESULTS: Construct and predictive validity were strongly demonstrated for FLS tasks but only incompletely for LapVR.

CONCLUSIONS: Efforts should be focused on using the well-validated lower-cost FLS video trainer for assessment of laparoscopic skills. The high-cost LapVR remains experimental in resource-constrained training programs.

© 2015 Elsevier Inc. All rights reserved.

This work was supported by a research grant from the faculty members of the Department of Surgery at the University of Manitoba. Ashley Vergis receives grants from Covidien ($170,000 per year). Covidien sponsors the MIS fellowship of the University of Manitoba. He is the program director of the fellowship. The other authors declare no conflicts of interest.

* Corresponding author. Tel.: +1-204-237-2574; fax: +1-204-237-3429.
E-mail address: avergis@sbgh.mb.ca
Manuscript received April 23, 2014; revised manuscript August 1, 2014

Laparoscopic surgery has become widely used and integrated into surgical practice over the last 2 decades. This has presented a unique set of educational challenges for training surgical residents. Laparoscopy uses long instruments and trocars that amplify movements and tremor. Additionally, there is decreased range of motion of the instruments and reduced tactile sensation. The specific technical challenges associated with laparoscopic surgery combined with reduced resident work hours, increased costs of operating room time, increased public awareness, and questions about the ethics of learning basic skills on patients have led to a re-evaluation of surgical education and how surgical skills are taught and assessed in trainees. This has fueled the development of basic skills laboratories and surgical simulation as means of training and assessing residents.

Simulation-based skills training and assessment for laparoscopic surgical skills include lower-fidelity physical box or video trainers and higher-fidelity virtual-reality systems. The physical box trainers use actual laparoscopic instruments and an optical system that allow the trainee to perform tasks under videoscopic guidance. Virtual-reality trainers use computer-generated environments to perform simulated tasks.

Training in laboratory-based settings with either virtual-reality simulators or video trainers results in improvement of skills both in the simulation setting and the operating room. Simulation is now established as an effective method of skill acquisition. At this time, the benefit of one simulation method over the other for training purposes has not been clearly established.

Both video trainers and virtual-reality simulators have demonstrated significant correlations between operative performance and psychomotor performance in laboratory-based settings and can be used to assess operative laparoscopic skills. To our knowledge, no investigation has compared physical box trainers to virtual-reality systems head to head for operative assessment purposes. This has widespread implications for surgical training programs. Currently, most programs in North America use a combination of simulation technology for training and assessing their residents. If either video trainers or virtual-reality simulators are shown to be superior at assessing laparoscopic operative skill, resources could be directed toward using that modality for evaluation purposes. This is particularly important as there are ever-increasing time and fiscal constraints in surgical training.

Methods

Participants

Twenty-six general surgery and urology residents at the University of Manitoba participated in this study. All available residents were approached and included. The University of Manitoba Health Research Ethics Board approved this study, and informed consent was obtained from all participants.

Study design

Data collection occurred between September 2011 and January 2013. Residents did the peg transfer, pattern cutting, ligating loop, and intracorporeal knot-tying manual skills tasks from the Fundamentals of Laparoscopic Surgery (FLS) on the video trainer. Additionally, all participants performed 5 of the level 1 essential tasks on the LapVR virtual-reality simulator (CAE Healthcare, Saint-Laurent, Quebec, Canada) that most closely approximated the FLS tasks (peg transfer, pattern cutting, clip application, needle driving, and knot tying).

All participants also performed a human laparoscopic cholecystectomy. Half of the residents performed the video trainer tasks first, whereas the other half performed the virtual-reality simulator tasks first to control for any potential learning crossover between the 2 simulation modalities. The performance of the laparoscopic cholecystectomy and laboratory-based simulation activities were done as close in time as possible.

Before each session, participants completed a questionnaire outlining their prior laparoscopic experience and any practice on video trainers or virtual-reality simulators, as well as any interim practice between the sessions.

Outcome measures

FLS task performance was evaluated using standardized FLS metrics. Validated LapVR evaluation parameters were manually extracted from the computer. An attending surgeon who practiced laparoscopic surgery evaluated the live laparoscopic cholecystectomy using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale. A total of 3 surgeons participated in the evaluation process. All had been instructed on the use of the assessment tool. All forms of outcome measures have been previously validated.

