Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality

Simulation in medical education

A novel full-scale immersive team-based simulation environment for objective performance assessment of surgical excellence

Patrick Wucherer

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r): Univ.-Prof. Dr. Hans Michael Gerndt

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Nassir Navab
2. Prof. Pierre Jannin, Ph.D. Université de Rennes 1
3. Prof. Roy Eagleson, Ph.D. University of Western Ontario

Die Dissertation wurde am 02.05.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 12.11.2016 angenommen.

Abstract

In traditional surgical education the Halstedian approach is applied for the transfer of surgical skills from the expert to the novice. The trainee has to acquire technical and non-technical skills within a high-risk environment and in direct contact with patients. Obviously this approach collides with the Hippocratic oath: "Primum non nocere - First, do no harm" and hence is ethically questionable. Further the knowledge gained by the trainee is limited to the experience of the mentor and the cases the trainee is exposed to during surgery. The training model "proficiency-based progression" (PBP) addresses this circumstance and illustrates a paradigm shift in surgical education from the traditional surgical curriculum towards a "simulation-based medical education" (SBME). It focuses on the training of technical skills namely the craftsmanship e.g. usage of instruments, procedural knowledge e.g. workflow and cognitive skills. These can be achieved by using a variety of simulation models ranging from bench-top models e.g. suture pads, virtual reality (VR) emulators/simulators, full-scale VR simulators, and human cadavers to living animals. However apart from the required technical skills it is a fact that non-technical skills cannot be neglected as non-technical failures are often root causes for harm and near-misses in the operating room (OR).

Thus the research community is requesting novel approaches for multidisciplinary team training. (i) Appropriate high fidelity VR simulators for team-based training are non-existant and (ii) there is a lack of knowledge on correlating technical and non-technical performance indicators hindering efficient and effective training support for individuals and teams.

The aim of my research work was to investigate the development of a simulated OR theatre which could be an option for conducting assessment and training of surgical skills within a multidisciplinary controlled environment. For the first time a VR surgical procedural simulator and an anesthesia computerized mannequin were fully integrated and functioned as one simulation setup. In a pilot study expert surgeons were immersed in the medical simulation environment through task and crisis scenarios of a typical vertebroplasty workflow. The face validity of the simulation environment was confirmed by investigating surgeon behavior and workflow response. The result of the conducted user-study corroborated the unique medical simulation concept of combining VR and human multisensory responses into surgical workflow. A second study with residents was conducted (i) to provide a qualitative measure of usability, (ii) to assess vertebroplasty technical performance of the surgeon, and (iii) to explore the relationship between mental workload and surgical performance during crisis. The results indicated that (a) the surgeons scored the face validity of the modeled simulation environment very highly, (b) surgeon training enabled completion of tasks more quickly, and (c) the introduction of crisis scenarios negatively affected the surgeons' objective performance. Taken together, the results underscored the need to develop realistic simulation environments that prepare young residents to respond to emergent events in the OR.

These findings will enable other national/international research and simulation centers to develop more sophisticated training environments and the hope exists that this work will trigger a new generation of simulators which will eventually improve medical education of multidisciplinary teams and hence increase patient safety and raise the quality of patient care.

Zusammenfassung

In der traditionellen medizinischen Ausbildung wird der Ansatz von Halsted verwendet, um chirurgische Fähigkeiten vom Experten zum Anfänger zu übertragen. Der Auszubildende muss technische und nicht-technische Fähigkeiten in einem risikoreichen Umfeld und im direkten Kontakt mit Patienten erlernen. Offensichtlich kollidiert dieser Ansatz mit dem Hippokratischen Eid: "Primum non nocere - Erstens, nicht schaden" und ist dementsprechend ethisch fraglich. Des Weiteren ist das vom Auszubildenen erlernte Wissen begrenzt durch die Erfahrung vom Mentor und die Fälle, die der Auszubildende mit behandelt. Das Trainingsmodell "proficiency-based progression" (PBP) befasst sich mit diesem Umstand und stellt einen Paradigmenwechsel dar in der chirurgischen Ausbildung von dem traditionellen Curriculum zu einer "simulation-based medical education" (SBME). Der Fokus liegt auf dem Training von technischen Fähigkeiten nämlich dem Handwerk z.B. Nutzung der Instrumente, dem prozeduralen Wissen z.B. Arbeitsablauf und kognitive Fähigkeiten. Diese können erlernt an verschiedenen Simulationsmodellen werden angefangen von Tischmodellen z.B. Nahtkissen, Virtual Reality (VR) Emulatoren/Simulatoren, full-scale VR Simulatoren, und menschlichen Leichen bis hin zu lebenden Tieren. Jedoch ist es Fakt, dass neben den benötigten technischen Fähigkeiten, die nicht-technischen Fähigkeiten nicht vernachlässigt werden dürfen, da diese oft die Hauptursachen für Schäden und Beinahe-Schäden in dem Operationssaal sind.

