



HAL
open science

Achieving Interface and Environment Fidelity in the Virtual Basic Laparoscopic Surgical Trainer

Amine Chellali, Helena Mentis, Amie Miller, Woojin Ahn, Venkata S. Arikatla, Ganesh Sankaranarayanan, Suvranu De, Steven D. Schwaitzberg,
Caroline G. L. Cao

► **To cite this version:**

Amine Chellali, Helena Mentis, Amie Miller, Woojin Ahn, Venkata S. Arikatla, et al.. Achieving Interface and Environment Fidelity in the Virtual Basic Laparoscopic Surgical Trainer. *International Journal of Human-Computer Studies*, 2016, 96, pp.22-37. 10.1016/j.ijhcs.2016.07.005 . hal-01344244

HAL Id: hal-01344244

<https://hal.science/hal-01344244v1>

Submitted on 21 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Achieving Interface and Environment Fidelity in the Virtual Basic Laparoscopic Surgical Trainer

Amine Chellali^{* 1,2}, Helena Mentis^{2,3}, Amie Miller^{2,4}, Woojin Ahn⁵, Venkata S. Arikatla⁵, Ganesh Sankaranarayanan⁵, Suvaranu De⁵, Steven D. Schwartzberg², Caroline G. L. Cao⁶

¹ Department of Computer Engineering, IBISC Laboratory, University of Evry, Evry, France

² Department of Surgery, Cambridge Health Alliance, Harvard Medical School, Cambridge, MA, USA

³ Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA

⁴ Department of Surgery, Wright State University, Dayton, OH, USA

⁵ Center for Modeling, Simulation and Imaging in Medicine, Rensselaer Polytechnic Institute, NY, USA

⁶ Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton, OH, USA

*** Corresponding author:** Dr. Amine CHELLALI

E-mail: amine.chellali@ibisc.fr

Current Address: IBISC Laboratory - University of Evry
40, Rue du Pelvoux
CE1455
91020 Courcouronnes

Phone #: +33-1-69-47-75-33

ABSTRACT

Virtual reality trainers are educational tools with great potential for laparoscopic surgery. They can provide basic skills training in a controlled environment and free of risks for patients. They can also offer objective performance assessment without the need for proctors. However, designing effective user interfaces that allow the acquisition of the appropriate technical skills on these systems remains a challenge. This paper aims to examine a process for achieving interface and environment fidelity during the development of the Virtual Basic Laparoscopic Surgical Trainer (VBLaST). Two iterations of the design process were conducted and evaluated. For that purpose, a total of 42 subjects participated in two experimental studies in which two versions of the VBLaST were compared to the accepted standard in the surgical community for training and assessing basic laparoscopic skills in North America, the FLS box-trainer. Participants performed 10 trials of the peg transfer task on each trainer. The assessment of task performance was based on the validated FLS scoring method. Moreover, a subjective evaluation questionnaire was used to assess the fidelity aspects of the VBLaST relative to the FLS trainer. Finally, a focus group session with expert surgeons was conducted as a comparative situated evaluation after the first design iteration. This session aimed to assess the fidelity aspects of the early VBLaST prototype as compared to the FLS trainer.

The results indicate that user performance on the earlier version of the VBLaST resulting from the first design iteration was significantly lower than the performance on the standard FLS box-trainer. The comparative situated evaluation with domain experts permitted us to identify some issues related to the visual, haptic and interface fidelity on this early prototype. Results of the second experiment indicate that the performance on the second generation VBLaST was significantly improved as compared to the first generation and not significantly different from that of the standard FLS box-trainer. Furthermore, the subjects rated the fidelity features of the modified VBLaST version higher than the early version. These findings demonstrate the value of the comparative situated evaluation sessions entailing hands on reflection by domain experts to achieve the environment and interface fidelity and training objectives when designing a virtual reality laparoscopic trainer. This suggests that this method could be used successfully in the future to enhance the value of VR systems as an alternative to physical trainers for laparoscopic surgery skills. Some recommendations on how to use this method to achieve the environment and interface fidelity of a VR laparoscopic surgical trainer are identified.

Keywords: Iterative design, Interaction design, surgical training, Virtual reality (VR), Simulator fidelity

1 INTRODUCTION

Laparoscopic surgery is a minimally invasive surgical technique that requires surgeons to view the operating field on a monitor in order to manipulate surgical instruments and tissues. This technique differs from open surgery in that the surgeons operate through small incisions, use specialized instruments, and do not have direct vision of their actions. In spite of its many advantages for patients – faster recovery, less damage for tissues, smaller scars, less pain, and less need for drugs postoperatively – laparoscopic surgery requires the surgeon to learn new and challenging skills such as hand-eye coordination skills, translation of two dimensional video images to a three dimensional working area, and dealing with the fulcrum effect (Cao et al., 1996; Cao and MacKenzie, 1997; Crothers et al., 1999; Figert et al., 2001; Nguyen et al., 2013). Because of these technical proficiency challenges, the traditional apprenticeship model “See one, do one, teach one” is insufficient for laparoscopic skills training. Moreover, this model introduces legal, ethical and patient safety issues because the apprenticeship model uses patients as practice platforms (Stassen et al., 2005). Finally, work hour restrictions, the cost of operating rooms time, and surgical complications are additional concerns challenging this longstanding dogma (Roberts et al., 2006; Buckley et al., 2012). Thus, one of the most common methods for training laparoscopic surgeons is the use of simulators or virtual reality trainers. These technologies have been identified as potential methods to reduce risks to both medical students and patients by allowing learning, practice, and testing of skills in a protected environment prior to real-world exposure (Aucar et al., 2005; Roberts et al., 2006; Buckley et al., 2012).

The effectiveness of a virtual reality (VR) training system, though, is determined by the ability of the trainee to transfer the knowledge learned during training to a real world task (Liu et al., 2008). Building efficient VR training systems depends on designing an effective interaction for these systems by ensuring its fidelity – i.e. the similarity between the knowledge taught in a simulator to the one used in the real world environment (Stoffregen et al., 2003). What is currently lacking are guidelines or processes for how to go about achieving fidelity for surgical trainers. There is the strong belief that the development of effective VR trainers is solely an engineering challenge. In fact, much of the engineering approach is technology driven and has been focusing on high (visual) environment fidelity and realism (Drews and Bakdash, 2013). However, the engineering approach has many limitations including the realization of complete realism being far into the future (Gibson, 1971; Stappers et al., 2003; Stoffregen et al., 2003; Kim et al., 2003) and the high cost of high-fidelity simulators (Hopkins, 1974).

One approach to achieving higher-fidelity simulations that achieve training objectives is a well-known process within HCI – multidisciplinary iterative design approaches. There is a growing interest in leveraging the principles of the iterative design for medical systems and VR surgical trainers (Mégard et al., 2009; Forsslund et al., 2009; Johnson et al., 2011; Chellali et al., 2012; Yang et al., 2012; Barkana Erol and Erol, 2013; von Zadow et al., 2013). The guidelines and best practices around iterative design, though, have primarily discussed the development of systems that are not VR, haptic, and 3D-based interaction environments. In particular, addressing how to make the best use of various user groups throughout different stages of the design and development process and how those evaluation points actually affect various aspects of fidelity have not been articulated.

In this paper, we describe our own experience with developing the Virtual Basic Laparoscopic Surgical Trainer (VBLaST). VBLaST is a VR computer-based basic laparoscopic skills trainer that is meant to mimic an existing physical basic laparoscopic skills simulator without the resource limitations. It enables self-administration, savings on consumables, and automatic processing of performance through objective interpretation of results (Arikatla et al., 2013). Aspects such as speed and efficiency of movements can be measured and displayed in real-time for the trainees and the instructor with an objective score and used to assess the learning of the targeted skills. The design objective of the VBLaST system is to obtain a *training performance* not significantly different from that of a validated physical part-task laparoscopic skills trainer while overcoming its limitations. . To achieve this goal, it was important to improve the fidelity of the system and to determine the characteristics of the virtual training environment that are most important for VR simulations to be effective in training laparoscopic skills. From our experience in utilizing an iterative design and development process that capitalized on the felt experience of expert surgeons through a situated comparative method, we are able to show how we determined the appropriate levels of fidelity and thus achieved a comparable training outcome to that of a physical training environment.

2 RELATED WORK

2.1 Types of laparoscopic surgery trainers and related simulators

There are currently two types of laparoscopic surgery simulators: full procedural trainers and part-task trainers.

2.1.1 Full procedural trainers

Full procedural trainers provide a realistic practice environment to conduct an entire surgical procedure. This includes for instance, live animal models and VR procedural simulators. While animal models offer the most realistic, non-patient environment for laparoscopic training, cost and ethical issues are the main reasons why they are not fully integrated into most surgical curricula (Roberts et al., 2006). VR trainers offer an alternative training environment and are currently considered as the most promising tools for surgical simulation. These systems digitally recreate the procedures, the specific anatomy, and the environment of laparoscopic surgery. This allows the practicing of all skills necessary to perform a particular laparoscopic operation. Some examples of commercially available full procedural VR simulators include LAP-Mentor (Simbionix USA, Cleveland, OH) (Simbionix, 2016), LapVR (CAE Healthcare Inc., Sarasota, FL) (CAE Healthcare, 2016), and LapSim (Surgical Science, Göteborg, Sweden) (Surgical Science, 2016). These systems offer training modules for full laparoscopic procedures, such as appendectomy (vermiform appendix removal), cholecystectomy (gallbladder removal), or laparoscopic ventral hernia repair.

The potential advantage of these simulators is that well designed full procedural simulation should teach skills, anatomy, and the nuances of a complete surgical procedure in an environment where errors can be made without consequences. In addition, VR trainers offer the possibility of evaluating the surgical performance based on objective measurements. Indeed, all movements and actions performed by the trainee can easily be recorded by the system allowing for immediate and objective feedback.

Despite these advantages, none of these VR systems is currently adopted as a standard for laparoscopic surgical training (Zevin et al., 2014; Yiannakopoulou et al., 2015). This can be explained by several factors. First, these systems are extremely expensive compared to physical trainers, with the average cost of a VR trainer ranging from approximately \$80,000 USD to \$120,000 USD (Palter, 2011; Steigerwald, 2013). That does not include additional maintenance fees, the technical support, or the purchase of any optional modules for the system. This increase in cost would inevitably lower the availability of these training systems to most hospitals and learning centers (McMains and Weitzel, 2008). In addition, a recent study has shown that even when these systems are available at the teaching hospitals, they are underused because they are physically locked away, perhaps owing to their cost (Brennan et al., 2014). The second issue is related to the lack of validity of measurements used to assess the surgical performance on these systems (Stefanidis and Heniford, 2009; Thijssen and Schijven, 2010; Våpenstad and Buzink, 2013; Stunt et al., 2014). This is perhaps due to the novelty of these systems and suggests that additional research is needed to validate these measurements. Finally, there is

currently a lack of fidelity of the visual and more particularly of the haptic feedback (Iwata et al., 2011; Steigerwald, 2013; Hennessey and Hewett, 2013; Sánchez-Margallo et al., 2014; Zevin et al., 2014). For instance, a study evaluating the haptic feedback of the Lap Mentor trainer has shown that the presence of the haptic feedback has no significant effect on the performance of novice trainees suggesting that a better haptic feedback is still needed on this system (Salkini et al., 2010).

All these issues contribute to a decrease in the acceptance of these systems as efficient training tools and slow down their incorporation into surgical training curricula (Reznick and MacRae, 2006; Steigerwald, 2013; Shaharan and Neary, 2014).

2.1.2 Part-task trainers

Most of the currently existing trainers for basic laparoscopic technical skills are part-task trainers. In contrast to full procedural trainers, these systems are designed to simulate a specific surgical task (such as cutting tissue or suturing), and improve the trainees' psychomotor skills in this task (Forsslund et al., 2011). This is because a surgical procedure (e.g. gallbladder removal) is considered to be a combination of basic tasks (e.g. cutting, grasping, suturing) that can be learned separately and out of context, and then put to use in the operating room during a complete surgical procedure (Johnson, 2007). Although these trainers only cover a small part of the surgical curriculum, they have great value for training basic laparoscopic skills (Giles, 2010; Buckley et al., 2012).

There are currently two types of part-task trainers for laparoscopic surgery; physical box-trainers and VR task trainers (Munz et al., 2004).

Physical (also called video) box-trainers incorporate real laparoscopic instruments providing realistic haptic feedback, enable training on inanimate synthetic models, and can also be coupled with cadaveric material to provide realistic anatomy. They have already shown a great value for training laparoscopic skills (Fried, 2008). For instance, the Fundamentals of Laparoscopic Skills (FLS) box-trainer is the only trainer that has been validated and adopted in North America as the standard for training and assessing basic psychomotor skills needed in laparoscopic surgery (Palter, 2011; Skinner et al., 2013; Valdivieso and Zorn, 2014; Pitzul et al., 2012). This box trainer is the technical component of the FLS curriculum and is based on the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) (Derossis et al., 1998). The five standard tasks of the FLS curriculum are the peg transfer, the pattern cutting, the ligating loop, and intracorporeal and extracorporeal suturing (Fried et al., 2004). The peg transfer task is described in details in section 3.2. The pattern cutting requires cutting a precise circular pattern from suspended gauze along a premarked template. The

learning objective is to develop skills in accurate cutting and applying appropriate traction and countertraction working through the constraints of fixed trocar positions. The ligating loop task aims to learn how to correctly and accurately place an endoloop around a tubular structure (foam appendage). In fact, the endoloop is a convenient tool in laparoscopic surgery to securely control a hollow tubular structure, such as a blood vessel. Finally, the two last tasks aim to teach the trainee how to perform suturing using either an intracorporeal tie (Task 4), or an extracorporeal tying technique with the help of a knot pusher (Task 5). Indeed, placing a stitch with precision and tying a knot securely are essential skills for a laparoscopic surgeon.