Data analysis

All analyses were carried out using Statistical Analysis System SAS 9.2 (SAS Institute, Cary, NC). Comparisons between junior residents (postgraduate training years [PGYs] 1 and 2) and senior residents (PGYs 3 to 5) were conducted using t tests or the Wilcoxon-Mann-Whitney test for continuous variables and chi-square tests or the Fisher exact test for categorical variables. Multivariate regression analysis was performed to examine the association between outcome variables (OSATS score) and predictor variables (FLS overall score or LapVR data from each domain: peg transfer, cutting, clipping, needle driving, and knot tying). The $R^2$ values from these multivariate regression analyses were used as a measure of how strongly FLS and LapVR...
evaluation metrics predict intraoperative performance. Two-tailed $P$ values of less than .05 were considered statistically significant.

We hypothesized that both forms of simulation would demonstrate construct and predictive validity and that there would not be a significant difference between the 2 methods.

**Results**

**Participant characteristics**

A total of 26 subjects (14 female, 12 male) participated in the study (Table 1). Participants included PGY 1 ($n = 12$), PGY 2 ($n = 4$), PGY 3 ($n = 3$), PGY 4 ($n = 6$), and PGY 5 ($n = 1$). General surgery ($n = 23$) residents of all PGY levels were included and urology ($n = 3$) PGY 1 residents were included. PGY 1 and 2 residents were considered juniors and PGY 3 to 5 residents were considered seniors on the basis of case numbers. The PGY 2 residents in our study performed an average of 15.2 operations as the primary operator and the PGY 3 residents an average of 78.3 procedures. Participant characteristics were generally similar between the junior and senior groups. Half of the juniors and half of the seniors performed the FLS video trainer tasks before the LapVR virtual simulator. The order of the laparoscopic cholecystectomy and simulation activities varied as they were limited by opportunity and scheduling. Seventeen residents performed the simulation tasks before the laparoscopic cholecystectomy and 9 performed the laparoscopic cholecystectomy first. This was evenly divided between junior and senior residents. All previous and interim experience was documented through a questionnaire. Time between activities was also recorded.

**Descriptive statistics of predictor and outcome variables**

Senior residents had significantly higher scores than the junior residents on the overall FLS score and 3 of the 4 FLS tasks (peg transfer, ligating loop, and intracorporeal suturing), with the fourth task (pattern cutting) approaching significance. For the LapVR tasks, there were no significant differences between the senior and junior residents for any of the peg-transfer parameters. For the cutting task, the senior residents scored significantly better on time to completion and unsuccessful cuts. The seniors also scored significantly better on the time to completion for clipping. For needle driving, time and dominant path length were statistically significant, with senior residents outperforming the junior residents. The seniors had significantly better scores on the time to completion and dominant path length knot-tying parameters, but the juniors scored significantly better on nondominant path length (Table 2; time measured in seconds, path length in meters). Senior residents also had significantly higher OSATS scores than junior residents (Table 2).

**Multivariate regression analysis for OSATS**

Overall FLS scores were positively and significantly associated with OSATS scores. All 4 of the individual FLS tasks also demonstrated significance. For LapVR, cutting,

![Table 1: Participant characteristics](https://example.com/table1.png)