Folglich fordert die Forschungsgemeinschaft neue Ansätze für das multidisziplinäre Team Training. (i) Geeignete High-Fidelity VR Simulatoren für team-basiertes Training sind nicht vorhanden und (ii) es besteht eine Wissenslücke technische und nicht-technische Leistungsindikatoren miteinander in Beziehung zu setzen. Dies verhindert eine effektive und effiziente Trainingsunterstützung von Einzelpersonen und Teams.

Das Ziel meiner wissenschaftlichen Arbeit bestand darin die Entwicklung eines simulierten Operationssaals zu untersuchen, welcher eine Option für die Durchführung von Leistungserfassung und Training von chirurgischen Fähigkeiten in einer multidisziplären kontrollierten Umgebung darstellt. Zum ersten Mal wurden ein VR chirurgischer prozeduraler Simulator und eine computerisierte Anästhesiepuppe vollständig zusammengeführt und als ein Simulationsaufbau eingesetzt. In einer Pilotstudie wurden erfahrene Chirurgen in die Simulationsumgebung durch Aufgaben und Krisenszenarien in einen typischen Vertebroplastie-Arbeitsablauf hineinversetzt. Die "Face Validity" der Simulationsumgebung wurde bestätigt durch die Betrachtung der chirurgischen Handlungen und Abarbeitung der Arbeitsschritte. Das Ergebnis der durchgeführten Benutzerstudie untermauert das einzigartige medizinische Konzept, die Kombination - virtuelle Realität und menschliches multisensorisches Ansprechen in einem chirurgischen Arbeitsablauf. Eine zweite Studie wurde mit Assistenzärzten durchgeführt, um (i) eine qualititative Aussage zur Nutzbarkeit, (ii) die technische Leistung eines Chirurgen während einer Vertebroplastie zu erfassen, und (iii) die Relation zwischen Arbeitsbelastung und chirurgischer Leistung in einem Krisenszenario zu erforschen. Die Ergebnisse zeigen, dass (a) die Chirurgen die "Face Validity" als sehr hoch einstufen, (b) das Training von Chirurgen die Durchführung der Aufgaben beschleunigen konnte, und (c) die Einführung von Krisenszenarien sich negativ auf die chirurgische objektive Leistung auswirkte. Insgesamt unterstreichen die Ergebnisse den Bedarf an der Entwicklung von realistischen Simulationsumgebungen, welche junge Assistenzärzte auf auftretende Ereignisse im Operationssaal vorbereiten.

Diese Erkenntnisse ermöglichen anderen nationalen/internationalen Wissenschafts- und Simulationszentren die Entwicklung von komplexen Trainingsumgebungen und die Hoffnung besteht, dass diese Arbeit eine neue Generation von Simulatoren auslöst, welche letztendlich die medizinische multidisziplinäre Ausbildung verbessern und dadurch die Patientensicherheit verbessern und die Qualität der Patientenversorgung erhöhen.

Acknowledgements

I would like to start by acknowledging my supervisor and my mentor, Prof. Dr. Nassir Navab, whose expertise, understanding, and patience, added considerably to my graduate experience. Throughout my Ph.D., he has always been a source of guidance and support, and I truly feel honoured to have had the opportunity to learn from him and to benefit from his breadth of knowledge and deep insights.

