LACK OF TRANSFER OF SKILLS AFTER VIRTUAL REALITY SIMULATOR TRAINING WITH HAPTIC FEEDBACK

Running title: Study on skills transfer from virtual reality simulator

Cecilie Våpenstad (MSc)* 1,2,3,4,5, Erlend Fagertun Hofstad (MSc) 2,4,5, Lars Eirik Bø (MSc) 1,2,3,4,5, Esther Kuhry (PhD) 1,6, Gjermund Johnsen (MD) 4,6, Ronald Mårvik (PhD) 1,4,5,6, Thomas Langø (PhD) 2,5, Toril Nagelhus Hernes (PhD) 1,5

1Norwegian University of Science and Technology (NTNU), Faculty of Medicine, Trondheim, Norway
2SINTEF Technology and Society, Department of Medical Technology, Trondheim, Norway, 3The Central Norway Regional Health Authority, Trondheim, Norway
4St. Olavs Hospital, Trondheim University Hospital, The Norwegian National Advisory Unit for Advanced Laparoscopic Surgery, Trondheim, Norway
5St. Olavs Hospital, Trondheim University Hospital, The Norwegian National Advisory Unit for Ultrasound and Image-Guided Therapy, Trondheim, Norway
6St. Olavs Hospital, Trondheim University Hospital, Department of Gastrointestinal Surgery, Trondheim, Norway

Abstract:

**Background and Objective:** Virtual Reality (VR) simulators enrich surgical training and offer training possibilities outside of the operating room (OR). In this study, we created a criterion-based training program on a VR simulator with haptic feedback and tested it by comparing the performances of a simulator group against a control group.

**Methods:** Medical students with no experience in laparoscopy were randomly assigned to a simulator group or a control group. In the simulator group the candidates trained until they reached predefined criteria on the LapSim® VR simulator (Surgical Science AB, Sweden) with haptic feedback (Xitact™ IHP, Mentice AB, Sweden). All candidates performed a cholecystectomy on a porcine organ model in a box trainer (the clinical setting). The performances were video rated by two surgeons blinded to subject training status.

**Results:** In total, 30 students performed the cholecystectomy and had their videos rated (N=16 simulator group, N=14 control group). The control group achieved better video rating scores than the simulator group (p<0.05).

**Conclusions:** The criterion-based training program did not transfer skills to the clinical setting. Poor mechanical performance of the simulated haptic feedback is believed to have resulted in a negative training effect.

**Keywords:** virtual reality, simulator, haptic feedback, laparoscopy, surgical education

* Corresponding author: Cecilie Våpenstad, SINTEF Technology and Society, Department of Medical Technology, PB 4760 Sluppen, 7465 Trondheim, Norway, e-mail: cecilie.vapenstad@sintef.no
Introduction

The ability to perform minimally invasive surgery safely is dependent on technical skills, an essential aspect of surgical proficiency [1, 2]. Both inanimate and animate models have been developed to train, the often counter-intuitive skills of minimally invasive surgery outside of the operating room (OR) [3]. The inanimate models are usually referred to as simulators, and the complexity of these varies from video boxes with off-the-shelf web cameras, sticks and holes, to advanced virtual reality (VR) simulators [4]. Today, simulators are valuable tools for training and assessment of technical skills [1, 4, 5]. A VR simulator with objective assessment tools can ensure consistent and proficiency-based training by having the trainees train until they reach criterion-based levels [2, 6]. It also enables exact replication of set-ups, repetitive training with limited need for preparation, and thus fair and reliable tests of the candidates [6]. Several studies show that skills acquired in a simulated environment transfer to clinical settings [1, 5, 7]. However, whether “training on simulators transfer to improved clinical performance”, is still a high priority research area [8].