The FLS curriculum including the previous five tasks underwent a rigorous validation processes demonstrating its reliability (Vassiliou et al., 2006), construct validity (McCluney et al., 2007), improved technical performance in the operating room as measured using a validated assessment tool (Sroka et al., 2010), and retention of learned skills (Edelman et al., 2010). The Proficiency criteria were set (Fraser et al., 2003) and the FLS curriculum was endorsed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) and the American College of Surgeons (ACS). Currently, completion of the FLS program is an eligibility requirement in order to sit for the American Board of Surgery (ABS) qualifying examination in the United States (Zevin et al., 2014) and SAGES recommends that all senior residents demonstrate laparoscopic skill competency by taking this exam. In addition to showing their validity in training basic laparoscopic skills, the other main advantage of the physical part-task trainers is their relatively low acquisition cost (approximately \$2,000 USD for the FLS box trainer) (Steigerwald, 2013). This makes them the most widely available tools in residency training programs (Steigerwald, 2013). However, these trainers are resource-intensive in terms of consumables which increase the annual training cost. A recent study has demonstrated that the annual cost for training 10 residents using the FLS box trainer can be increased up to \$12, 000 USD (Orzech et al., 2012). More importantly, physical box trainers are time-consuming to administer and process, and subjective when the results are interpreted (Botden and Jakimowicz, 2009; Chellali et al., 2014). In fact, a teacher/evaluator needs to be present to ensure the tasks are being performed properly, to give feedback to the trainees, and to evaluate them.

On the other hand, VR part-task trainers also replicate simple laparoscopic tasks such as cutting, grasping and suturing enabling the operator to learn how to perform basic skills. They are based on three-dimensional computer generated images with which the user can interact through a physical interface. Previous research has shown that virtual task trainers can actually improve intraoperative skills in minimally invasive surgery similar to that of physical task trainers (Seymour et al., 2002; Chellali et al., 2012; Ullrich and Kuhlen, 2012; Yiannakopoulou et al., 2015).

The following commercially available VR trainers incorporate basic laparoscopic tasks comparable to the FLS tasks; LAP-Mentor, and LapVR, LapSim, and Lap-X (Medical-X, Rotterdam, the Netherlands) (Medical-X, 2016).

LAP Mentor consists of a custom hardware interface with haptic feedback. It includes a library of modules designed for training basic laparoscopic skills such as camera manipulation, hand–eye coordination, clipping, grasping, cutting and electrocautery. The system provides also tasks similar to those of the FLS box such as the peg transfer, the pattern cutting and ligating loop tasks. Preliminary validations of these three tasks have been conducted (Pitzul et al., 2012) and the results have shown a moderate concurrent validity for the three tasks with significant differences in completion time between the FLS box-trainer and the Lap Mentor. In addition, the three tasks do not use the same scoring metrics that are used in the FLS. On the other hand, another study comparing the two systems shows no significant difference in training performance between the two systems (Beyer et al., 2011). A more recent study comparing the Lap Mentor to the FLS trainer has shown that the performance on both systems is correlated but not interchangeable (Brinkman et al., 2013). The trainees have preferred the FLS box trainer over the VR system. The authors have concluded that Lap Mentor training alone might not suffice to pass an assessment on the validated FLS box trainer (Brinkman et al., 2013).

The LapVR consists of laparoscopic instruments connected to a desktop computer. The system provides also a set of basic skills comparable to the FLS tasks (peg transfer, cutting and clip application). A recent comparative study has shown the lack of validity of the scoring system of the LapVR tasks as compared to the FLS trainer (Steigerwald et al., 2015). Because of the lack of validity and the high cost associated with the LapVR, the study concluded that the FLS trainer is currently a more cost-effective and a better tool for training and assessing basic laparoscopic skills.

LapSim is another training system that features training for basic laparoscopic skills. The system has different modules mimicking different tasks such as camera navigation, instrument navigation, coordination, grasping, cutting, clip applying and more advanced modules like suturing and “running the small bowel”. Time, instrument path length and procedure specific errors are measured. In a recent study, the performance on this simulator has not been shown to be correlated with the FLS scores (Hennessey and Hewett, 2014). Some comparative studies and reviews have shown also the benefit of the FLS trainer compared to LapSim for training basic laparoscopic skills (Fairhurst et al., 2011; Tan et al., 2012; Hennessey and Hewett, 2014), while another study has concluded that both systems are efficient for training laparoscopic skills with an advantage to the FLS box trainer in regard to its cost effectiveness (Orzech et al., 2012). Indeed, the annual cost of VR training for 5

residents was estimated to \$80,000 USD while it was estimated to \$12,000 for the FLS box training (Orzech et al., 2012).

Lap-X also offers similar FLS tasks like peg transfer, pattern cutting and suturing. Lap-X design is centered on portability and hence features a light user interface with haptic feedback. Therefore, the system is less expensive than the previous VR trainers. However, to the best of our knowledge, Lap-X has not been clinically validated and only a preliminary evaluation study has been conducted recently (Kawaguchi et al., 2014). No comparative study with other existing systems such as the FLS trainer has been conducted to show his concurrent validity. Therefore, more research is needed to validate this system.

In summary, the brief review above shows that both physical box trainers and VR simulators could be efficient for training and assessing basic laparoscopic skills. However, there is currently a significantly larger body of evidence supporting this for physical box trainers than VR simulators. This is confirmed by a recent systematic review showing an advantage for box trainers over VR trainers for training and assessing basic laparoscopic skills regarding the learners' satisfaction and task time (Zendejas et al., 2013). In addition these trainers are less expensive than VR systems. As a consequence, they are the most widely available tools in laparoscopic surgery training programs. For instance, the FLS system which is the only ACS and ABS approved trainer for training and assessing basic laparoscopic skills in North America is a box trainer. However, the evaluation of skills on these systems is subjective and requires the presence of a trained proctor. In this context, VR technologies can overcome this limitation. However, the current VR laparoscopic trainers (both box trainers and full procedural trainers) suffer from some issues such as their high acquisition cost, the lack of validity of their performance measurements, or their lack of visual and haptic fidelity. Our focus in this paper is more particularly on this last issue.

2.2 Impact of simulator fidelity

The effectiveness of a training system is usually determined by the ability of the trainee to transfer the knowledge learned during training to a real world task (Liu et al., 2008). This is particularly true with virtual training environments that replicate real world environments. Nevertheless, building efficient VR training systems depends on designing an effective interaction for these systems. Complex and inappropriate user interfaces make an interactive system likely to be underused or misused with frustrated trainees maintaining their current training methods or not acquiring the targeted skills (Maguire, 2001). To design a system that overcomes some of the previous issues associated with VR surgical trainers, it is important to ensure its fidelity. Simulator fidelity can be defined as the similarity between the knowledge taught in a simulator to the one used in the real

world environment (Stoffregen et al., 2003). From the fidelity perspective, part-task trainers can be considered as low-fidelity simulators and full procedural trainers as high-fidelity simulators. While it is intuitive to think that higher fidelity will naturally be necessary to increase the training effectiveness (Hays and Singer, 1989; Hamblin, 2005), other researchers have shown that low-fidelity trainers could be sufficient for an efficient and effective training experience (Kim et al., 2003; McMains and Weitzel, 2008; Moroney and Moroney, 1999). Our previous review of the existing laparoscopic simulators confirms that low-fidelity box trainers are sufficient and sometimes more suitable for training basic laparoscopic skills than high-fidelity full procedural trainers.

However, it is important to distinguish between two different aspects of simulator fidelity – interface fidelity and environment fidelity (Waller and Hunt, 1998). Interface fidelity is defined as the degree to which the input and output interaction devices used in the virtual training environment function similarly to the way in which the trainee would interact with the real world (Waller and Hunt, 1998). This concept is similar to the concept of interaction fidelity defined by Ragan et al. – the objective degree of exactness with which real-world interactions can be reproduced (Ragan et al., 2015). Interface fidelity plays a central role in the transfer of knowledge from a VR trainer to the real world (Drews and Bakdash, 2013). It is affected by the ease of interaction and level of user control of the system (Hamblin, 2005). For instance, an efficient interaction requires that any action from the operator in the virtual world generates an instantaneous multimodal response from the virtual environment.

On the other hand, the environment fidelity is related to the realism of the virtual environment (Hamblin, 2005). It is defined as the degree of mapping between the real world environment and the training environment. This can be linked to the concept of display fidelity defined by Ragan et al. – the objective degree of exactness with which real-world sensory stimuli are reproduced (Ragan et al., 2015). Environment fidelity depends on a subjective judgment of similarity between the real and the simulated worlds rather than an objectively quantifiable correspondence between values of variables (Waller and Hunt, 1998). It is affected by the system's quality of the visual, auditory, and haptic feedback. However, according to Dieckmann (2009), the realism of a virtual simulator must be a means to serve specific learning objectives of the trainer and not the goal per se of designing a simulation system. In fact, environment fidelity should be used to positively affect the interface fidelity. For instance, learning how to handle tissue in laparoscopic surgery requires knowledge and skills to apply the appropriate amount of force for pulling or cutting. Previous studies have shown that virtual simulators can efficiently train these skills (Kim et al., 2003; Al-Kadi et al., 2012). Moreover, the addition of haptic

feedback and animation to the surgical surface also improves the acquisition of these skills because it allows one to correctly perceive the elasticity and other tissue properties (Basdogan et al., 2004).

However, guidelines or processes for how to go about achieving interface and environment fidelity for surgical trainers have not been fully realized. There is the strong belief that the development of effective VR trainers is solely an engineering challenge. In fact, much of the engineering approach is technology driven and has been focusing on high (visual) environment fidelity and realism (Drews and Bakdash, 2013). For instance, a system that has visual feedback with more realistic graphics is usually preferred over a system with less realistic graphics, even when the later leads to a better interaction and training (Smallman and St. John, 2005). However, the engineering approach has many limitations. First, despite the advances in computer graphics technology and processing power, complete physical realism towards environment fidelity in simulation will not be accomplished in the near future (Gibson, 1971; Stappers et al., 2003; Stoffregen et al., 2003; Kim et al., 2003). Second, high-environment fidelity simulators are usually much more expensive than low-environment fidelity simulators (Hopkins, 1974). As discussed above, this is the case of most available VR laparoscopic trainers. Finally, the relationship between high-environment fidelity simulators and transfer of knowledge to the real world has not been fully established (Drews and Bakdash, 2013). Again, this is confirmed by recent reviews on VR laparoscopic simulators (Våpenstad and Buzink, 2013; Yiannakopoulou et al., 2015). On the contrary, if the environment fidelity in simulation does not match the interaction aspects of performing the actual task; high environment fidelity even may be detrimental to training (Kozolowski and DeShon, 2004).

Cost-effective low-fidelity surgical trainers could be efficient in surgical training (Kim et al., 2003; McMains and Weitzel, 2008). However, the levels of interface and environment fidelity required and the characteristics of the virtual training environment that are most important for VR simulations to be effective in training laparoscopic skills have not been defined. Hence, the design of training systems requires defining the learning objectives of the system, the most important part of the surgical task to simulate, and to what extent it can be simulated successfully (Forsslund et al., 2009). One approach to achieve these requirements is the use of the multidisciplinary iterative design approaches. There is a growing interest in leveraging the principles of the iterative design for medical systems and VR surgical trainers (Mégard et al., 2009; Forsslund et al., 2009; Johnson et al., 2011; Chellali et al., 2012; Yang et al., 2012; Barkana Erol and Erol, 2013; von Zadow et al., 2013). Nevertheless, none of these studies have addressed how to leverage the design process to achieve interface and environment fidelity.

2.3 Objectives

In our development effort, human factors engineers, task designers, software developers, and physicians worked together to design and build the VBLaST system. Despite starting off with a concerted effort to gather a definitive set of requirements, our first prototype had reduced performance in comparison to the physical box trainer. Our previous studies (Arikatla et al., 2013; Chellali et al., 2014) have shown that the system has several limitations and suggested that improvements are necessary. However, the exact changes required to improve the system were not identified.

In the current paper, the objective was to point out these issues which led to a decrease in the trainees' performance/learning on our system. Through a focus group with expert laparoscopic surgeons evaluating the VBLaST system side-by-side with the physical skills trainer, we began to uncover the intangible qualities that contribute to interface and environment fidelity in a virtual skills trainer. The findings of this comparative situated evaluation session were then rolled into a second design iteration of our system and the evaluation of that system proved its validity against the physical skills trainer. We discuss these findings in light of the need for an iterative design process to elicit not only the cognitive and functional requirements from the standpoint of the trainees, but also the reflective 'felt-experience' of the experienced surgeons in order to round out the intangible properties that enhance the simulator's functional validity.

In the following, we describe two separate experiments that were conducted to evaluate the two different iterations in the design process of the VBLaST system. In each experiment, the VBLaST prototype was compared to the accepted standard in the surgical community for training and assessing basic laparoscopic skills, the FLS box-trainer. The peg transfer task, the first of the five standard psychomotor tasks of the FLS program was used as the test case in the current study. All of the studies conducted within the scope of this work were approved by the local Institutional Review Boards (IRB).

3 FIRST DESIGN ITERATION

The first iteration of our design process was conducted by performing observations of the use of the physical task trainer, a task analysis from those observations, a prototype design and implementation and an evaluation session. The observations consisted of observing residents during their training for the peg transfer task, using the physical skills trainer. The goal was to conduct a task analysis to identify the learning objectives, the subtasks and actions necessary to complete the main peg transfer task, the instruments used to perform the task as well as the constraints when performing it.

3.1 Physical task trainer

The FLS box-trainer (Figure 1) includes a set of accessories used to simulate specific basic laparoscopic tasks, a source light and a built-in camera capturing the objects/tools movement while a trainee is performing the surgical tasks. The video is displayed on a monitor for the trainee to watch. For the peg transfer laparoscopic task, a peg board is placed in the center of the FLS box-trainer with 12 pegs and six rings. The six rings are on the left side of the peg board at the start of each trial. Two laparoscopic graspers are used to manipulate the rings.