Further I would like to express my sincere gratitude to my loyal colleague Philipp Stefan, the best senior research scientist ever Dr. Pascal Fallavollita and the inspiring surgeon Dr. med. Simon Weidert. I am truly thankful for their help in writing papers and conducting experiments, and for their assistance at all levels of this research project. This work would have not been possible without them. In particular I enjoyed the great research trips to Japan, Canada and France. Thank you for those memorable days and experiences. Then I would like to express my sincere gratitude to my co-supervisor, Prof. Dr. med. Ekkehard Euler for his support, the opportunities to get inside a real OR and his expertise on procedural workflows.

Special thanks goes out to Dr. Matthias Weigl, Dr. med. Marc Lazarovici, Dr. med. Ulla Stumpf, Dr. med. Anna von der Heide for sharing their expertise and providing advice and assistance at every level of this research work. It has truly been a privilege to work with and learn from them.

Moreover I would like to thank Prof. Roy Eagleson, as well as Dr. Sandrine de Ribaupierre for their support on setting up studies, fruitful discussions and valuable advice along the way.

Furthermore I would like to say thank you to Prof. Russell Taylor who invited me to his research laboratory LCSR (Laboratory for Computational Sensing and Robotics) at the Johns Hopkins University in Baltimore. It was a pleasure to work with him. Additionally I would like to thank Prof. Pierre Jannin for the warm welcome at his research laboratory LTSI (Laboratoire Traitement du Signal et de l'Image) at the University of Rennes I in Rennes.

I like to thank Prof. Dr. med. Wolf Mutschler for his support to conduct the study with 19 surgeons.

Further I would like to thank the NARVIS lab crew consisting of Severine Habert, Ma Meng, Ulrich Eck and former members Dr. Tobias Blum, Dr. Ahmad Ahmadi, Dr. Tassilo Klein as well as many others at the Chair for Computer Aided Medical Procedures (CAMP) for their thoughtful discussions, helpful criticisms, and, most of all, for their friendship.

I also like to thank all the surgeons, residents, and students who participated in my studies.

Thank you, Sarah Leese, for accompanying and supporting me during my Ph.D. adventure.

Finally, I would like to thank my parents, Monique and Andreas, my sister Isabelle and my brother Frédéric for their unconditional love and support. I am truly grateful for having them by me and none of my accomplishments would have been possible without them.

Contents

"The art of medicine is to be learned only by experience, 'tis not an inheritance; it cannot be revealed. Learn to see, learn to hear, learn to feel, learn to smell, and know that by practice alone can you become an expert." [Sir William Osler, M.D.; 1920]

"No industry in which human lives depend on the skilled performance of responsible operators has waited for unequivocal proof of the benefits of simulation before embracing it." [David M. Gaba, M.D.; 1992]