A VR simulator is composed of a computer together with physical interfaces that represent surgical instruments and instrument ports. Together they create a virtual surgical environment with which the user can interact [9-11]. In addition, most VR simulators have teaching resources with assessment tools that track and display metrics related to time, instrument handling and predefined errors [6, 9]. An important factor of VR simulation is the reproduction of technical challenges that a surgeon encounters during surgery, e.g. challenges related to touch and proprioception, i.e. haptic feedback [12-15]. If the simulator reproduces technical aspects that are relevant for a surgical setting, and the trainee is exposed to them when training on the simulator, one might expect that skills acquired and objectively assessed
on the simulator can be transferred to, and measured, in a clinical setting. This is what is tested in a predictive validity study [16]. Several VR simulators have shown transfer of skills from the simulated environment to a clinical setting [1, 5, 7]. However, these results are only valid for specific simulator set-ups and training curricula [17].

We created a criterion-based training program using the LapSim® VR simulator from Surgical Science LTD and performed a study on its predictive validity. Criterion-based training, as opposed to time-based training, makes sure that the trainees all attain the same level on the simulator [1]. To our knowledge this is the first study that examines predictive validity of the LapSim® simulator equipped with the Xitact® IHP handles with haptic feedback [7].

**Materials and Methods**

Medical students were randomized into two groups, where one group underwent a training program on the laparoscopic simulator and the other did not receive any practical training. The laparoscopic skills of all candidates were tested on a simulated clinical setting using a porcine organ model in a box.

**Subjects**

All aspects of the study were approved by the Norwegian Data Protection Agency, and all subjects gave written informed consent to participate. To have a homogeneous group of participants, they were all recruited among medical students in their 5th or 6th year of study, or interns in their first year of practice after medical school. None of the participants had any experience with laparoscopic surgery. At initial enrolment, the participants answered a
questionnaire with background information, and they were randomly assigned to either a simulator group or a control group.

**VR Simulator training**

The simulator used in this study was the LapSim® VR simulator (Surgical Science Ltd., Gothenburg, Sweden). The system consisted of a software program (LapSIM 2009) running on a computer with the Windows XP operating system (Microsoft Corporation, Redmond, WA, USA), a 3 GHz Intel Core 2 Duo processor (Intel Corporation, Santa Clara, CA, USA), 3.25 GB RAM and a GeForce 8600 GTS graphics card (NVIDIA Corporation, Santa Clara, CA, USA). The system had a 19-inch TFT monitor, a diathermy foot pedal interface and two Xitact® IHP haptic feedback instrument ports (Mentice AB, Gothenburg, Sweden) (Fig. 1).

![Fig. 1 The VR simulator set-up](image)

The criterion-based training program was put together using five basic tasks: *coordination, clip applying, lifting and grasping, fine dissection* and *pattern cutting* (Table 1). The settings configuration and passing levels of each task was set as presented in table 1. Passing levels for the assessed parameters (Table 1) were derived from the work of Ahlberg et al. [1] and a pilot study with experienced surgeons that we conducted prior to this study. The participants in the simulator group performed each task until they attained the respective criteria.
Table 1 The five basic tasks with configurations and passing levels

<table>
<thead>
<tr>
<th>Task</th>
<th>Configuration</th>
<th>Passing levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordination</td>
<td>5 balls, 10 mm ball size, no wide spread, scope angle 0°, field of view 60°, left camera hand, no ball timeout</td>
<td>20 s, no misses, 1.3 m instrument path length, 275° instrument angular path, no instrument outside view, 0.2 m camera path length, 65° camera angular path, maximum one tissue damage with maximum 1 mm damage</td>
</tr>
<tr>
<td>Clip Applying</td>
<td>no moving camera, clip and area size on the vessels 3 with low stretch sensitivity, no spontaneous bleeding, no exercise timeout</td>
<td>55 s, no incomplete target areas, no badly placed clips, no dropped clips, no maximum stretch damage and no blood loss, 5 and 3 m left and right instrument path length respectively, 300° left and right angular path length respectively</td>
</tr>
<tr>
<td>Lifting and Grasping</td>
<td>no moving camera, no rotation, 3 left objects and 3 right objects, object size 10 mm, target size 20 mm and timeout after 12 s</td>
<td>60 s, no instrument misses, no tissue damage, 1.5 m left and right path length respectively, 320° left and right angular path length respectively</td>
</tr>
<tr>
<td>Fine Dissection</td>
<td>2 blood vessels, 3 small vessels around each blood vessel, stretch sensitivity 3 on a scale from 1-5, grasper in left hand, thermo hook as cutter instrument, no moving camera, no rotation, no timeout</td>
<td>100 s, no ripped or burned blood vessels, no energy damage on blood vessels, burned all small vessels with proper stretch, no instrument outside view, 0.5 m grasper and cutter path length respectively, 80° grasper and cutter angular path length respectively</td>
</tr>
<tr>
<td>Pattern Cutting</td>
<td>no contour rotation, contour size 30 mm, 20 mm cut tolerance</td>
<td>110 s, no minimum nor maximum cut error, no mean cut error, no edge cuts</td>
</tr>
</tbody>
</table>