Figure 1: The FLS box-trainer

3.2 Task to be simulated

The task analysis was validated by two different certified SAGES FLS proctors. The learning objectives of this task are to develop bimanual dexterity, hand-eye coordination, speed and precision. The learners are trained to perform the task as quickly as possible with fewer errors. A hierarchical task analysis (HTA) at the tasks level and based on video analysis was conducted following the same method used in previous studies (Cao et al., 1999; Chellali et al., 2014). The HTA permitted to decompose the FLS peg transfer task into the following subtasks: using two laparoscopic graspers (one in each hand), operators pick up the rings (one at a time) with the left hand, transfer them to the right hand, and place them on the opposite side of the peg board. Once all the rings are transferred, the process is repeated to transfer the rings back to the other side of the peg board to complete one trial. The instrument tool tip can be rotated using a rotation wheel in order to facilitate the manipulation of

the rings. A total of 6 rings with different colors have to be transferred back and forth for each trial. The color and order of transfer of the rings are of no importance. If a ring is dropped on the board, the operator has to pick it up with the same hand and continue the transfer. However, if the ring is dropped outside of the board, the operator has to leave it behind and continue the task. Timing starts once the first ring is grasped and ends after the last ring is transferred.

The previous task analysis suggests that to achieve fidelity of the environment, it is important to replicate the physical workspace by simulating the surgical tools, the rings, the pegs and their interaction in the virtual world. Moreover, the analysis shows that haptic interactions and physical responses of the virtual objects are essential to allow the user to manipulate the virtual objects and to achieve an efficient interaction with the system. These findings were the foundation for the design of the early VBLaST prototype (Arikatla et al., 2013). This prototype is described in the following section. The first iteration concluded with an experimental evaluation session in which the early VBLaST prototype was compared to the FLS box-trainer.

3.3 VBLaST System Design

The early VBLaST prototype (Figure 2) consisted of computational software to simulate the FLS pegs and rings, and a physical user interface to connect two laparoscopic graspers handles (the same as those used for the FLS box-trainer) to two PHANTOM Omni haptic devices (Geomagic Inc., Boston, MA, USA). These haptic devices have 6 Degrees Of Freedom (DOF) positional sensing, 3-DOF force feedback and a removable stylus for end-user customization. They have a workspace size of 160mm width x 120mm height x 70mm depth. These commercially available haptic devices have a relatively low cost (around \$1800 USD each).

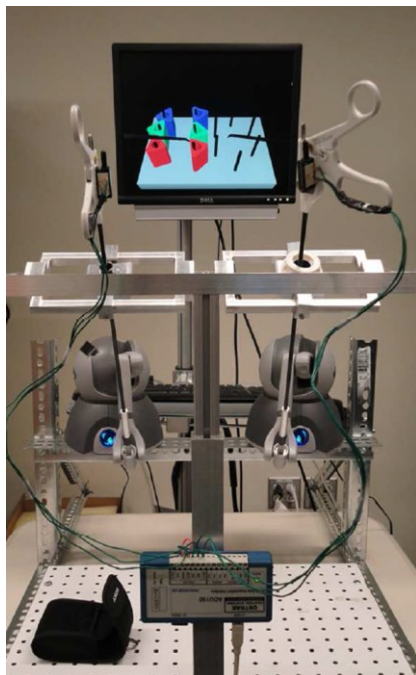


Figure 2: the early VBLaST prototype (VBLaST 1)

Using this interface, the users can hold onto the instruments handles with their hands, moving them in the real space while watching what they are doing inside the virtual space on the computer screen. This replicates the setup of the FLS box trainer where the users move the instruments while watching on a monitor the tools/objects movements captured through a built-in camera. This is also similar to what surgeons experience during laparoscopic surgery and thus ensures the ecological validity of the study (Breedveld and Wentink, 2001). The haptic devices simulate and render contact forces with the virtual objects. Finally, two linear slider potentiometers (one mounted on each handle) were used to allow the user rotating the virtual tool tip using the rotation wheels of the handles (Figure 2).

3.4 First experimental evaluation session

The goal of the first experiment was to verify the design decisions for the early VBLaST version and to get input for modifications in the future iterations. Twenty general surgery residents and attendings (N = 20) were asked to participate in the study (25-55 years old, 13 male-7 female, 2 left handed). Five postgraduate year 1 (PGY1) general surgery residents, five postgraduate year 3 (PGY3) general surgery residents, five postgraduate year 4 (PGY4) general surgery residents, and five fellows and attendings from different hospitals in the Greater Boston Area were recruited and categorized into four groups depending on their expertise level (five subjects in each group), as shown on Table 1.

Table 1: Subjects' groups in the first experiment

Expertise level	PGY1	PGY3	PGY4	Attendings
Group name	Group1	Group2	Group3	Group4
Number of subjects	5	5	5	5

3.4.1 Apparatus

The FLS box-trainer and the early VBLaST prototype were used to perform the peg transfer task in this experiment. A digital video capture device (AVerMedia, Milpitas, CA, USA) was used to record subjects' performance inside the task space for the FLS box-trainer. The video was used to calculate the task completion time and errors (the number of rings dropped outside of the board) for data analysis. The performance measures for the VBLaST were automatically recorded by the system on a log file.

3.4.2 Evaluation procedure

The standard FLS peg transfer task was used in this experiment. Before the start of the experimental session, subjects were asked to complete the consent form and to fill-in a questionnaire detailing the demographics and their previous laparoscopic surgery and laparoscopic simulators experience. The subjects were then shown an instructional video describing the peg transfer task, accompanied by a verbal explanation. Subjects were given one practice trial on each system to become familiar with the procedure and the two systems. Each participant was then asked to perform ten trials of the standard FLS peg transfer task both on the FLS box-trainer and the VR trainer. The study was counterbalanced so that half of the participants performed the task on the FLS box-trainer first, while the other half performed the task on the virtual trainer first. Subjects were assigned randomly to each group. At the end of the session, they were asked to fill-in an evaluation questionnaire and to comment their experience with the system and their opinions about what design aspects of the VBLaST system could be changed.

3.4.3 Evaluation dependent measures

For the FLS box-trainer, the dependent measure (obtained from the captured videos) consisted of a total raw score (ranging from 0 to 300) calculated using the standard metric for the FLS training program (Fraser et al., 2003) obtained from the SAGES FLS committee. This is the current assessment method of the proficiency of

laparoscopic technical skills adopted by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES), the American College of Surgeons (ACS) and the American Board of Surgery (ABS) for general surgeons (Fried, 2008).

The raw score is based both on accuracy and speed (Derossis et al., 1998). The timing score is calculated by subtracting the time to complete the task from a preset cutoff time (300 sec): $\text{timing score} = 300 \text{ (sec)} - \text{time to complete the task (sec)}$. This scoring method rewards faster performance with higher scores. If the time to complete the task surpasses the preset cutoff time, a timing score of zero is given to avoid negative scores. Accuracy is also objectively scored by calculating penalty score: percentage of pegs dropped outside the field of view. Finally a total score combining the two above scores is calculated for each trial performed by subtracting the penalty from the timing score: $\text{raw score} = \text{timing score} - \text{penalty score}$. Thus, the more accurately and quicker a task was completed, the higher the score.

Similar to what have been used in previous studies (Arikatla et al., 2013; Chellali et al., 2014), these raw scores were divided by a normalization factor that was calculated using the best expert FLS score obtained in the FLS condition in the two experiments. The FLS normalized scores ranged from 0 to 100. The same FLS normalization factor was used for experiment 1 and experiment 2 so that the scores are comparable.

For the VBLaST trainer, the raw scores were automatically calculated by the system using the same formula. These raw scores were then divided by a normalization factor calculated using the best expert score obtained from the VBLaST condition in the two experiments. The VBLaST normalized scores also ranged from 0 to 100. The same normalization factor was used for both the early and modified VBLaST prototypes.

Finally, the subjects evaluated different features of the VBLaST system relative to the FLS trainer using a 5-point Likert scale post experimental questionnaire. The ratings were ranging from 1 (not realistic/not useful) to 5 (very realistic/useful).

3.4.4 Experimental design and data analysis

The participants' performance on each simulator was compared using a mixed design (Split-Plot) ANOVA. The study was a 2x4 within-between subjects design. The within subjects factor was Trainer with two levels (FLS Vs VBLaST1). The between subjects factor was Experience group with four levels (1, 2, 3 and 4). Moreover, The Pearson's correlation tests were used to assess the correlation between the FLS and the VBLaST scores. Finally, descriptive analyses were used to present the results of the post experimental questionnaire. All analyses were performed using SPSS v.21.0 (IBM Corp., Armonk, NY, USA).

3.5 Results of the first system's evaluation

3.5.1 Effects of trainer and expertise level on performance

All assumptions for the Split-Plot ANOVA (the normality of data, the equality of covariance, and the homogeneity of variances) were checked using the appropriate statistical tests (Shapiro-Wilk test, the Box's test and the Levene's test of equality of variances, respectively). The results indicate that all assumptions are met.

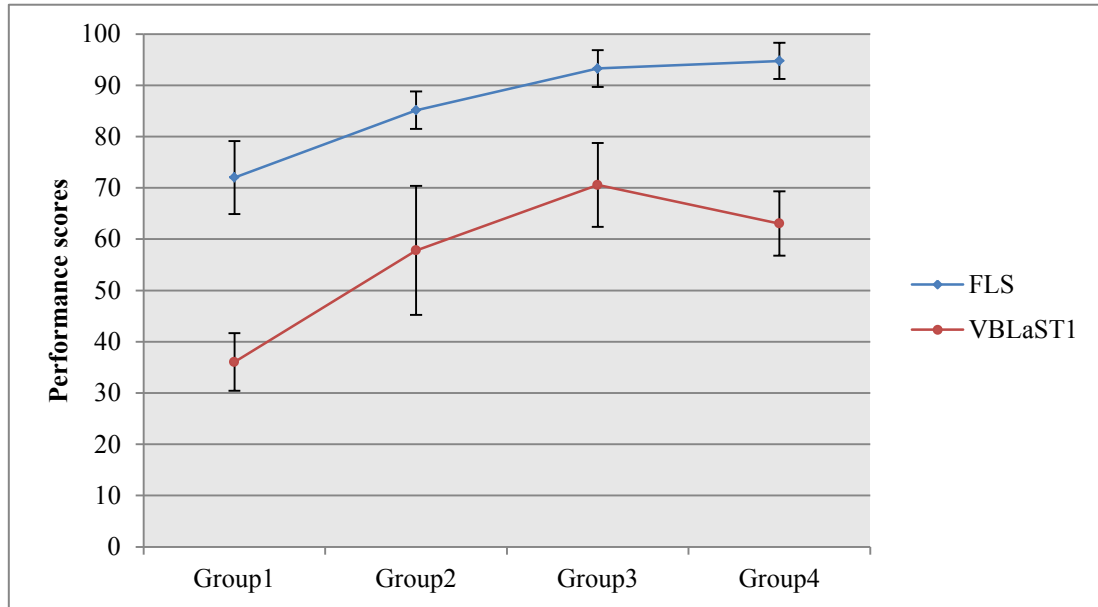


Figure 3: Comparing the early VBLaST prototype and the FLS box-trainer (error bars represent the standard error)

The results of the ANOVA show that there is a significant main effect of Trainer ($F(1, 16) = 51.24$, $p < 0.001$, $\eta_p^2 = 0.76$) and a significant main effect of Experience group ($F(3, 16) = 5.00$, $p = 0.012$, $\eta_p^2 = 0.48$) on the performance scores (Figure 3). There is no significant interaction between the two factors ($F(3, 16) = 0.48$, $p = 0.70$). These results indicate that the subjects' performance on the FLS box-trainer was significantly higher than their performance on the VBLaST1 simulator.

Moreover, the Post-Hoc tests with Bonferroni correction show that group 1 performed (mean = 72.04 ± 15.9 ; mean = 36.05 ± 12.57 ; respectively for FLS and VBLaST1) significantly lower than group 3 (mean = 93.27 ± 8.06 ; mean = 70.56 ± 18.31 ; respectively for FLS and VBLaST1) and group 4 (mean = 94.76 ± 7.83 ; mean = 63.05 ± 13.97 ; respectively for FLS and VBLaST1) at the .05 level of significance. No other significant differences were observed among the three other groups (Figure 3).

Finally, the Pearson's correlation test shows that the FLS and VBLaST1 mean scores had a correlation of 0.60 (Pearson's $r(18) = 0.60, p = 0.005$).

3.5.2 Subjective evaluation

The results of the subjective evaluation are presented on Table 2. The results show that the rating scores were relatively low more particularly for the realism of instruments handling, the quality of the haptic feedback and the trustworthiness of the system in quantifying the laparoscopic performance as compared to the FLS trainer. The overall realism of the simulation and the usefulness of the system for learning the different laparoscopic skills (hand-eye coordination, bimanual dexterity and speed and accuracy) were rated average. Finally, the realism of virtual objects and the usefulness of the haptic feedback were rated high.

Table 2: subjective rating of the early VBLaST prototype relative to the FLS box-trainer

Questions: scale varies from 1 (not realistic/not useful) to 5 (very realistic/useful).	Mean Rating	SD
1. Realism of the virtual objects (the rings, the pegs...etc.)	3.65	±1.18
2. Realism of instrument handling compared to FLS	2.80	±0.95
3. Overall realism compared to FLS	3.05	±0.82
4. Quality of Haptic feedback	2.80	±1.15
5. Usefulness of Haptic Feedback	3.85	±1.38
6. Usefulness for learning hand-eye coordination skills compared to FLS	3.25	±1.06
7. Usefulness for learning bimanual dexterity skills compared to FLS	3.20	±1.05
8. Overall usefulness for learning laparoscopic skills compared to FLS	2.95	±0.94
9. Trustworthiness in quantifying laparoscopic performance	2.75	±1.20

SD = standard deviation.