1. Introduction

In the last century the surgical training model focused on the knowledge transfer during surgeries and the teaching of scientific principles. In the past few years it has started to evolve towards a simulation-based training model [98] also known as simulation-based medical education (SBME) [89], implying 'devices, trained persons, lifelike virtual environments and contrived social situations that mimic problems, events, or conditions that arise in professional encounters' [8]. Hence surgical trainees can (i) "practice clinical skills under safe, controlled, forgiving conditions". Additionally they benefit from undergoing (ii) "formative assessment", and (iii) "receiving focused feedback with the aims of acquiring and maintaining clinical competence" [89].
Nowadays the use of simulation in surgical training has gained significant momentum [27]. It was only recently that the prominent American Academy of Orthopedic Surgeons (AAOS), in conjunction with the American Board of Orthopedic Surgery (ABOS) and the Arthroscopy Association of North America (AANA), acknowledged the importance of simulation-based education [7].
First, it is a fact that "practice is the key for the acquisition and retention of medical skills" [117]. Second, simulation has shown promising results in several other high-risk industries including the military, aviation, and astronautics and is well established [27].
Additionally the ethical issue is a valid argument against the traditional surgical education. It is based on the Halstedian approach of 'see one, do one, teach one' also known as 'apprenticeship model', 'surgical preceptorship' [9] [115] and 'learning by doing' [70]. The technical and procedural knowledge is passed by the expert to the trainee during surgeries [9]. This leads to the inevitable exposure of patients to inexperienced practitioners which does not correspond to one of the principal beliefs of the Hippocratic Oath: "Primum non nocere - first, do no harm."
Further work hour restrictions e.g. the European Union Working Time Directive (EWTD) was applied in 2009 which included the reduction of working time of medical doctors to 48 hours per week with a voluntary option of working 56 hours. The intent was to ensure patient safety and quality of care by providing medical doctors a better work-life balance. However, it limits the time of exposure to clinical practice of surgical trainees [46][9].
Moreover the cost pressures and patient safety concerns exert pressure on surgical trainees to acquire increasingly complex surgical skills in shorter time periods, with fewer opportunities to operate [79]. Cost analysis of surgical resident training has shown that the cost per resident has increased, primarily as a result of increased operative times [36].
Then the rapid introduction of minimal invasive surgeries (MIS) in the 1990s radically changed the perspective on surgical training [46]. Enthusiastic adopters of new technology for new ways of patient treatment quickly spread the news at international congresses and symposiums and the media helped by establishing a catchy term "the keyhole surgery" to capture the world's attention. The advantages of performing MIS are the smaller incisions, less pain for patients and faster recovery of patients and less costs due to much shorter hospital stay. However there are limitations the new treatment was and is still facing, since it requires from the surgeon psychomotoric and perceptual skills e.g. moving long instruments inside the human body with limited tactile and haptic feedback while looking at a monitor. Probably the biggest difficulty is the counterintuitive movement of surgical instruments in particular in laparoscopy. Thus the introduction of MIS was associated with a higher complication rate than compared to open surgery and it became apparent that the Halstedian approach of training is not a viable training model any more.
And more than a decade ago public scrutiny of the 'To Err is Human' report has been unprecedented [71]. It is now widely accepted that nearly 10% of all patients admitted to hospital will likely be unintentionally harmed in some way. There are more deaths annually as a result of health care errors than other forms. Beyond their cost in human lives, preventable medical errors also take other heavy toll. Errors have been estimated to result in costs totalling between $17-29 billion per year in hospitals in the US, and from a recent 2013 study, between 210.000-400.000 deaths annually [63]. In the European Union Member nations, facts consistently show that medical errors and health-care related crisis events occur in 8-12% of hospitalizations. For example, the United Kingdom Department of Health estimated about 850.000 crisis events per year, with Spain, France and Denmark having published incidence studies with similar results [108]. Studies continuously show human

factors as the key contribution for adversity in the operating room (OR) theatre [49][80][116][52][54]. These factors are commonly known as *non-technical skills* which are recognized as key root causes of surgical errors worldwide consisting of *cognitive* e.g. situational awareness and decision-making, *personal resource* e.g. coping with stress and fatigue and *interpersonal skills* e.g. communication and teamwork [62]. A study conducted in 2008 by [21] in an emergency department underscore those assumptions that most of the errors for mortality cases involved several contributing factors as shown in Table 1.2, with communication and assessment as the leading causes.

2004 through June 2013 [N=845] *The majority of events have multiple root causes*	
Communication	634
Assessment	619
Human Factors	545
Leadership	535
Information Management	247
Continuum of Care	212
Care Planning	141
Physical Environment	134
Medication Use	61
Patient Rights	20

Table 1.2: Root causes of medical errors.

Obviously surgical competence which is required to perform a successful operation consists of non-technical skills e.g. teamwork, communication, leadership and judgement (see Table 1.2), technical skills and procedural knowledge [126] [4]. Yet surgical excellence is merely associated with technical performance neglecting the importance of non-technical skill. Furthermore performance measurements are traditionally focusing on assessing technical proficiency either based on rating scales and/or motion analysis [132][25] [86] [34]. This situation can be compared to the situation which existed over 30 years ago in the aviation industry. It concentrated on the "training of technical skills" of pilots and the "technological development" [95]. However since research findings sponsored by the National Aeronautics and Space Administration revealed that 70% of errors were routed to human causes e.g. failed communication, interpersonal communication and leadership [3] [95], the focus was put on the design and development of crew resource management (CRM). CRM provides a curriculum for flight crews ranging from seminars, lectures to simulation-based training in flight decks [100]. The application of CRM led to a profound knowledge of the existing "limitations of human performance" and furthermore fostered a culture of safety [3]. This program spread due to its success to other high reliability organizations e.g. nuclear industries, military [38] and even medicine. Within the medical domain anesthetists adapted it and created anesthesia CRM (ACRM). It is derived from the principle of repeated practice as CRM [61]. The participants are exposed to crisis scenarios making best use of the available resources. According to this principle today anaesthetists are trained on computerized mannequins, which are replications of human patients, in common and rare crisis scenarios [96] having the primary objective of error management [59] which is a crucial aspect of surgical performance just as in aviation [96].