**Table 1** Participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (n = 26)</th>
<th>Juniors (n = 16)</th>
<th>Seniors (n = 10)</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>29.7 (3.4)</td>
<td>29.3 (3.9)</td>
<td>30.5 (2.5)</td>
<td>.19</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>14.0 (53.9)</td>
<td>10.0 (62.5)</td>
<td>4.0 (40.0)</td>
<td>.42</td>
</tr>
<tr>
<td>M</td>
<td>12.0 (46.1)</td>
<td>6.0 (37.5)</td>
<td>6.0 (60.0)</td>
<td></td>
</tr>
<tr>
<td>Handedness, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>2.0 (7.7)</td>
<td>1.0 (6.3)</td>
<td>1.0 (10.0)</td>
<td>.99</td>
</tr>
<tr>
<td>R</td>
<td>24.0 (92.3)</td>
<td>15.0 (93.7)</td>
<td>9.0 (90.0)</td>
<td></td>
</tr>
<tr>
<td>Laparoscopic experience (procedures as primary operator), mean (SD)</td>
<td>60.8 (83.5)</td>
<td>4.8 (9.1)</td>
<td>150.4 (68.9)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Video/box trainer experience (hours), mean (SD)</td>
<td>10.1 (29.7)</td>
<td>10.5 (37.2)</td>
<td>9.3 (12.0)</td>
<td>.09</td>
</tr>
<tr>
<td>Virtual simulator experience (hours), mean (SD)</td>
<td>.3 (.5)</td>
<td>.4 (.6)</td>
<td>.3 (.4)</td>
<td>.69</td>
</tr>
<tr>
<td>Time between activities, (days), mean (SD)</td>
<td>93.1 (113.7)</td>
<td>149.4 (112.9)</td>
<td>3.1 (4.2)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interim laparoscopic experience (procedures as primary operator), mean (SD)</td>
<td>1.7 (3.0)</td>
<td>2.1 (3.2)</td>
<td>1.0 (2.8)</td>
<td>.03</td>
</tr>
<tr>
<td>Interim video/box trainer experience (hours), mean (SD)</td>
<td>.7 (1.2)</td>
<td>1.1 (1.3)</td>
<td>.0 (.0)</td>
<td></td>
</tr>
<tr>
<td>Interim virtual simulator experience (hours), mean (SD)</td>
<td>.1 (.3)</td>
<td>.1 (.3)</td>
<td>.0 (.0)</td>
<td></td>
</tr>
</tbody>
</table>

Bold face highlights significant values.

$F = $ female; $L = $ left; $M = $ male; $R = $ right; SD = standard deviation.
clipping, and knot tying were significantly associated with operative OSATS scores (Table 3).

Comment

To be able to decide if one form of simulation is superior to the other, we must first look at the validity of both forms of simulation. The primary outcome of this study was predictive validity with the secondary outcome being construct. Both are 2 of the more objective forms of validity.

Construct validity is the degree to which inferences made by a measurement tool actually represent the theoretical “construct” being investigated. Construct validity was demonstrated for overall FLS score, with the senior group scoring significantly higher than the junior group. It was also demonstrated for the OSATS evaluation of the laparoscopic cholecystectomy. These findings are consistent with the literature and support past research using these forms of assessments in these contexts.3,22,29–31

In the classic model of validity, construct validity is one of 3 main types of validity. Over the last 2 decades, there has developed a more unified concept where construct validity embodies the whole of validity with multiple facets.32 Uniform acceptance of one theory does not currently exist.

For the LapVR tasks, there were significant differences demonstrated between the junior and senior residents for all the clipping and knot-tying parameters. However, for one of the knot-tying parameters, nondominant path length, the junior residents had significantly better scores than the senior residents. This is not expected, as senior residents should have a more efficient path length than the juniors. A possible explanation is that junior residents who are not usually as efficient at coordinating efforts between their 2 hands may be neglecting their nondominant hand, thus giving a falsely improved score.33

None of the parameters evaluated for the peg transfer were significant. This is in contrast to the study validating LapVR by Iwata et al.27 Possible reasons for this could be that the peg-transfer task was the first task that the residents performed.
As majority of the residents had spent minimal time on a virtual simulator, a lack of familiarity with the technology could be partially responsible. Second, at our institution, the peg-transfer task seemed to be the task the junior residents practiced most often on the video trainer. A crossover training effect between the video trainer and virtual simulator has been demonstrated, and this may have also been a factor.

Two of the parameters measured for the cutting task, nondominant hand path length and comprehensive score, showed no significant difference between the junior and senior residents. This is again in contrast to previous findings by Iwata et al. Junior residents would be expected to be less adept at skillfully manipulating their nondominant hand when compared with the seniors and should have an overall lower success rate. This could be a result of poor performance by the senior residents, and the sample size was not large enough to show a true effect. One of the validated needle-driving parameters, nondominant hand path length, also did not show a significant difference. This was the most difficult task according to participants, and many senior residents struggled to orient their instruments correctly. The lack of significance could also be related to sample size.

Construct validity for LapVR was completely demonstrated for 1 of the 5 tasks, partially demonstrated for 3 of the 5, and not demonstrated for 1 of the 5. Compared with that for FLS, construct validity was much weaker and less thoroughly demonstrated for LapVR. This was contrary to our hypothesis. LapVR performance was demonstrated on overall FLS score and OSATS, with the overall FLS score explaining 45.1% of the variation in OSATS operative score. This is consistent with previous literature. These findings again support and add to the body of literature in this area.