**Clinical evaluation of skills**

Laparoscopic skills of the candidates were evaluated, using a porcine organ model in a box simulating a cholecystectomy (the clinical setting). All candidates in both groups were given a theoretical lecture of about one hour on the procedural steps; they observed an expert surgeon that performed part of the procedure in the box model; and with the help of one of the researchers they had ten minutes to get acquainted with the instruments inspecting them and testing their functionality by e.g. opening and closing them. The instruments they used were
graspers (Endo Clinch II), a clip applier (Endo Clip II ML 10 mm), an ultrasound hook (Auto Sonix Hook Probe 5 mm), all from Covidien Ltd. (Dublin, Ireland), and scissors (Metzenbaum, Ergo handle, 19 mm jaws) from Olympus GmbH (Hamburg, Germany). The box model consisted of a pig liver with an intact gallbladder placed in a Pulsating Organ Perfusion trainer (Optimist Hg.m.b.H, Innsbruck, Austria) (Fig. 2) (The flow functionality was not in use).

The participants were asked to expose and open Calot’s triangle, clip and cut the cystic artery and the cystic duct and remove the gallbladder from the liver. The camera was held in a fixed position during the test by a mechanical camera holder. The laparoscopic videos were recorded and rated independently by two expert surgeons blinded to training status. The expert surgeons used the parameters depth perception, bimanual dexterity, efficiency and tissue handling of the Global Operative Assessment of Laparoscopic Skills (GOALS) tool [18]. The video rating was based on the full length of each video and each parameter was scored from one till five.

![Fig. 2 The simulated clinical setting. The candidates performed a cholecystectomy on a porcine organ model.](image)

Statistical analysis

The data were analysed using SPSS 20.0 (IBM Corporation, Armonk, USA). The Mann–Whitney U test was used to explore for differences in the distributions of scores between simulator and control group, and between frequent and non-frequent video game players.
Analysis was repeated for each video rater to evaluate consistency across raters. The Spearman’s rank order correlation was computed to investigate correlation between experience with computer games and box model performance, and to investigate correlation between performance on the VR simulator and the box model.

**Results**

Thirty participants were included in the analysis: 16 in the simulator group and 14 in the control group. There were 13 women and 17 men, evenly distributed between the two groups. The participants were under 35 years old, and the age distribution was equal between both groups. One of the candidates in the simulator group had previous experience with VR simulators (between 6-10 hours). Experience with computer games was evenly distributed between both groups. A mean of the two raters was calculated as the video raters had comparable standard deviations (Fig. 3).

The participants in the control group achieved significantly (p<0.05) better scores on the laparoscopic video rating, compared to the participants in the simulator group, for three of the four parameters (depth perception, p=0.025, bimanual dexterity, p=0.031, efficiency, p=0.047) and an average of the four scores (p=0.047), based on the mean of the two video raters (Fig. 3). The fourth parameter, *tissue handling*, did not show a significant difference between the groups (p=0.208). The results were consistent in separate analysis of judgments of both raters, with higher median scores for the control group than for the simulator group (Fig. 4). The candidates in the simulator group completed the criterion-based training program with a median of 79 trials (approximately 3 hours) to pass the test (range 32-162), and there
was a correlation between a low number of trials to pass the VR training program and a high video rating score ($\rho=-0.66$, $p=0.01$).

Fig. 3 Boxplot of scores for depth perception, bimanual dexterity, efficiency, tissue handling and average of the four scores, based on the mean of the two video raters, for the simulator and the control group. Maximum video rating score 5. Statistical significantly differences are marked with (*). The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as stars.