Thus novel approaches for training of surgical competence have to be formulated. The general consensus from the literature and simulation program directors [77] is that simulation has become an important and necessary component of medical student and resident training, to target skills and practices that are difficult to acquire through traditional training, to reduce risks to patients, and potentially, to deliver more cost-effective training. Gallagher et al. proposed in 2012 a novel training paradigm called "proficiency-based progression (PBP)" training which integrates simulation-based training into the surgical curriculum though mainly considering the technical skill acquisition and neglecting the non-technical skill training. Though a recent study in 2015 recommended that: ”it is imperative that surgical residents undergo simulation training directly linked to their hospital responsibilities so as to provide immediate performance improvement and reduce errors in the clinical environment” [15]. Medical simulation training with computer-controlled equipment provides an environment for acquiring knowledge, skills and attitudes without putting patients' health at risks [75]. It offers a highly standardized environment for objective technical and non-technical performance assessment [46]. The possibility to repeatedly practice procedures enables correction of mistakes, mediation of error recovery strategies, skill amelioration and clinical outcome optimization [51]. Moreover medical experience can be gained by conducting difficult procedures or even by inducing complications affecting the workflow of the procedure.

Thesis Objectives The aim of the presented fundamental research work is the exploration and implementation of the requirements and technological necessities of a low-cost, yet highly realistic VR surgical procedural simulator for inter-disciplinary/multidisciplinary medical team training. It provides the basis for ”understanding of the interactions between specific skills and technical performance” and therefore ”tailor-made training packages aiming at specific skills for specific grades of surgical expertise can be developed, implemented, and evaluated” [62]. Furthermore multidisciplinary team training in a simulated environment including crisis scenarios has the potential to reduce the occurrence of adverse events in the real OR. It can enable the medical team to function in a more effective and efficient manner when crises occur in real OR environments. Therefore the key contributions are:

- Scientific evaluation and validation of the combination of a VR surgical procedural simulator and computerized mannequin in a novel training setup for medical assessment and training (see Fig. 1).
- Research and development of a surgical simulator reflecting complete workflows of surgical procedures, including intra-operative crisis scenarios which potentially result in adverse events. The capability to deliberately expose the trainee to realistic adverse events facilitates the research on robust error metrics for performance assessment in adverse or critical situations, as well as on strategies for augmentation and amplification of error perception and mediation of error recovery strategies, e.g. in deliberate practice or procedure rehearsal scenarios.
- Establishment of structured assessments of surgeons' and surgical team's technical and non-technical skills during simulation-based training that enhance impact on the acquisition, application, and retention of teamwork skills in healthcare.