The LapVR tasks had to be compared with the overall operative score individually as currently there is no validated overall summary score for the tasks as there is with FLS. Additionally, the individual LapVR tasks could not be compared with the individual FLS tasks as each LapVR task had anywhere from 1 to 4 scoring parameters contributing to the task evaluation. Looking at the relationship between individual LapVR tasks and OSATS scores, clipping, knot tying, and cutting were statistically significant, explaining 45.1%, 35.9%, and 39.9% of the variation in OSATS operative scores, respectively. Therefore, predictive validity was demonstrated between 3 of the 5 LapVR tasks and OSATS scores. Both predictive and discriminate validity have been demonstrated for virtual-reality simulators; however, the MIST VR (Mentice Inc., San Diego, CA) and LapSim (Surgical Science, Goteborg, Sweden) virtual-reality simulators are the most frequently studied.

To our knowledge, this is the first study using LapVR specifically. In this study, predictive validity was incompletely demonstrated for LapVR and strongly demonstrated for FLS. This was in contrast to our hypothesis. LapVR performed more poorly than was expected. Potential reasons for this such as the lack of a task-specific summary metric and overall summary score have been outlined previously.

When looking at an assessment tool, it is also important to look at the feasibility of the evaluation. The video trainer and FLS are practical, easily portable, and relatively inexpensive. The average setup for the video trainer and FLS practice materials is approximately $2,000. In comparison, the virtual-reality simulator is more difficult to transport, requires technical support, and costs on average $80,000 to $120,000 to purchase, as well as additional maintenance and software costs. These are important considerations to factor in when deciding on an assessment tool.

The present study has some limitations. First, the results are drawn from a single institution with 26 participants. Although this number is consistent with related educational literature and recommended guidelines, a multicenter study with a larger sample size would help strengthen the results. Second, there is no overall comprehensive LapVR score as in FLS. Therefore, individual LapVR tasks must be compared with the overall FLS score. The development of an overall LapVR score would allow a more direct comparison between the 2 forms of simulation. Third, there was some variability

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Multiple correlations between OSATS scores and predictors from multiple regression analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall (n = 26)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
</tr>
<tr>
<td>FLS</td>
<td></td>
</tr>
<tr>
<td>Overall score</td>
<td>.415</td>
</tr>
<tr>
<td>Peg transfer</td>
<td>.254</td>
</tr>
<tr>
<td>Pattern cutting</td>
<td>.262</td>
</tr>
<tr>
<td>Ligating loop</td>
<td>.206</td>
</tr>
<tr>
<td>Intracorporeal suturing</td>
<td>.377</td>
</tr>
<tr>
<td>LapVR</td>
<td></td>
</tr>
<tr>
<td>Peg transfer</td>
<td>.172</td>
</tr>
<tr>
<td>Cutting</td>
<td>.399</td>
</tr>
<tr>
<td>Clipping</td>
<td>.451</td>
</tr>
<tr>
<td>Needle driving</td>
<td>.257</td>
</tr>
<tr>
<td>Knot tying</td>
<td>.359</td>
</tr>
</tbody>
</table>

Bold face highlights significant values.

FLS = Fundamentals of Laparoscopic Surgery.
in the time between laparoscopic cholecystectomy and simulator tasks. Given the complicated nature of surgical resident schedules, opportunity sampling was used. Thus, setting up a specific time frame between the activities was not possible. There was a longer time interval for the junior than the senior residents; however, many of the junior residents were rotating on nonsurgical off-service rotations, and it was felt that a longer interval between activities for the juniors would not be as detrimental as a significant period of time between tasks for the seniors. Additionally, the number of interim laparoscopic procedures was relatively low for both the senior and junior resident groups. Therefore, this should have had an insignificant impact on the results of this project.

Conclusion

Both construct and predictive validity were more thoroughly demonstrated for FLS and the video trainer than the LapVR virtual-reality simulator. Overall FLS score explains more of the variation in operative score than the LapVR tasks. It is hard to recommend using the LapVR virtual-reality simulator for assessment of laparoscopic skills without further research, development, and evaluation, especially when taking immense costs into account. On a practical level and in terms of wide implementation, most training programs may continue to be better served using well-validated, lower-cost, more feasible forms of simulation like FLS for laparoscopic skills training and assessments at this time.

References