4 EXPERT FOCUS GROUP REFLECTION AND FEEDBACK

The objective of the first study was to evaluate the early prototype of our system and identify its validity and problems. The results of this evaluation have shown that the subjects' performance on the early VBLaST prototype was significantly lower than their performance on the FLS box-trainer. Moreover, although the scores

on the VBLaST and FLS are significantly correlated, the correlation coefficient is medium ($r = 0.6$). This suggested that the system required improvements to meet the users' needs in terms of interface and environment fidelity. The analysis of the subjective evaluation gave us some indications on what should be improved.

After this first evaluation session, we held a focus group session with expert surgeons ($n = 4$), software developers and designers. The aim of this session was to analyze the various dimensions of the early VBLaST prototype, to identify the issues that may have contributed to decrease the users' performance on this early version and attempt to correct them. As we identified earlier from the literature, fidelity can significantly affect the performance of users in a simulation environment. We know that one could perform the same tasks in the two systems, so it was not a difference in unmet manual or cognitive requirements. Thus, we turned our attention to the felt experience of using the system.

For that purpose, the expert surgeons were asked to test successively the two systems (both the physical and the virtual trainer) and describe the differences between them. The feedback during this comparative situated evaluation was provided verbally and the session was videotaped. Video analyses combined with the findings of the previous preliminary studies (Arikatla et al., 2013; Chellali et al., 2014) and the analyses of the questionnaires results and comments provided by subjects during the first experiment, led us to identify different items that required improvements. In the following, we describe the main issues with regards to environment and interface fidelity that were identified and the related design decisions to overcome them during the second design iteration.

4.1 Incorrect mapping of the workspace

The first identified issue is related to the size and mapping of the workspace. In fact, haptic devices are usually used as position control devices in which the displacement of the end-effector (the device stylus) is directly correlated with the displacement of the virtual tool displayed on the screen (Conti and Khatib, 2005). This displacement correlation may not be a one-to-one correspondence, since the tool position may be scaled according to a constant mapping from the device position. In this case, the end effector may be moved a distance of one centimeter in the workspace which causes the virtual tool to move ten centimeters in the virtual space. This can be useful for some applications in which the size of the device workspace is smaller than the size of the virtual space. Indeed, this allows the user to reach objects in all areas of the virtual space. However, this may not be appropriate in the context of learning a surgical motor skill in which fine and accurate positioning of tools is an important task. For these tasks, the large scaling of device movement to virtual tool movement makes the task difficult to perform and may lead to learning the wrong motor skills. Rather, the system needs to simulate

faithfully the real world to ensure the correct transfer of skills such as accuracy and speed of movement. In McMahan's framework of interaction fidelity analysis (FIFA) (2011), this is referred to as the Transfer Function Symmetry under the control symmetry concept. McMahan argues that a one-to-one position-to-position mapping between the virtual and the real-world results in a higher level of transfer function symmetry which in turn increases the interaction fidelity (2011).

After testing our system, the expert surgeons commented that although the size of the laparoscopic tool handles connected to the haptic devices were similar to the physical tools in the FLS box-trainer, the combination of the virtual tool tips and the physical handles resulted in longer tools than those used in the FLS box-trainer. Therefore, users had to move them deeper into the virtual 'box' in order to reach and manipulate the virtual rings. Moreover, they felt that the virtual space was larger than the FLS trainer workspace resulting in much bigger tool movements to manipulate the rings.

Software developers subsequently confirmed that the size of the virtual tools and space in this prototype were different from the size of the physical tools and workspace. These interface fidelity issues led to an increase in the distance traveled by the tools which would invariably increase the time to complete the task and alter the accuracy of movement. This also affects bimanual dexterity, speed and precision, which were identified as important learning objectives of our system.

This led to the second round design decision of reducing the size of the virtual space so that it matches exactly the size of the physical workspace. Moreover, it was decided to use shorter tool handles connected to the haptic devices so that their combination with the virtual tool tips matches the exact length of the actual tools. For that purpose, some measurements on the FLS box and laparoscopic tools were performed to match the virtual environment to the real-world environment.

4.2 Lack of depth cues

The second identified issue is related to the availability of spatial cues. In fact, grasping, moving and placing objects involve fine motor skills that are dependent on precise perceptual judgments (Hu et al., 2000). In laparoscopic surgery, the 2D image provided on the monitor lacks binocular depth cues. Subsequently, surgeons compensate this lack of spatial cues by relying on other information such as the pictorial depth cues available from the video display (Shimotsu and Cao, 2007). These include the monocular cues that are preserved during the 3D to 2D transition from the real image to the monitor and are not substantially degraded in the video image (Wickens and Hollands, 2000). One example of these monocular cues used by the surgeons to perceive depth is shadows cast by the laparoscopic instruments (Shimotsu and Cao, 2007). These shadows are created when an

object blocks the illumination of a surface by obscuring the light source. The shape of the shadow provides clues on the shape of the object and its spatial position and orientation. Moreover, the motion of shadows generated by the movement of the shadow-casting objects can give the observer important cues about the trajectory of the moving object and is linked with the object movement to perceive depth (Imura et al., 2006). Several studies have shown that these shadows are useful depth cues and that humans begin to use them from infancy (Puerta, 1989; Castiello, 2001; Imura et al., 2006).

When testing the early VBLaST prototype, the expert surgeons noticed that no shadows were displayed. During the discussions, they confirmed that the shadows casted by the surgical instruments are an important depth cue and distance judgment information that helps them during picking, transfer and placement of the rings in the peg transfer task. In fact, when moving an instrument, cast shadows allow them to know when the tool is in close proximity to the surface of an object before contact occurs. They can then timely decide to open the tool tip to grab, transfer or release the rings they are manipulating.

They commented that the absence of shadows in our system altered their perception of depth and imminent contact with objects during the manipulation of the rings. This suggests that this issue has also contributed to decrease the task performance as shown in previous research (Hu et al., 2000). This is a visual fidelity problem that we could not have known or ascertained from observations how important this was in skills training. This led to the design decision of including simulated cast shadows in the virtual environment with a particular focus on displaying them in the right position so that they can give the appropriate depth cues to the users.

4.3 Lack of haptic feedback fidelity

The last identified issue is related to the quality of the haptic feedback. In fact, in real laparoscopic surgery haptic perception is highly impaired as interaction with organs is performed through rigid instruments (Klatzky et al., 2003; Brydges et al., 2005). Several studies have investigated the impact of haptic interaction modality in surgical training. Results from these studies show that haptics is an essential component of minimally invasive surgery VR trainers (Dang et al., 2001; Seymour et al., 2002; Reich et al., 2006; Ström et al., 2006; Panait et al., 2009; Chellali et al., 2012), and particularly more so for less experienced surgeons (Cao et al., 2007).

In the past, there was an agreement among experts that training on the box trainer is preferable over virtual reality simulator systems for some laparoscopic surgical tasks because of the lack of fidelity of haptic feedback in VR trainers (Botden et al., 2007). Nevertheless, with the advent of haptic rendering technology it is now possible to simulate faithfully haptic interactions with complex virtual objects in real-time and then overcome this limitation of VR trainers.

After testing our system during the focus group session, our experts pointed out some issues related to the quality of the haptic feedback of the system. Indeed, they noticed that there is a mismatch between the haptic feedback felt when interacting with the virtual rings and the corresponding visual feedback. For instance, they commented that sometimes it seemed like the virtual tool tip was visually in contact with the ring while no haptic feedback was felt. Moreover, the forces felt when touching the virtual rings were much stronger than those felt by the users when interacting with the actual rings. Therefore, they needed to apply stronger forces in order to move the virtual rings. Finally, they noticed that grasping the rings caused sometimes strong vibrations on the devices which disturbed them when performing the task. The software engineers found that this was due to some computational errors related to the collision detection algorithm which caused the instability of the haptic device. Therefore, it was decided to improve the collision detection algorithm as well as the fidelity of the haptic feedback in order to overcome these limitations. Improvements included for instance, measuring the forces generated when manipulating the actual rings so that similar force amplitudes can be displayed on the haptic devices when interacting with virtual rings.

5 SECOND DESIGN ITERATION

A modified VBLaST prototype was designed and implemented. This modified prototype leveraged the lessons learned from the early prototype evaluation as described above. The new prototype is described in the apparatus section below.

5.1 Second experimental evaluation session

To assess the design improvements between the early and the modified versions of the VBLaST system, a second experiment took place using the modified VBLaST prototype. This time twenty-two Obstetrics and Gynecology (OB/GYN) residents and attendings (N = 22) were recruited (26-45 years old, 2 male-20 female, 2 left handed). They consisted of 9 first year residents, 4 third year residents, 5 fourth year residents, and 4 attendings all from a teaching hospital in the Greater Boston Area. The subjects were categorized into four groups depending on their expertise level and comparable (in terms of residency years) to the four groups previously defined for general surgery residents in the first experiment as shown on Table 3.

Table 3: Groups' of subjects in the second experiment

Expertise level	First year	Third year	Forth year	Attendings
Group name	Group1	Group2	Group3	Group4
Number of subjects	9	4	5	4

5.1.1 Apparatus

In this second experiment, the same FLS box-trainer and set of laparoscopic tools were used and compared to the modified VBLaST prototype (Figure 4).

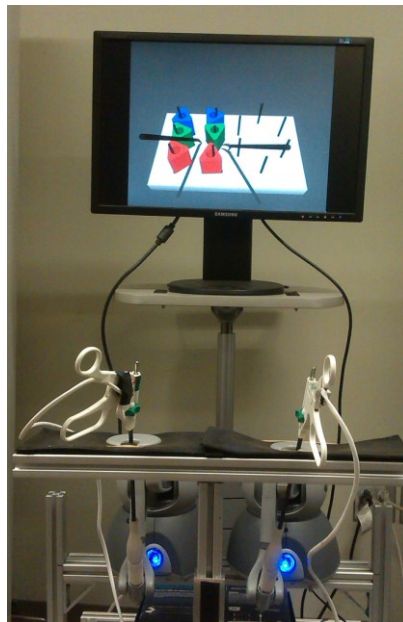


Figure 4: The modified VBLaST prototype (VBLaST 2)

5.1.2 Task and procedure

The same task and procedure as experiment 1 were used in this experiment.

5.1.3 Dependent measures

We used the same dependent measures as those used in the previous experiment. Moreover, the same normalization factors were used for this experiment. Finally, the same questionnaires and scales were used for the subjective evaluation.

5.1.4 Experimental design and data analysis

The participants' performance on each simulator was compared using a mixed design (Split-Plot) ANOVA. The study was a 2x4 within-between subjects design. The within subjects factor was "trainer" with two levels (FLS Vs VBLaST2). The between subjects factor was "experience group" with four levels (1, 2, 3 and 4). Moreover, the participants' performance on each simulator was compared over the two experiments using two 2-way ANOVAs. The study was a 2x4 between subjects design. The first factor was "trainer" with two levels (FLS1 Vs FLS2 for the first comparison, and VBLaST1 Vs VBLaST2 for the second comparison). The second factor was "experience group" with four levels (1, 2, 3 and 4) for both comparisons. Versions 1 and 2 of the FLS box-trainer (FLS1 and FLS2) were exactly the same, whereas the modified VBLaST prototype (VBLaST2) was the improved version of the early VBLaST prototype (VBLaST1) as described above. In addition, a correlation analysis (Pearson's test) was performed to investigate the relationship between the FLS and VBLaST2 scores. Finally, Mann Whitney non parametric test was used to compare the subjective ratings of the two VBLaST prototypes. All analyses were performed using SPSS v.21.0 (IBM Corp., Armonk, NY, USA).

5.2 Results of the second system's evaluation: comparative analyses

5.2.1 The effect of the expertise of participants

Since the two pools of subjects in experiments 1 and 2 were from two different medical specialties (General surgery and OB/GYN), the first step in our analysis was to determine whether the performance of these two populations on the FLS box-trainer were not significantly different in our two experiments. Because the FLS box-trainers used in the two experiments were similar and the expertise groups not significantly different, we hypothesized that the performance of both populations would not be significantly different on this trainer. To verify this hypothesis, a 2-way between subjects ANOVA was conducted.

All assumptions for this 2-way ANOVA (the normality of data, and the homogeneity of variances) were checked using the appropriate statistical tests (Shapiro-Wilk test and the Levene's test of equality of variances, respectively). The results indicate that the normality assumption is met. However, the Levene's test reveals that the assumption of equality of variances is not met. According to Tabachnick and Fidell (1989), the inequality of variances increases Type I error risk (page 45). To overcome this issue, Keppel (1991) suggests correcting the violation of the equal variance assumption in a two-way ANOVA by substituting alpha by alpha/2. In our case, this leads to an alpha level of 0.025.

The results of the ANOVA (Figure 5) show no significant main effect of the FLS trainer on the performance scores ($F(1, 34) = 2.51, p = 0.12, \eta_p^2 = 0.06$), while a significant main effect of the Expertise group on the performance scores was found ($F(3, 34) = 22.91, p < 0.001, \eta_p^2 = 0.66$). Finally, no significant interaction between the two factors was found (Figure 5). The Post-Hoc tests with Bonferroni correction show that group 1 performed (mean = 72.04 ± 15.9 ; mean = 63.56 ± 6.72 ; respectively for FLS1 and FLS2) significantly lower than group 2 (mean = 85.14 ± 8.12 , mean = 86.62 ± 3.71 ; respectively for FLS1 and FLS2) group 3 (mean = 93.27 ± 8.06 ; mean = 90.70 ± 2.69 ; respectively for FLS1 and FLS2) and group 4 (mean = 94.76 ± 7.83 ; mean = 88.08 ± 0.77 ; respectively for FLS1 and FLS2) at the .025 level of significance. No other significant differences were observed among the three other groups (Figure 5).