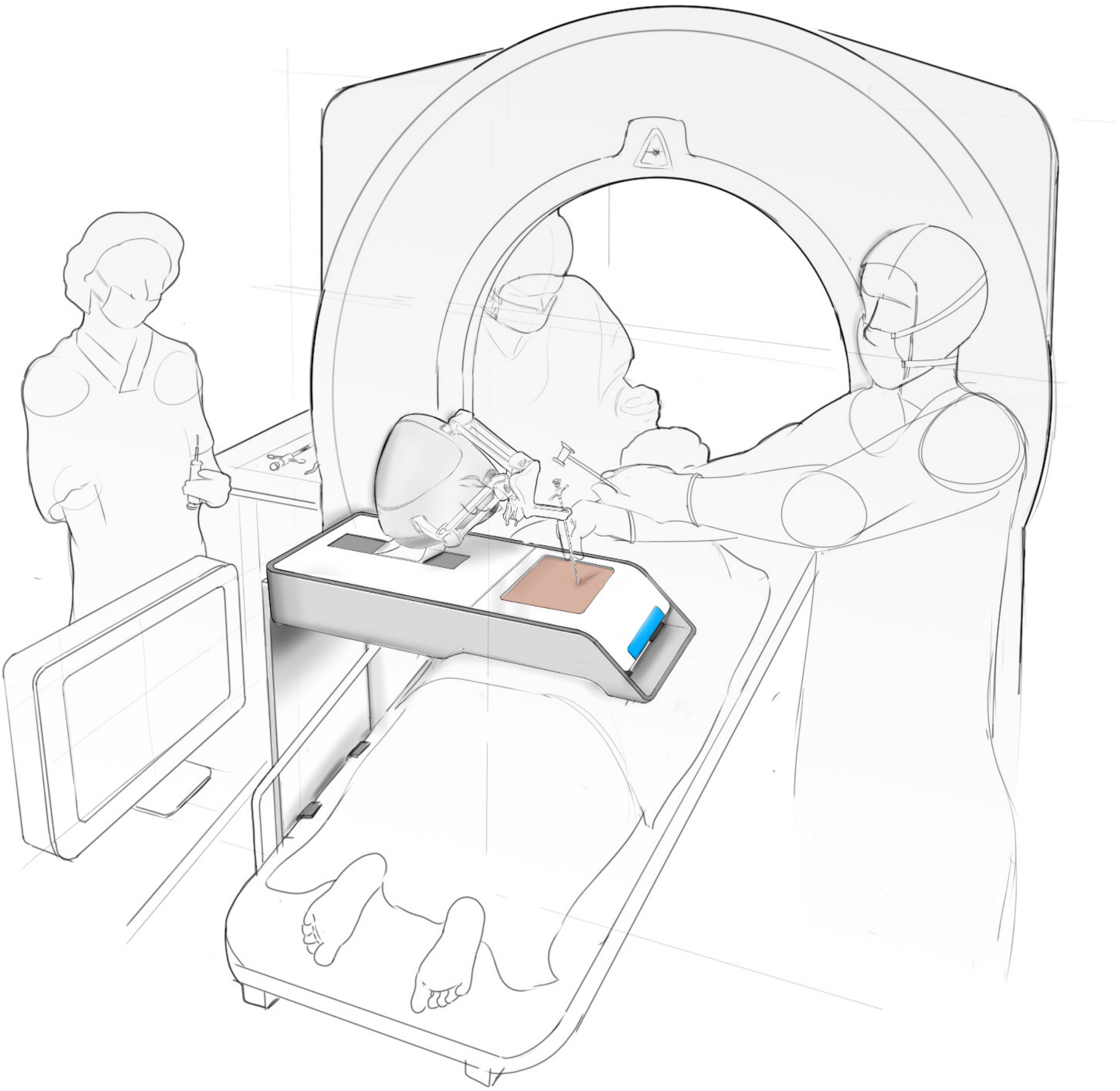


Fig. 1: Sketch of simulated operating room theatre with CT scanner mock-up.

Finally, when designing simulators it is important to adhere to the rule of Dr. David Gaba, pioneer in simulation and Director of the Center for Immersive and Simulation based Learning (CISL) at Stanford Medical School who states: "simulation is a technique, not a technology, to replace or amplify real experiences with guided experiences, often immersive in nature, that evoke or replicate substantial aspects of the real world in a fully interactive fashion"[44].

Thesis structure First an overview of the traditional surgical training is given. Then a novel paradigm for surgical education is described called "proficiency-based progression" (PBP) aiming at efficiency and effectiveness. It comprises seven steps to wisdom including the education bandwidth ranging from learning from text books to learning on live patients. The state of the art research within the time frame 1996-2012 on team-based training is summarized with the conclusion that mainly three key problem statements exist: (i) lack of immersive full-scale simulation environment for multidisciplinary team training due to non-existing integration of surgical and patient simulator [23], (ii) absence of validated and robust technical and non-technical performance metrics [65] and (iii) lack of knowledge on correlating technical and non-technical performance indicators hindering efficient and effective training support for individuals and teams.
Hence this thesis proposes a framework for designing and validating multidisciplinary team-based training technical setups consisting of a virtual reality simulator and a mannequin, which could be used in future to validate metrics for the assessment of technical or non-technical performance and furthermore provide an environment for testing technology e.g. novel imaging devices, robots or instruments.
The main part of this thesis consists of two studies in which the simulated OR is validated and a procedure-specific performance metric is developed [148] [149] [147] [141]. Additionally a study is described on assessing the visuospatial reasoning of surgeons using the virtual reality surgical simulator component of the multidisciplinary team training environment.
In the summary the broad picture of the paradigm shift in medical education is drawn with the results of the conducted research work. Then the next steps for the development and the studies are illustrated. Additionally research questions are posed. Finally an outlook for further research work is depicted on how to fully make use of the simulation environment with currently existing barriers for simulation-based training of surgeons being highlighted.