Fig. 4 Boxplot of scores for depth perception, bimanual dexterity, efficiency and tissue handling for video rater 1 and video rater 2. Maximum video rating score 5. Statistical significantly differences are marked with (*). The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles. Outliers are plotted as circles and extreme outliers as stars.
Participants, regardless of simulator and control group, that played computer games (action games, simulation games or games with movements, such as Nintendo Wii or Microsoft Kinect) (N=7) weekly or more had significantly better depth perception than the others on the video rating (p<0.05) (Fig. 5).

Fig. 5 Boxplot of scores for depth perception, bimanual dexterity, efficiency, tissue handling and average of the four scores based on the mean of the video raters, for participants that plays (yes) or do not play (no) action games, simulation games or games with movements (Wii, kinect) more than once a week. Maximum video rating score 5. Statistical significantly differences are marked with (*). The middle band shows the median value, the bottom and the top of the boxes show the 25th and the 75th percentiles, and the ends of the whiskers show the 5th and the 95th percentiles.

Discussion

The OR is not the ideal learning environment to train basic technical skills due to the nature of minimally invasive surgery, ethical considerations, working time directives and the increased focus on efficient surgical production [4, 19]. Therefore, new tools that facilitate training outside of the OR have been developed, such as VR simulators and box trainers [2, 4, 19, 20]. Any simulator that enhances surgical education and eventually surgical safety is useful regardless of complexity [2, 20]. Although several studies show skills transfer from simulators to a clinical setting [5, 7], the question of whether they do so is still one of the highest prioritized questions in this field of research [8]. A validation study is only valid for a specific training set-up, where type of simulator, software, hardware interfaces and curricula
are aspects that play a role. Unfortunately, simulators are often only described by brand name in the literature, without specification of important hardware components, such as instrument ports [9, 17]. In addition, a large number of simulated basic tasks and procedures are not validated. In 2013, Våpenstad et al. found that 33 simulated laparoscopic procedural tasks were available on the market, but only eight of them had been part of a validation study [9]. This makes it difficult for the trainer to create valid curricula based on what is available on the market and validation studies found in the literature. It is positive that an increasing number of procedures are simulated and that improved hardware is introduced on the market, but simulators as a training and assessment tool are not validated per se. It is an ongoing task to validate simulators and simulated tasks prior to implementing them in surgical education, of which this study is an example.

We found, using a previous untested type of handles, that the skills acquired and tested with our simulator set-up, did not transfer to a clinical setting (the box model). Those in the control group performed statistically significantly better than those in the simulator group for three out of four parameters. This is in contradiction to other studies on the LapSim® VR simulator [1, 7]. Hogle et al. [3] found similar results as we did, and were also not able to establish predictive validity for the LapSim® VR simulator. Unfortunately, they did not specify which handles they used. One might ask if the participants in the simulator group that trained on the VR simulator with the specified haptic handles, acquired skills that had a negative impact on their performance in the box model?

Our study had possible weaknesses that may have influenced the results. One is the use of medical students, who may differ more in innate technical abilities than surgical residents who have chosen to become surgeons. The training program did e.g. not include any
procedural tasks, nor did it include training of non-technical skills, which are important parts of surgical competence [20]. The simulator group may have acquired an artificially high self-confidence from the training, or they might have been more stressed because they had the impression that they were expected to perform better. Low sample size and/or limited training time on the VR simulator (median training time approximately 3 hours) might explain lack of power to statistically differentiate when analyzing each rater separately. The simulator training program was based on the study of Ahlberg et al. [1], which showed predictive validity. The main differences between the two studies were that Ahlberg et al. [1] included surgical residents whereas we included medical students; the clinical performance was assessed differently and we used different handles. Ahlberg et al. used a real clinical setting, i.e. cholecystectomies on patients, whereas we used a box model with animal organs. Since we included medical students as candidates, a box model was the closest we could get to a clinical setting. Nevertheless, we believe that the box model resembled a clinical setting close enough to be used in this study. Ahlberg et al. used handles from immersion incorporation and turned off the force feedback during their study [1]. We argue that the most important difference between the set-ups, was the handles, and if we take into account results from other studies that our group has performed on the same simulator set-up (see [12, 17]), the lack of predictive validity in this study seems to be related to the handles. The handles with haptic feedback used in this set-up probably did not simulate reality well enough, and we were therefore not able to establish either construct validity, presented in a previous study [17], nor predictive validity, presented in this study.