Because the objective was to determine whether the two populations are significantly different from each other, we have also performed direct comparisons. First, we have conducted a simple independent samples t-test to determine whether the performance of all the participants on FLS1 and FLS2 was significantly different. The results show that the assumption of equality of variances is met and that there is no significant effect ($t(41) = 1.90, p = 0.06$) of the system on the performance (mean = 86.3 ± 13.39 ; mean = 78.4 ± 13.48 , for FLS1 and FLS 2, respectively).

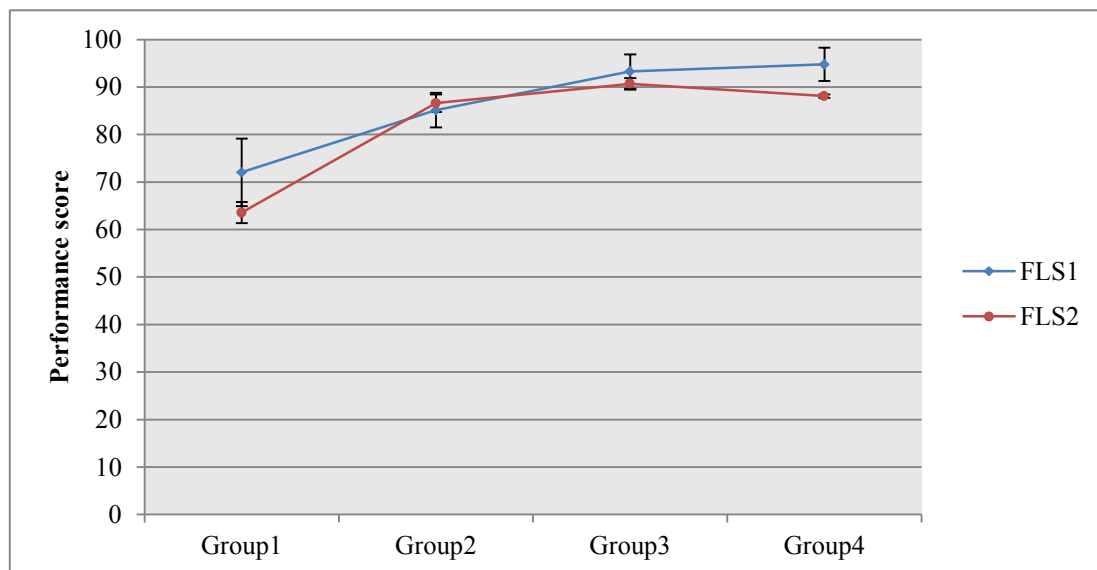


Figure 5: comparing the performance of the two populations of subjects on the FLS trainer (error bars represent the standard error)

We have also conducted a one way ANOVA to compare the scores on the two systems for each group of expertise. The Levene's test in this case indicates that the assumption of equality of variances is met only for group 2. Therefore, the Welch correction was used to overcome the violation of equality of variances

assumption. The one-way ANOVA with Welch correction results indicate no main effect of the system for all of the four groups ($F(1, 13) = 2.01, p = 0.18$; $F(1, 8) = 0.11, p = 0.74$, $F(1, 9) = 0.45, p = 0.51$, $F(1,8) = 2.80, p = 0.14$; respectively for group 1, group 2, group 3 and group 4).

These results confirm that the two populations are not significantly different, both regarding the overall performance scores and the experience groups. The validation of our first hypothesis permits to conduct further comparisons as described in the next sections.

5.2.2 The effect of the design iteration on the VBLaST trainer

To investigate the effect of the design process and determine whether the performance was improved on the modified version of the VBLaST prototype as compared with the early VBLaST prototype, a 2-way between subjects ANOVA on the VBLaST scores was conducted.

All assumptions for this 2 way ANOVA (the normality of data, and the homogeneity of variances) were checked using the appropriate statistical tests (Shapiro-Wilk test and the Levene's test of equality of variances, respectively). The results indicate that all assumptions are met.

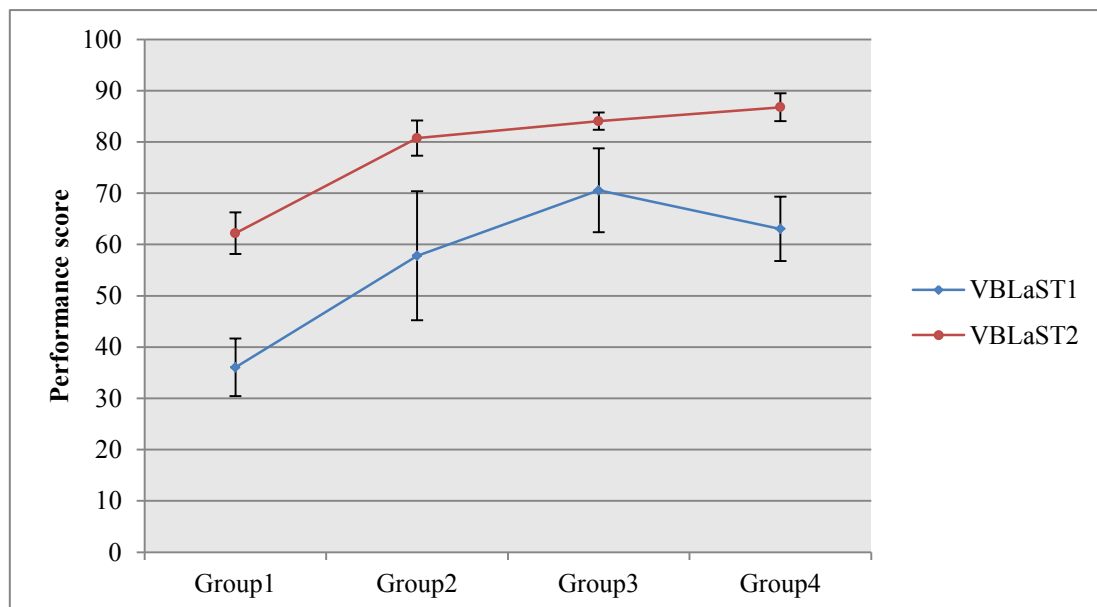


Figure 6: Comparing the performance of the two populations of subjects on the two VBLaST prototypes (error bars represent the standard error)

The results of the ANOVA (Figure 6) show a significant main effect of the VBLaST version ($F(1, 34) = 21.29, p < 0.001, \eta_p^2 = 0.38$) and a significant main effect of the Experience group ($F(3, 34) = 8.81, p < 0.001, \eta_p^2 = 0.43$) on the VBLaST performance scores. No significant interaction was found between the two

factors. The results (Figure 6) indicate that the subjects performed significantly better on the modified VBLaST version. Overall, the mean performance scores were increased by 38% in the new VBLaST version. Moreover, the Post-Hoc tests with Bonferroni correction show that experience group 2 (mean = 57.79 ± 28.12 ; mean = 80.74 ± 6.82 ; respectively for VBLaST1 and VBLaST2), group 3 (mean = 70.56 ± 18.31 ; mean = 84.06 ± 3.70 ; respectively for VBLaST1 and VBLaST2) and group 4 (mean = 63.05 ± 13.97 ; mean = 86.76 ± 5.43 ; respectively for VBLaST1 and VBLaST2) performed significantly better than group 1 (mean = 36.05 ± 12.57 ; mean = 62.20 ± 12.12 ; respectively for VBLaST1 and VBLaST2) at the .05 level of significance. All other comparisons were not statistically significant (Figure 6).

5.2.3 The effect of the trainer (FLS vs VBLaST2)

Our design objective was to obtain similar performance on the virtual and the physical trainer. To compare the subjects' performance on the modified VBLaST and the FLS box-trainer, a mixed-design (Split-Plot) ANOVA was conducted.

All assumptions for this Split-Plot ANOVA (the normality of data, the equality of covariance, and the homogeneity of variances) were checked using the appropriate statistical tests (Shapiro-Wilk test, the Box's test, the Levene's test of equality of variances, respectively). The results indicate that all assumptions are met.

We expected the subjects' performance on the modified VBLaST prototype to be not significantly different from to the subjects' performance on the FLS box-trainer.

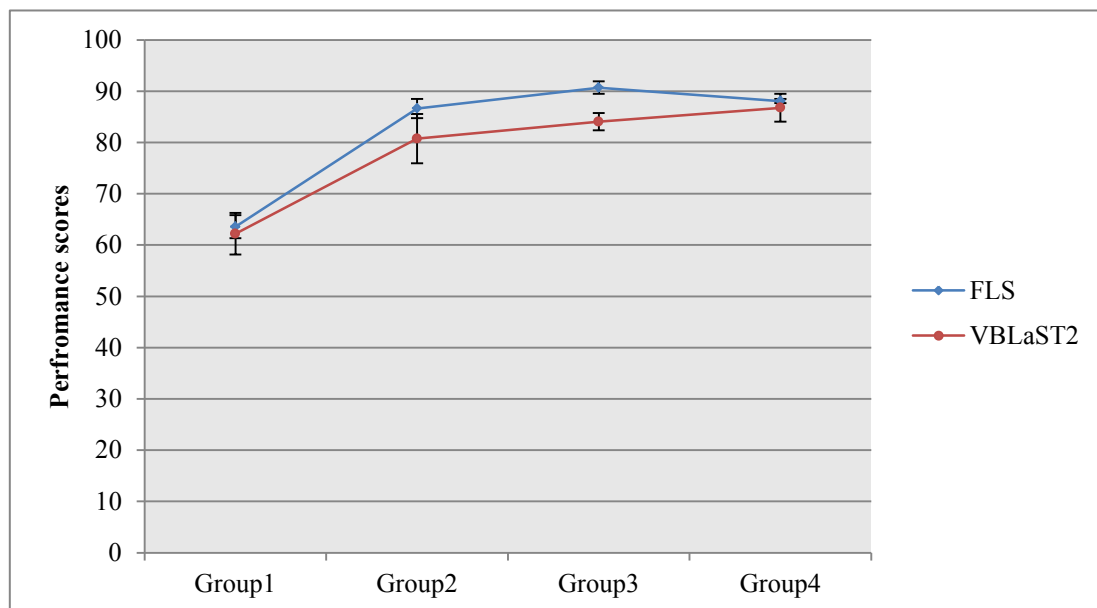


Figure 7: Comparing the modified VBLaST prototype and the FLS box-trainer (error bars represent the standard error)

The results (Figure 7) show no significant main effect of the Trainer ($F(1, 18) = 3.62, p = 0.073, \eta_p^2 = 0.16$) and no significant interaction between the Trainer and Expertise level ($F(3, 18) = 0.56, p = 0.65, \eta_p^2 = 0.08$) on the performance scores. This confirms that the performance on the two trainers (VBLaST2 and FLS) was not significantly different (Figure 7). On the other hand, there is a significant main effect of Experience group ($F(3, 18) = 29.83, p < 0.001, \eta_p^2 = 0.83$) on the performance scores (Figure 7). The Post-Hoc tests with Bonferroni correction show that group 1 (mean = 63.56 ± 6.72 ; mean = 62.20 ± 12.12 ; respectively for FLS and VBLaST2) performed significantly lower than group 2 (mean = 86.62 ± 3.71 ; mean = 80.74 ± 6.82 ; respectively for FLS and VBLaST2) group 3 (mean = 90.70 ± 2.69 ; mean = 84.06 ± 3.70 ; respectively for FLS and VBLaST2) and group 4 (mean = 88.08 ± 0.77 ; mean = 86.76 ± 5.43 ; respectively for FLS and VBLaST2) at the .05 level of significance. No other significant differences were observed among the three other groups (Figure 7).

To further check whether the performance on the two systems is significantly different, we have also performed direct comparisons. First, we have conducted a simple paired-samples t-test to determine whether the performance of all the participants on FLS and VBLaST2 was significantly different. The results show that the assumption of equality of variances is met and that there is no significant effect of the system ($t(21) = 1.84, p = 0.08$) on the performance (mean = 78.4 ± 13.48 ; mean = 75.0 ± 13.85 ; for FLS and VBLaST2, respectively).

Finally, the Pearson's correlation test shows that the FLS and VBLaST2 mean scores had a correlation of 0.80 (Pearson's $r(20) = 0.80, p = 0.005$).