2. State of the Art: Surgical Education

2.1. Traditional surgical education

Traditional surgical education consists of three steps (see Fig. 2) (i) simple explanation (a)) of how to conduct tasks e.g. knot tying, (ii) the attendance of a mandatory lecture (b)), (iii) and finally the novice is introduced into the operating room (OR) (see Fig. 3): Within the OR the surgical trainee is educated on the patient using the Halstedian approach "See one, do one, teach one" (c)) also called 'apprenticeship model'. It served surgery well for more than a century as this training model gradually transfers the responsibility for the patient treatment in the OR to the trainee under supervision of an expert surgeon.

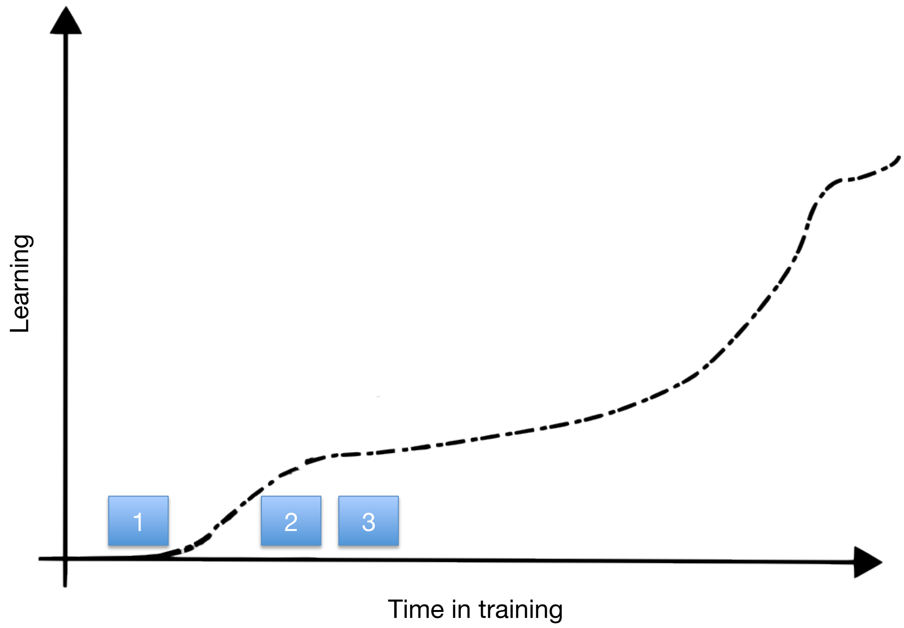


Fig. 2: Traditional medical curriculum (image taken from Gallagher 2012).

Fig. 3: From left to right: Textbook[a], lecture[b], operating room[c].

[a] http://amberb907.wikispaces.com/file/view/Anatomy-and-Physiology-book.jpg accessed Apr/21/2016.
[b] http://amberb907.wikispaces.com/file/view/Anatomy-and-Physiology-book.jpg accessed Apr/21/2016.
[c] http://www.surgeons.org/media/18938218/timor_surgeons_photo_ellen_smith_497x330.jpg accessed Apr/21/2016.

2.2. Towards proficiency-based progression learning: Seven steps to wisdom

Gallagher et al. [46] proposed in 2012 the 'proficiency-based progression (PBP) training paradigm' (see Fig. 4 and Table 2.1) which is an augmentation of the traditional surgical curriculum (see Fig. 2), as it integrates simulation-based training into the traditional training curriculum. It is grounded on a quantitive definition of proficiency using metric-based performance units. This means that the learning process is supported with relevant and timely information with the main focus to be efficient and effective. The training is of reinforcing nature, as the feedback is constructive and formative.