Haptic sensations have been found important in laparoscopic surgery and in skills training [13, 14, 21-23]. It has been shown that tasks such as grasping and pulling are better retained when learned on simulators with realistic haptic feedback [22, 24-26]. Most surgeons also believe that haptic feedback is an important part of a VR simulator [12, 27], and surgical
performance scores on a simulator are influenced by whether haptic feedback is simulated or not [15, 28]. In a study by Chmarra et al. [22] the participants either performed three tasks on a box trainer first and then three tasks on a VR simulator without haptic feedback, or the other way around. They found that training on the VR simulator first had a negative effect on one of the box trainer tasks in which force application (pulling and pushing) was required, indicating that the unrealistic haptic feedback, in this case the lack of it, resulted in a negative training effect [22]. Surgeons hands both sense and act upon tissue, as opposed to eyes which only sense the surgical field [27]. Is haptic feedback realism thereby more challenging compared to visual realism? Visual realism can be stylistic forms and still show transfer of skills [7], but VR simulators that do and do not simulate haptic feedback have also shown transfer of skills [5, 7]. The influence of simulated haptic feedback is still not well understood [15]. In this study, we found that training on a simulator with unrealistic haptic feedback seem to have had a negative training effect. Haptic devices try to simulate, in the case of laparoscopy, the sensations felt by kinesthetic receptors in muscles and bones, mediated through the laparoscopic instruments [12, 15, 27]. Although tool-mediated, it has been difficult to simulate realistic haptic feedback [11, 12, 15] and it is usually an expensive add-on to VR simulators [13]. The mechanical performance of the haptic device such as frequency response, fidelity in force reproduction and force resolution can be insufficient [10, 12, 15] and unrealistic friction can be introduced, which we believe had a negative training effect in this study.

In the literature, a frequent topic of interest is the existence of a correlation between experience with computer games and surgical abilities [29-38]. We found in this study a significant positive correlation between experience with computer games and better scores on depth perception (Fig. 4). This is in accordance with the findings of other similar studies [30-
33], whereas yet other studies did not find any statistically significant correlation between computer games experience and better clinical performance [34-38]. A question of interest is whether playing computer games results in better visual-spatial abilities or manual dexterity, or if those who play computer games have innate abilities that make them interested in activities that let them explore their abilities, such as computer games. These innate abilities could then be an advantage when learning surgical skills. We found that those that performed best on the simulator also scored well on the box model, which indicates that they had the advantage of innate abilities when performing on the simulator and when being tested on the box model. Hassan et al. [39] found that those that scored well on spatial perception tests also scored well on the LapSim® VR simulator. The role of innate abilities and (innate) potential to acquire surgical skills are of relevance when discussing selection criteria for surgical training [40].

We found that skills acquired through criterion-based training on a VR simulator did not transfer to a clinical setting. VR simulators as a training and assessment tool has advantages, but to provide educational benefits, training programs need to use validated simulators. This study raises several questions about possible limitations of VR technology, including whether training on VR simulators with unrealistic haptic feedback can have a negative training effect. More studies are needed to investigate the role of simulated haptic feedback, an important part of VR simulation.

**Acknowledgements**

We thank all those who participated in the study, in addition to Kirsten Rønning for help with preparing the box models.
The work has been supported by a grant from the Central Norway Regional Health Authority, SINTEF, the Norwegian University of Science and Technology (NTNU), the Norwegian National Advisory Unit for Advanced Laparoscopic Surgery and the Norwegian National Advisory Unit for Ultrasound and Image-Guided Therapy at St. Olavs Hospital (all Trondheim, Norway).

Disclosure of Interest

Cecilie Våpenstad, Erlend Fagertun Hofstad, Lars Eirik Bø, Esther Kuhry, Gjermund Johnsen, Ronald Mårvik, Thomas Lango and Toril Nagelhus Hernes have no conflict of interest or financial ties to disclose.

References