5.2.4 Subjective evaluation

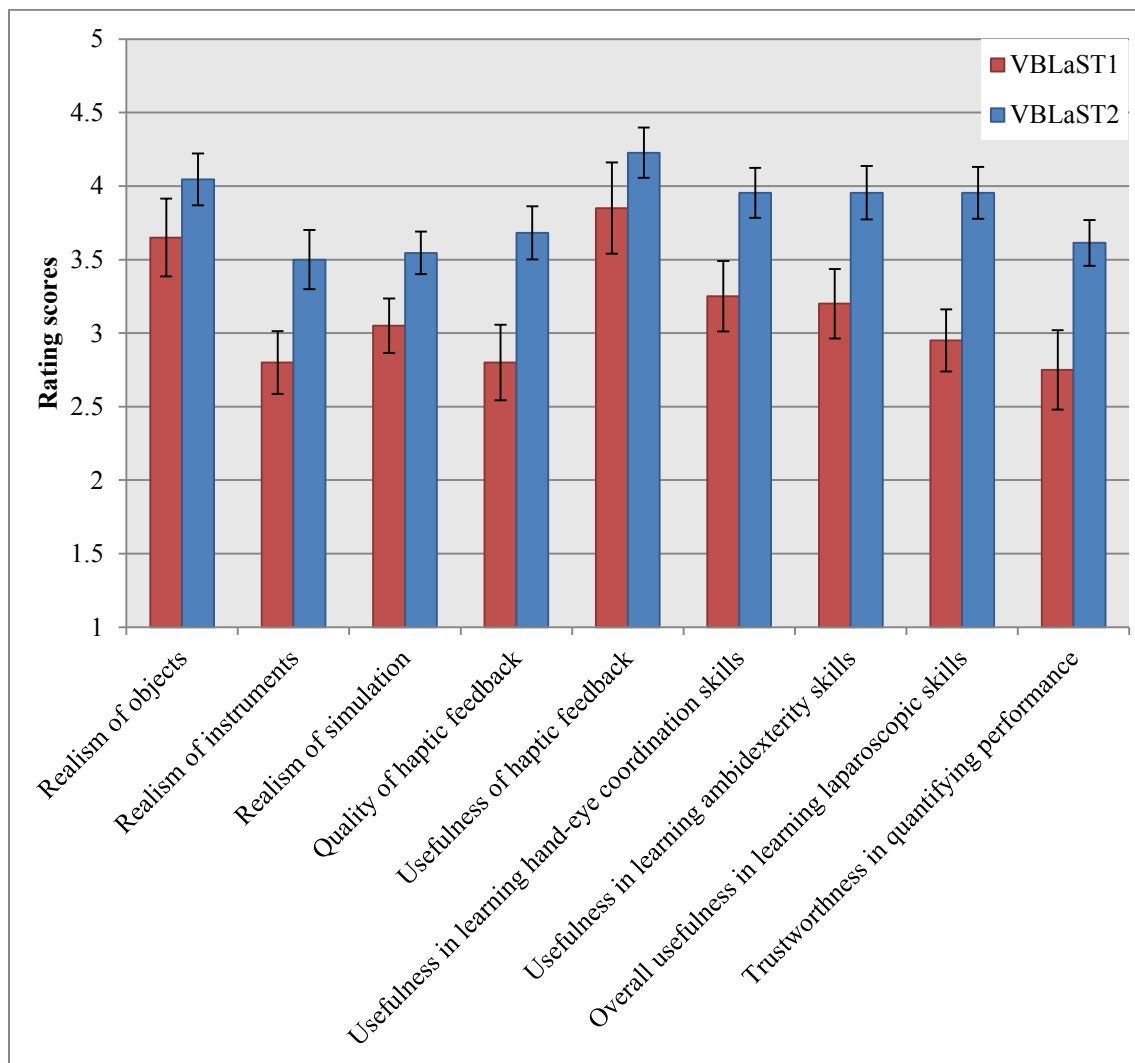


Figure 8: comparison of subjective evaluations of the two VBLaST prototypes (error bars represent the standard error)

The results show that overall; the ratings of all the evaluated aspects of the system were increased in the second experiment (Figure 8).

The Mann-Whitney tests (Table 4) show a significant effect of the VBLaST prototype version on the ratings of the realism of instruments, the quality of the haptic feedback, the usefulness of learning the different laparoscopic skills and the trustworthiness in quantifying the laparoscopic performance. No other significant differences were found.

Table 4: Results of the Mann-Whitney test comparing the ratings of the two VBLaST prototypes

	<i>Mann Whitney U</i>	<i>Z</i>	<i>P-value</i>
	<i>VBLaST 1 Vs</i>	<i>VBLaST 1 Vs</i>	<i>VBLaST 1 Vs</i>
	<i>VBLaST2</i>	<i>VBLaST2</i>	<i>VBLaST2</i>
1. Realism of the virtual objects	180.00	-1.06	0.28
2. Realism of instrument handling	142.50	-2.04	0.04*
3. Overall realism	154.50	-1.79	0.07
4. Quality of Haptic feedback	123.00	-2.55	0.01*
5. Usefulness of Haptic Feedback	123.00	-2.55	0.01*
6. Usefulness for learning hand-eye coordination skills	197.50	-0.60	0.54
7. Usefulness for learning bimanual dexterity skills	136.00	-2.24	0.02*
8. Overall usefulness for learning laparoscopic skills	135.50	-2.21	0.02*
9. Trustworthiness in quantifying performance	96.50	-3.26	0.001*

* $P < 0.05$ =significant effect

6 DISCUSSION AND DESIGN IMPLICATIONS

Virtual reality is a promising technology for developing laparoscopic skills task trainers. The long term goal of this research is to design and develop the VBLaST system. This system is aimed to reach similar objectives in terms of basic technical laparoscopic skills as the current accepted training standard in the surgical community (the FLS box-trainer) while overcoming some of its limitations. The advantage of replicating an existing physical trainer is twofold. First, the FLS trainer has been previously validated and has widely proven its efficacy in training basic laparoscopic skills (Fried, 2008). Second, the assessment of laparoscopic performance on this trainer is also based on validated metrics (Fraser et al., 2003; Fried, 2008).

6.1 Summary of findings

In this work, we have presented two evaluation studies that allowed us to compare two prototypes of the VBLaST system built after two successive iterations of the design process. The results of these studies demonstrate that the subjects' performance measured using the current standard scoring method (Fried, 2008), was significantly increased on the modified VBLaST prototype as compared with the early VBLaST prototype. The results of our second experiment demonstrate that the subjects' performance on the modified VBLaST prototype assessed using the current standard method (Fried, 2008), was not significantly different from their performance on the FLS box-trainer. Moreover, the scores on both systems were significantly correlated with a high coefficient of correlation ($r = 0.80$). This suggests that the second design iteration permitted to attain our main design objective; having a surgical performance not significantly different from that of the FLS box-trainer in assessing laparoscopic skills. This further demonstrates that inputs from the experts were determinant to improve the system fidelity and training performance and thus, attain the design objectives.

Moreover, while there is no significant difference between the VBLaST and the FLS scores, the results show that the performance scores on the two systems are much closer for group 1 (first year residents) and group 4 (fellows and attendings). These results can be explained by the fact that these groups have reported less exposure to the peg transfer task on the FLS trainer as compared with the two other groups. This suggests that the current version of the VBLaST could be more comparable to the FLS trainer for training novices and users with no previous experience with laparoscopic trainers. The strong correlation between the scores on the two systems further suggests that the current prototype is an efficient tool to assess the laparoscopic performance for the peg transfer task.

Finally, the subjective evaluation of the two prototypes indicates that the perceived realism of the virtual environment was not affected by the new design. This result is not surprising since these aspects of the environment fidelity were not modified and were already rated high after the first design iteration. On the other hand, other aspects of the environment and interface fidelity, such as the perceived realism of instruments and the quality of the haptic feedback were rated significantly higher in the second experiment. Moreover, the perceived usefulness of the system for learning basic laparoscopic skills increased over the two experiments. This also increased the subjects' assessment of how trustworthy the VBLaST is to quantify accurate measures of laparoscopic performance. These results demonstrate that the design improvements have a direct positive effect on the subjects' subjective assessment of the environment and interface fidelity and therefore, on the VBLaST as

a training system for basic laparoscopic skills. Nevertheless, it is important in the future to evaluate the new version of the VBLaST also with general surgery residents and fellows. This will further validate our system.

6.2 Comparative situated evaluation for expert surgeons and the present-at-hand tool

Our findings have a significant impact on the design methods used in effecting successful felt experiences in VR trainers. We show the importance of involving expert surgeons (i.e. domain experts) during comparative situated evaluation sessions in order to analyze the fidelity component of the system. While the design aspects we uncovered were important in our system, the initial field studies did not permit us to identify them (Arikatla et al., 2013; Chellali et al., 2014). In fact, these design elements are all related to movement in the 3-dimensional space and belong then to the procedural knowledge of experts, as defined by Rasmussen (1983). And discussions and observations of *residents* during their training for the peg transfer task did not allow us to extract these elements. From the novices' perspective, the first prototype of the system replicated faithfully the physical trainer that they were familiar with because it simulated the real world objects (environment fidelity) and mimicked the interaction techniques (interface fidelity). However, as we showed, this in and of itself did not lead to a successful VR trainer. Upon testing the system, *the expert surgeons* provided a different point of view. During the comparative situated evaluation session, expert surgeons were able to note the absence of elements that are inherent in the real world. As we showed with the increased performance of our system to effect similar training outcomes as the physical box training, these elements have a direct impact on learning laparoscopic skills and so ignoring them may result in designing an ineffective training system.

The crucial element to reiterate is that expert surgeons are not the end users of the training system; surgical training simulators are designed for use by novices. However, because of their expertise, expert surgeons have the domain knowledge that designers and the end users (surgical trainees) do not possess and can provide invaluable input to the design process. The comparative situated evaluation session allows an expert in the felt experience of the real world trainer to evaluate the systems at different stages of the design process to identify the potential usability and fidelity issues. The first design iteration is still very much necessary because the success of the comparative situated evaluation sessions is dependent upon the availability of a working prototype of the system. It is this hands-on experience of the system that allowed for us to elicit the nuanced interactions that bridge the gap of fidelity, with a focus on one component of fidelity at a time.

For instance, haptic sensations – information about the amount of forces to apply – are difficult to describe verbally. Allowing the expert surgeons to act while evaluating the two systems allowed them to compare the

haptic sensations on both systems and to describe verbally the relative differences (harder, softer...etc.) between them. This, combined with the actual measurements of forces being used permitted us to enhance the haptic fidelity of the system. It was also useful to investigate the visual fidelity. The comparative situated evaluation highlighted the tight link between seeing and doing. In our case, this was observed when the expert surgeons noted while acting that the lack of visual landmarks and depth cues altered their visual perception and prevented them from correctly performing the surgical tasks. Other aspects of the visual fidelity were judged to be less relevant to the learning objectives of the peg transfer task. This is the case for the visual deformation of the virtual rings while manipulating them using the virtual tools. The slight flexibility of the plastic rings in the FLS box trainer was judged by the expert surgeons not to be relevant to the task of picking and transferring them. We have therefore modeled the virtual rings as rigid bodies in the virtual environment, which significantly improved the computational efficiency of the software. Finally, the auditory feedback was judged to be not relevant for this training task and therefore was not included. However, the expert surgeons commented that the auditory feedback could be more important for some of the remaining FLS tasks, such as the pattern cutting.

The evaluation highlighted also issues related to the interface fidelity. During the comparative evaluation of the physical interface, we analyzed the shape, the size, the location, and the input position DOF of the surgical instruments. We also analyzed the fidelity of interaction (i.e., visuomotor mapping) between the physical handles and the virtual end effectors (tool tips) of these instruments. Results show that the movements and the handling of the tools were identical on both systems but their scales were different. This was then linked to the size of the physical workspace, which has a direct impact on the amplitude of movements. Correcting these issues led to significant improvements in the quality of the physical user interface.

The phenomenon that we are capitalizing on with the situated comparative method is the difference between “ready-to-hand” (zuhanden) and “present-at-hand” (vorhanden) as introduced by Martin Heidegger (1927). These concepts explain how we encounter the world and act through it. By ‘ready-to-hand’ Heidegger describes a tool that has become invisible, receding to the background of the work, and we are no longer conscious of it. This in turn allows our concentration to be focused on the work as opposed to the tools that we use to perform the work. But a tool will move from the background to the foreground and become an entity that is ‘present-at-hand’ when that tool breaks down or fails to function as expected.

For our specific situation, allowing for an expert surgeon, who is already accustomed to the tools of the trade being ready-to-hand, to have the felt experience of the simulated environment, allows him to not only identify the failings of the present-at-hand virtual environment, but also bring to the foreground the feeling of the

environment they are so used to. This in turn allows them to articulate the differences in these feelings and help identify the crucial elements to bridge the gap between the physical and virtual environment.

For surgical training system designers, our findings highlight the need for evaluation methodologies beyond heuristic evaluations, or simply observing the motor tasks associated with laparoscopic surgery. The comparative situated evaluation should provide means for designers to ‘see’ and ‘feel’ what the expert surgeons see and feel, and if necessary facilitate the experts’ articulation of their expert tacit knowledge.

6.3 Implications for the process of designing VR trainers

Our recommendations for the success of comparative situated evaluations to achieve fidelity of a virtual reality trainer are: (1) involve domain experts for a direct comparison of the designed system with an existing system or with a real life situation, (2) investigate each fidelity component separately: focusing on one fidelity aspect at a time can help isolate its effect, (3) collect domain experts’ subjective feedback on the targeted fidelity component while they are performing the surgical tasks, and (4) link the fidelity aspects to the learning objectives of the training task: does this fidelity component have a direct effect on the training objectives of the task?

Here are some examples of components of fidelity to investigate within a laparoscopic surgical simulator:

- Environment fidelity components: the visual and haptic fidelity are important for the training of laparoscopic surgical. Auditory fidelity could be important depending on the surgical task.
 - Visual fidelity: visual display size and quality, positions of the virtual camera and virtual light sources, visual landmarks and depth cues, physics and visual deformation of the virtual objects, colors, sizes and shapes of the virtual objects, position and size of cast shadows, visual interactions between the virtual tools and objects. It is to be noted that some other components identified by previous researchers for another kind of simulators (Bowman and McMahan, 2007) could be less relevant for a laparoscopic surgery trainer (stereoscopy and the field of view, for instance).
 - Haptic fidelity: amount, directions and amplitude of forces, rendering frame rate.
 - Auditory fidelity: sounds amplitudes and frequencies, position of the sound source, displayed sound channels, quality of displayed sounds.
 - Multi-sensory coherence: one should ensure that all the previous components of fidelity are displayed in coherent way to the user.

- Interface fidelity components: the size of the physical workspace, the size and the shape of surgical instruments, the relative position of instruments to the user. One suggestion for the analysis of the different components of the interface fidelity is to use a validated analysis framework such as the FIFA (McMahan, 2011).

In conclusion, our main contributions are: (1) designed and validated an effective VR simulator for training basic laparoscopic skills, and (2) provided valuable design recommendations on the practical application of comparative situated evaluation to achieve environment and interface fidelity during the design of a VR surgical trainer. This work demonstrates the value of involving expert surgeons in hands-on evaluation sessions during the iterative design process and the development of VR surgical simulators, as well as generalized guidelines for use in similar design efforts. The methodology presented supports designers in finding the required real needs of the surgical trainees in this relatively new domain (i.e., laparoscopic surgery skills training).

7 LIMITATIONS AND FUTURE WORK

Because of the relatively small number of available subjects within our partner hospitals (an unfortunate reality in conducting research in clinical settings that rely on highly specialized domain experts), our studies were conducted with few participants in each experience level. This can be a limitation for conducting reliable statistical analyses. Nevertheless, the collected data are sufficient to show that our methodology significantly improved the VBLaST system over the two design iterations, and thus emphasizes the value of this approach and its application in clinical settings.

For the peg transfer task, only the novice groups (group 1 in both experiments) performed lower than the other participants on all the systems tested in our studies. This finding can be explained by the relative simplicity of the peg transfer task. In fact, all but novice subjects acknowledged having performed the peg transfer task using the FLS box trainer prior to the experiments (at least 10 times during the year preceding the experiments). This suggests that novices can attain a high proficiency level for this task after a small amount of training. In the future, we are planning to run a learning curve study to confirm that the current version of the VBLaST trainer yields a learning curve comparable to the FLS box-trainer. This will further validate the VBLaST peg transfer trainer.

One of our design objectives was to reduce the training costs of the FLS box-trainer relative to the use of consumables. However, no formal cost analysis of the VBLaST system was performed within the scope of this work. A quick investigation indicates that the hardware interface of the current prototype is based on commercially available haptic devices, which have relatively low cost and customized laparoscopic instruments.

This physical interface is connected to a standard PC and the simulation software and algorithms were developed by our research team. Therefore, we can argue that the system cost would be much lower than other existing VR simulators. However, a more rigorous cost analysis must be conducted once the system is validated for a formal comparison with the FLS box trainer and the other available systems.

We are continuing to use our design methodology to develop the remaining technical tasks of the FLS curriculum on the VBLaST system. After a first design iteration, the new VBLaST prototype that includes the pattern cutting and the ligating loop tasks was built and evaluated with experts (Chellali et al., 2014). Further comparative situated evaluation sessions will be conducted in the future with expert surgeons and end-users in order to improve the overall quality of the VBLaST system. Future evaluations will focus on how skills gained from training on the VR simulator would transfer to the operating room following the Kirkpatrick's four levels model for assessing training effectiveness (Kirkpatrick and Kirkpatrick, 2005). This will permit, in the future, adoption of the system as a new standard for training laparoscopic technical skills.

8 ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH) Grant NIBIB R01 EB010037-01. The authors would like to thank all of the volunteers who participated in the studies.

9 REFERENCES

- Al-Kadi, A.S., Donnon, T., Oddone Paolucci, E., Mitchell, P., Debru, E., Church, N., 2012. The effect of simulation in improving students' performance in laparoscopic surgery: A meta-analysis. *Surgical Endoscopy*, 26(11), 3215-3224.
- Arikatla, V.S., Sankaranarayanan, G., Ahn, W., Chellali, A., De, S., Cao, C.G.L., Hwabejire, J., DeMoya, M., Schwaitzberg, S., Jones, D.B., 2013. Face and construct validation of a virtual peg transfer simulator. *Surgical Endoscopy*, 27(5), 1721-1729.
- Aucar, J., Groch, N., Troxel, S., Eubanks, S.A., 2005. Review of Surgical Simulation with Attention to Validation Methodology. *Surgical Laparoscopic Endoscopic Percutaneous Technology*, 15, 82-89.
- Barkana Erol, D., Erol, D., 2013. Applying a User-Centered Design Approach to Develop a Surgical Interface for Cryoablation of a Kidney Tumor. In *the proceedings of the International Workshop on Human-Machine Systems, Cyborgs and Enhancing Devices (HUMASCEND)*. Manchester, United Kingdom.

- Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H., Srinivasan, M.A., 2004. Haptics in Minimally Invasive Surgical Simulation. *IEEE Computer Graphics and Applications*, 24(2), 56–64.
- Beyer, L., Troyer, J.D., Mancini, J., Bladou, F., Berdah, S.V., Karsenty, G., 2011. Impact of laparoscopy simulator training on the technical skills of future surgeons in the operating room: a prospective study. *American Journal of Surgery*, 202, 265–272.
- Botden, S.M., Jakimowicz, J.J., 2009. What is going on in augmented reality simulation in laparoscopic surgery? *Surgical Endoscopy*, 23(8), 1693-1700.
- Botden, S.M., Torab, F., Buzink, S.N., Jakimowicz, J.J., 2007. The importance of haptic feedback in laparoscopic suturing training and the additive value of virtual reality simulation. *Surgical Endoscopy*, 22, 1214-1222.
- Bowman, D.A., McMahan, R.P., 2007. Virtual Reality: How Much Immersion Is Enough? *IEEE Computer*, 40(7), 36–43.
- Breedveld, P., Wentink, M., 2001. Eye-hand coordination in laparoscopy - an overview of experiments and supporting aids. *Minimally Invasive Therapy & Allied Technologies*, 10(3), 155-162.
- Brennan, P.M., Loan, J.J.M., Hughes, M.A., Hennessey, I.A.M., Partridge, R.W., 2014. Surgical training is undermined by inadequate provision of laparoscopic surgical simulators. *Bulletin of the Royal College of Surgeons of England*, 96(9), 304-307.
- Brinkman, W.M., Tjiam, I.M., Buzink, S.N., 2013. Assessment of basic laparoscopic skills on virtual reality simulator or box trainer. *Surgical Endoscopy*, 27(10), 3584-3590.
- Brydges, R., Carnahan, H., Dubrowski, A., 2005. Surface exploration using laparoscopic surgical instruments: the perception of surface roughness. *Ergonomics*, 48(7), 874-894.
- Buckley, C.E., Nugent, E., Ryan, D., Neary, P.C., 2012. Virtual Reality – A New Era in Surgical Training. In C. Eichenberg, ed. *Virtual Reality in Psychological, Medical and Pedagogical Applications*. InTech.
- CAE Healthcare, 2016. *LapVR*. [Online] Available at: <http://www.caehealthcare.com/eng/interventional-simulators/lapvr> [Accessed 31 March 2016].
- Cao, C.G.L., MacKenzie, C.L., 1997. Direct, 2-D vs. 3-D endoscopic viewing & surgical task performance. *A Symposium on Evolving Technologies: Surgeons' Performance of Surgical Tasks. Journal of Sport and Exercise Psychology*, 19.

- Cao, C.G.L., MacKenzie, C.L., Ibbotson, J.A., Turner, L.J., Blair, N.P., Nagy, A.G., 1999. Hierarchical decomposition of laparoscopic procedures. *Studies in Health Technology and Informatics*, 62, 83-89.
- Cao, C.G.L., MacKenzie, C.L., Payandeh, S., 1996. Precision and safety constraints in laparoscopic surgery. *Journal of Sport and Exercise Psychology*, 18.
- Cao, C.G.L., Zhou, M., Jones, D.B., Schwaitzberg, S.D., 2007. Can surgeons think and operate with haptics at the same time? *Journal of Gastrointestinal Surgery*, 11(11), 1564-1569.
- Castiello, U., 2001. Implicit Processing of Shadows. *Vision Research*, 41, 2305-2309.
- Chellali, A., Ahn, W., Sankaranarayanan, G., Flinn, J., Schwaitzberg, S.D., Jones, D.B., De, S., Cao, C.G.L., 2014. Preliminary Evaluation of the Pattern Cutting and the Ligating Loop Virtual Laparoscopic Trainers. *Surgical Endoscopy*, 29(4), 815-821.
- Chellali, A., Dumas, C., Milleville-Pennel, I., 2012. Haptic communication to support biopsy procedures learning in virtual environments. *Presence: Teleoperators and Virtual Environments*, 21(4), 470-489.
- Chellali, A., Schwaitzberg, S.D., Jones, D.B., Romanelli, J., Miller, A., Rattner, D., Roberts, K., Cao, C.G.L., 2014. Toward scar-free surgery: an analysis of the increasing complexity from laparoscopic surgery to NOTES. *Surgical Endoscopy*, 28(11), 3119-3133.
- Chellali, A., Zhang, L., Sankaranarayanan, G., Arikatla, V.S., Ahn, W., Derevianko, A., Schwaitzberg, S.D., Jones, D.B., Demoya, M., Cao, C.G.L., 2014. Validation of the VBLaST Peg Transfer Task: A First Step toward an Alternate Training Standard. *Surgical Endoscopy*, 28(10), 2856-2862.
- Conti, F., Khatib, O., 2005. Spanning large workspaces using small haptic devices. In *the proceedings of the Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. pp.183-188
- Crothers, I.R., Gallagher, A.G., McClure, N., James, D.T.D., McGuigan, J., 1999. Experienced Laparoscopic Surgeons are Automated to the “Fulcrum Effect”: An Ergonomic Demonstration. *Endoscopy*, 31(5), 365-369.
- Dang, T., Annaswamy, T.M., Srinivasan, M.A., 2001. Development and evaluation of an epidural injection simulator with force feedback for medical training. *Studies in health technology and informatics*, 81, 97-102.
- Derossis, A.M., Fried, G.M., Abrahamowicz, M., Sigman, H.H., Barkun, J.S., Meakins, J.L., 1998. Development of a Model for Training and Evaluation of Laparoscopic Skills. *The American Journal of Surgery*, 175(6), 482-487.

- Dieckmann, P., 2009. *Using Simulations for Education, Training and Research*. Pabst, Wolfgang Science.
- Drews, F.A., Bakdash, J.Z., 2013. Simulation Training in Health Care. *Reviews of Human Factors and Ergonomics*, 8(1), 191-234.
- Edelman, D.A., Mattos, M.A., Bouwman, D.L., 2010. FLS Skill Retention (Learning) in First Year Surgery Residents. *Journal of Surgical Research*, 163(1), 24-28.
- Fairhurst, K., Strickland, A., Maddern, G., 2011. The LapSim virtual reality simulator: promising but not yet proven. *Surgical endoscopy*, 25(2), 343-355.
- Figert, P.L., Park, A.E., Witzke, D.B., Schwartz, R.W., 2001. Transfer of training in acquiring laparoscopic skills. *Journal of the American College of Surgeons*, 193(5), 533-537.
- Forsslund, J., Sallnäs, E.-L., Lundin Palmerius, K.-J., 2011. Design of Perceptualization Applications in Medicine. In *the proceedings of the first workshop on Engineering Interactive Computing Systems for Medicine and Health Care (EICS4Med)*. Pisa, Italy. pp.42-47
- Forsslund, J., Sallnäs, E.-L., Palmerius, K.-J., 2009. A User-Centered Designed FOSS Implementation of Bone Surgery Simulations. In *the proceedings of Third Joint EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009*. Salt Lake City, UT. pp.391-392
- Fraser, S.A., Klassen, D.R., Feldman, L.S., Ghitulescu, G.A., Stanbridge, D., Fried, G.M., 2003. Valuating laparoscopic skills: setting the pass/fail score for the MISTELS system. *Surgical Endoscopy*, 17(6), 964-967.
- Fried, G.M., 2008. FLS assessment of competency using simulated laparoscopic tasks. *Journal of Gastrointestinal Surgery*, 12(2), 210-212.
- Fried, G.M., Feldman, L.S., Vassiliou, M.C., Fraser, S.A., Stanbridge, D., 2004. Proving the value of simulation in laparoscopic surgery. *Annals of Surgery*, 240, 518-528.
- Gibson, J.J., 1971. The information available in pictures. *Leonardo*, 4, 27-35.
- Giles, J.A., 2010. Surgical training and the European working time directive: The role of informal workplace learning. *International Journal of Surgery*, 8(3), 179-180.
- Hamblin, C.J., 2005. *Transfer of training from virtual reality environments*. PhD Dissertation. Wichita, KS, USA: Wichita State University.

- Hays, R.T., Singer, M.J., 1989. Simulation fidelity as an organizing concept. In R.T. Hays & M.J. Singer, eds. *Simulation Fidelity in Training System Design, Recent Research in Psychology*. New York, NY, USA: Springer. pp.47–75.
- Heidegger, M., 1927. *Being and Time*. New York: Blackwell Publishers.
- Hennessey, I.A.M., Hewett, P., 2013. Construct, Concurrent, and Content Validity of the eoSim Laparoscopic Simulator. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 23(10), 855-860.
- Hennessey, I.A., Hewett, P., 2014. Virtual reality versus box laparoscopic simulators in trainee selection and aptitude testing. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, 24(4), 318-321.
- Hopkins, C., 1974. How much should you pay for that box? In *the proceedings of the Human Factors and Ergonomics Society annual meeting*. pp.1-10
- Hu, H.H., Gooch, A.A., Thompson, W.B., Smits, B.E., Rieser, J.J., Shirley, P., 2000. Visual cues for imminent object contact in realistic virtual environment. In *the proceedings of the conference on Visualization '00*. Los Alamitos, CA, USA IEEE Computer Society Press. pp.179-185
- Imura, T., Yamaguchi, M., Kanazawa, S., Shirai, N., Otsuka, Y., Tomonaga, M., Yagi, A., 2006. Perception of motion trajectory of object from the moving cast shadow in infants. *Vision Research*, 46, 652-657.
- Iwata, N., Fujiwara, M., Kodera, Y., Tanaka, C., Ohashi, N., Nakayama, G., Koike, M., A., N., 2011. Construct validity of the LapVR virtual-reality surgical simulator. *Surgical Endoscopy*, 25(2), 423-428.
- Johnson, E., 2007. Surgical Simulators and Simulated Surgeons: Reconstituting Medical Practice and Practitioners in Simulations. *Social Studies of Science*, 37(4), 585-608.
- Johnson, S.J., Guediri, S.M., Kilkenny, C., Clough, P.J., 2011. Development and validation of a virtual reality simulator: human factors input to interventional radiology training. *Human Factors*, 53(6), 612-625.
- Kawaguchi, K., Egi, H., Hattori, M., Sawada, H., Suzuki, T., Ohdan, H., 2014. Validation of a novel basic virtual reality simulator, the LAP-X, for training basic laparoscopic skills. *Minimally Invasive Therapy*, 23(5), 287-293.
- Keppel, G., 1991. *Design and analysis: A researchers handbook*. 3rd ed. Englewood Cliffs, N.J: Prentice Hall.
- Kim, H., Rattner, D., Srinivasan, M., 2003. The role of simulation fidelity in laparoscopic surgical training. In *the proceedings of the 6th International Conference on Medical Imaging Computing and Computer-Assisted Intervention*. Montreal, Canada. pp.1-8

- Kirkpatrick, L.D., Kirkpatrick, J.D., 2005. *Evaluating Training Programs*. 3rd ed. San Francisco, CA: Berrett-Koehler Publishers.
- Klatzky, R.L., Lederman, S.J., Hamilton, C.L., Grindley, M., Swendsen, R.H., 2003. Feeling textures through a probe: Effects of probe and surface geometry and exploratory factors. *Perception & Psychophysics*, 65(3), 613-631.
- Kozolowski, S.W.J., DeShon, R.P., 2004. A psychological fidelity approach to simulation-based training. In E. Salas, S.G. Schflett & M.D. Coovert, eds. *Scaled Worlds: Development, Validation, and Applications*. Burlington, VT: Ashgate. pp.75-99.
- Liu, D., Macchiarella, N.D., Blickensderfer, E.L., Vincenzi, D.A., 2008. Transfer of Training. In P.A. Hancock, D.A. Vincenzi, J.A. Wise & M. Mouloua, eds. *Human Factors in Simulation and Training*. Boca Raton, FL: CRC Press. pp.49-60.
- Maguire, M., 2001. Methods to support human-centred design. *International Journal of Human-Computer Studies*, 55(4), 587-634.
- McCluney, A.L., Vassiliou, M.C., Kaneva, P.A., Cao, J., Stanbridge, D.D., Feldman, L.S., Fried, G.M., 2007. FLS simulator performance predicts intraoperative laparoscopic skill. *Surgical Endoscopy*, 21(11), 1991-1995.
- McMahan, R., 2011. *Exploring the Effects of Higher-Fidelity Display and Interaction for Virtual Reality Games*. PhD Dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.
- McMains, K., Weitzel, E., 2008. Low-Fidelity Simulation for Skill Attainment in Endoscopic Sinus Surgery. *The Internet Journal of Otorhinolaryngology*, 11(1).
- Medical-X, 2016. *Lap-X*. [Online] Available at: http://www.medical-x.com/products/lap_x [Accessed 31 March 2016].
- Mégard, C., Gosselin, F., Bouchigny, S., Ferlay, F., Taha, F., 2009. User-centered design of a maxillo-facial surgery training platform. In *the proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (VRST 2009)*. ACM. pp.265-266
- Moroney, W.F., Moroney, B.W., 1999. Flight simulation. In J.A. Wise, V.D. Hopkin & D.J. Garland, eds. *Handbook of Aviation Human Factors*. Mahwah, NJ: Lawrence Erlbaum. pp.355-388.
- Munz, Y., Kumar, B.D., Moorthy, K., Bann, S., Darzi, A., 2004. Laparoscopic virtual reality and box trainers: Is one superior to the other? *Surgical Endoscopy*, 18, 485-494.

Nguyen, D.M.P., Tonetti, J., Thomann, G., 2013. Virtual Reality Coupled with Adapted Physical Interface for a Better Evaluation of the Innovative Surgical Instrument. In M. Abramovici & R. Stark, eds. *Smart Product Engineering*. Springer Berlin Heidelberg. pp.927-936.

Orzech, N., Palter, V.N., Reznick, R.K., Aggarwal, R., Grantcharov, T.P., 2012. A comparison of 2 ex vivo training curricula for advanced laparoscopic skills: a randomized controlled trial. *Annals of Surgery*, 255(5), 833-839.

Palter, V.N., 2011. Comprehensive Training Curricula for Minimally Invasive Surgery. *Journal of Graduate Medical Education*, 3(3), 293–298.

Panait, L., Akkary, E., Bell, R.L., Roberts, K.E., Dudrick, S.J., Duffy, A.J., 2009. The Role of Haptic Feedback in Laparoscopic Simulation Training. *Journal of Surgical Research*, 156(2), 312-316.

Pitzul, K.B., Grantcharov, T.P., Okrainec, A., 2012. Validation of three virtual reality Fundamentals of Laparoscopic Surgery (FLS) modules. In *Medicine Meets Virtual Reality 19*. IOS Press. pp.349-355

Puerta, A., 1989. The power of shadows: shadow stereopsis. *Journal of the Optical society of America*, 6(2), 309-311.

Ragan, E.D., Bowman, D.A., Kopper, R., Stinson, C., Scerbo, S., McMahan, R.P., 2015. Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7), 794 - 807.

Rasmussen, J., 1983. Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions in Systems, Man, and Cybernetics*, 13(3), 257-266.

Reich, O., Noll, M., Gratzke, C., Bachmann, A., Waidelich, R., Seitz, M., Schlenker, B., Baumgartner, R., Hofstetter, A., Stief, C.G., 2006. High-level virtual reality simulator for endourologic procedures of lower urinary tract. *Urology*, 67(6), 1144-1148.

Reznick, R.K., MacRae, H., 2006. Teaching surgical skills--changes in the wind. *The New England journal of medicine*, 355(25), 2664-2669.

Roberts, K.E., Bell, R.L., Duffy, A.J., 2006. Evolution of surgical skills training. *World Journal of Gastroenterology*, 12(20), 3219-3224.

Salkini, M.W., Doarn, C.R., Kiehl, N., Broderick, T.J., J.F., D., Gaitonde, K., 2010. The role of haptic feedback in laparoscopic training using the LapMentor II. *Journal of Endourology*, 24(1), 99–102.

- Sánchez-Margallo, J.A., Sánchez-Margallo, F.M., Oropesa, I., Gómez, E.J., 2014. Systems and technologies for objective evaluation of technical skills in laparoscopic surgery. *Minimally Invasive Therapy & Allied Technologies*, 23(1), 40-51.
- Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Bri, B.M.K., Andersen, D.K., Satava, R.M., 2002. Virtual Reality Training Improves Operating Room Performance Results of a Randomized, Double-Blinded Study. *Annals of Surgery*, 236(4), 458–464.
- Shaharan, S., Neary, P., 2014. Evaluation of surgical training in the era of simulation. *World Journal of Gastrointestinal Endoscopy*, 6(9), 436–447.
- Shimotsu, R.T., Cao, C.G.L., 2007. The Effect of Color-Contrasting Shadows on a Dynamic 3D Laparoscopic Surgical Task. *IEEE Transactions in Systems, Man, and Cybernetics, Part A: Systems and Humans*, 37(6), 1047-1053.
- Simbionix, 2016. *Lap Mentor*. [Online] Available at: <http://simbionix.com/simulators/lap-mentor/> [Accessed 31 March 2016].
- Skinner, A., Auner, G., Meadors, M., Sebrechts, M., 2013. Ambidexterity in laparoscopic surgical skills training. In *Medicine Meets Virtual Reality 20*. IOS Press. pp. 412-416
- Smallman, H.S., St. John, M.S., 2005. Naive Realism: Misplaced Faith in Realistic Displays. *Ergonomics in Design*, 13(3), 6-13.
- Sroka, G., Feldman, L.S., Vassiliou, M.C., Kaneva, P.A., Fayez, R., G.M., F., 2010. Fundamentals of Laparoscopic Surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial. *The American Journal of Surgery*, 199(1), 115-120.
- Stappers, P.J., Gaver, W.W., Overbeeke, K., 2003. Beyond the limits of real-time realism: Moving from stimulation correspondence to information correspondence. In L.J. Hettinger & M.W. Haas, eds. *Virtual and Adaptive Environments*. Mahwah, NJ: Lawrence Erlbaum Associates. pp.91-110.
- Stassen, H.G., Bonjer, H.J., Grimbergen, C.A., Dankelman, J., 2005. The Future of Minimally Invasive Surgery and Training. In Dankelman, J., Grimbergen, C.A., Stassen, H.G. *Engineering for patient safety: issues in minimally invasive procedures*. London: Lawrence Erlbaum Associates Inc. pp.272-282.
- Stefanidis, D., Heniford, B., 2009. The Formula for a Successful Laparoscopic Skills Curriculum. *Archives of Surgery*, 144(1), 77-82.

Steigerwald, S., 2013. *Do Fundamentals of Laparoscopic Surgery (FLS) and LapVR Evaluation Metrics Predict Intra-Operative Performance?* Master thesis. Winnipeg: University of Manitoba.

Steigerwald, S.N., Park, J., Hardy, K.M., Gillman, L.M., Vergis, A.S., 2015. Does laparoscopic simulation predict intraoperative performance? A comparison between the Fundamentals of Laparoscopic Surgery and LapVR evaluation metrics. *American Journal of Surgery*, 209(1), 34-39.

Stoffregen, T.A., Bardy, B.G., Smart, L.J., Pagulayan, R., 2003. On the nature and evaluation of fidelity in virtual environments. In L.J. Hettinger & M.W. Haas, eds. *Virtual and Adaptive Environments*. Mahwah, New Jersey: Lawrence Erlbaum Associates. pp.111-129.

Ström, P., Hedman, L., Särnå, L., Kjellin, A., Wredmark, T., L., F.-T., 2006. Early exposure to haptic feedback enhances performance in surgical simulator training: a prospective randomized crossover study in surgical residents. *Surgical Endoscopy*, 20(9), 1383-1388.

Stunt, J., Wulms, P., Kerkhoffs, G., Dankelman, J., van Dijk, C., Tuijthof, G., 2014. How valid are commercially available medical simulators? *Advances in Medical Education and Practice*, 5, 385–395.

Surgical Science, 2016. *LapSim*. [Online] Available at: <http://www.surgical-science.com/portfolio-gallery/lapsim/> [Accessed 31 March 2016].

Tabachnick, B.G., Fidell, L.S., 1989. *Using multivariate statistics*. 2nd ed. New York: Harper & Row.

Tan, S.C., Marlow, N., Field, J., Altree, M., Babidge, W., Hewett, P., Maddern, G.J., 2012. A randomized crossover trial examining low- versus high-fidelity simulation in basic laparoscopic skills training. *Surgical Endoscopy*, 26(11), 3207-3214.

Thijssen, A.S., Schijven, M.P., 2010. Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? *The American Journal of Surgery*, 199(4), 529-541.

Ullrich, S., Kuhlen, T., 2012. Haptic Palpation for Medical Simulation in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 617-625.

Valdivieso, R.F., Zorn, K.C., 2014. Surgery: Urological laparoscopic training--practice makes perfect. *Nature reviews. Urology.*, 11(3), 138-139.

Våpenstad, C., Buzink, S.N., 2013. Procedural virtual reality simulation in minimally invasive surgery. *Surgical Endoscopy*, 27(2), 364-377.

- Vassiliou, M.C., Ghitulescu, G.A.F.L.S..S.D., Leffondré, K., Sigman, H.H., Fried, G.M., 2006. The MISTELS program to measure technical skill in laparoscopic surgery. *Surgical Endoscopy*, 20(5), 744-747.
- von Zadow, U., Buron, S., Harms, T., Behringer, F., Sostmann, K., Dachsel, R., 2013. SimMed: combining simulation and interactive tabletps for medical education. In *the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Paris ACM. pp.1469-1478
- Waller, D., Hunt, E., 1998. The Transfer of Spatial Knowledge in Virtual Environment Training. *Presence: Teleoperators and Virtual Environments* , 7(2), 129-143.
- Wickens, C.D., Hollands, J.G., 2000. Three dimensions and depth. In *Engineering Psychology and Human Performance*. 3rd ed. Upper Saddle River: Prentice-Hall. Ch. 4. pp.139-142.
- Yang, X., Lee, W., Choi, Y., You, H., 2012. Development of A User-Centered Virtual Liver Surgery Planning System. In *the proceedings of the 56th Human Factors and Ergonomics Society Annual Meeting*. Boston, MA. pp.1772-1776
- Yiannakopoulou, E., Nikiteas, N., Perrea, D., Tsigris, C., 2015. Virtual reality simulators and training in laparoscopic surgery. *International Journal of Surgery*, 13, 60-64.
- Zendejas, B., Brydges, R., Hamstra, S.J., Cook, D.A., 2013. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Annals of Surgery*, 257(4), 586-593.
- Zevin, B., Aggarwal, R., Grantcharov, T.P., 2014. Surgical simulation in 2013: why is it still not the standard in surgical training? *Journal of the American College of Surgeons*, 218(2), 294-301.

Table captions

Table 1: Subjects' groups in the first experiment

Table 2: subjective rating of the early VBLaST prototype relative to the FLS box-trainer

Table 3: Groups' of subjects in the second experiment

Table 4: Results of the Mann-Whitney test comparing the ratings of the two VBLaST prototypes

Figures Captions

Figure 1: The FLS box-trainer

Figure 2: the early VBLaST prototype (VBLaST 1)

Figure 3: Comparing the early VBLaST prototype and the FLS box-trainer (error bars represent the standard error)

Figure4: The modified VBLaST prototype (VBLaST 2)

Figure 5: comparing the performance of the two populations of subjects on the FLS trainer (error bars represent the standard error)

Figure 6: Comparing the performance of the two populations of subjects on the two VBLaST prototypes (error bars represent the standard error)

Figure 7: Comparing the modified VBLaST prototype and the FLS box-trainer (error bars represent the standard error)

Figure 8: comparison of subjective evaluations of the two VBLaST prototypes (error bars represent the standard error)