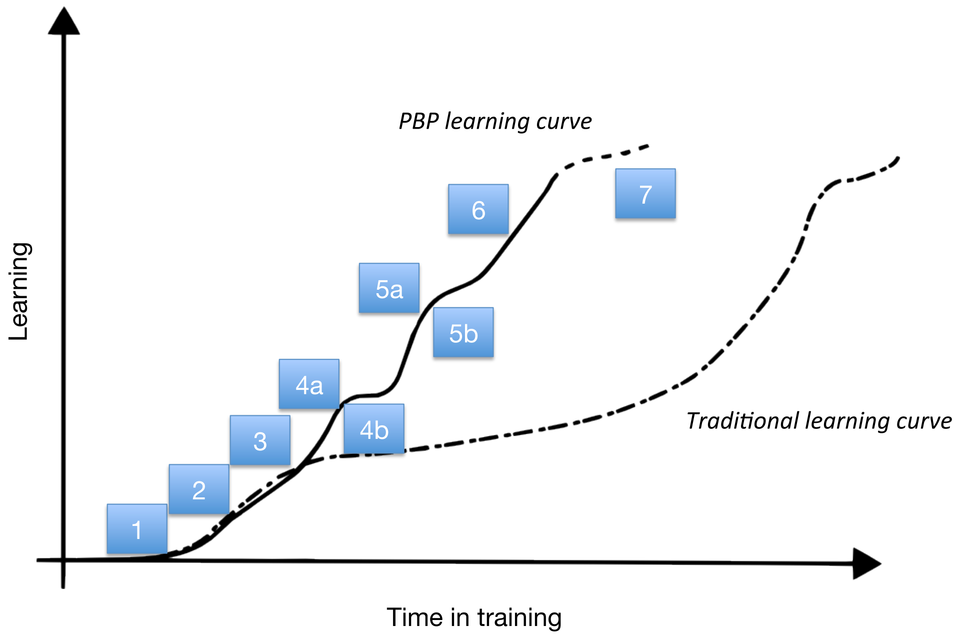


Fig. 4: Diagrammatic representation of the "estimated" speed of learning with proficiency-based progression (PBP) training and traditional training as a function of time in training adapted from Gallagher 2012.

Outline of the seven steps to wisdom

1. Textbook
2. Lecture
3. Online platforms for learning
4. Emulation models
   (a) Silicon, animal model tissue (no formative feedback)
   (b) Silicon, animal tissue, Minimally Invasive Surgical Trainer - Virtual Reality (MIST VR) (formative feedback)
5. VR Procedural simulation
   (a) Full procedural VR simulation with summative metrics
   (b) Full physics VR simulation with proximate formative and summative metrics
6. Real patients with good mentoring and feedback
7. Wisdom acquisition

Step	Simulation technique	Costs	Realism	Reusability	Hygiene	Ethics	Metrics
4	**Bench-top models**						
	Animal parts	+	o	-	-	+	-
	Synthetic materials	o	+	o	+	+	-
	Computer-based						
3	Online education/Simulation models						
4	*VR emulators*						
	e.g. MIST VR	o	-	o	+	+	+
5	*VR simulators*						
	Part-task e.g. LapSim	+	o	+	+	+	o
	High fidelity e.g. CAE, Simbionix, VirtaMed simulators	-	+	+	+	+	o
	Mannequins	-	+	+	+	+	o
	Full physics e.g. VIST	-	+	+	+	+	+
6	*Real tissue*						
	Animal models, human cadavers	-	o	-	-	-	-
	Real patients		+	-	-	-	-

Table 2.1: Simulations for the various steps of PBP learning curriculum. Table is adapted from Gallagher 2012. (+) advantage (o) no advantage or disadvantage (-) disadvantage.

2.2.1. Steps 1-3: Textbook, lectures and online platforms for learning

The PBP learning curriculum includes the first two steps of the traditional learning process, then it progresses to online education with formative assessed explanation (see Fig. 5).

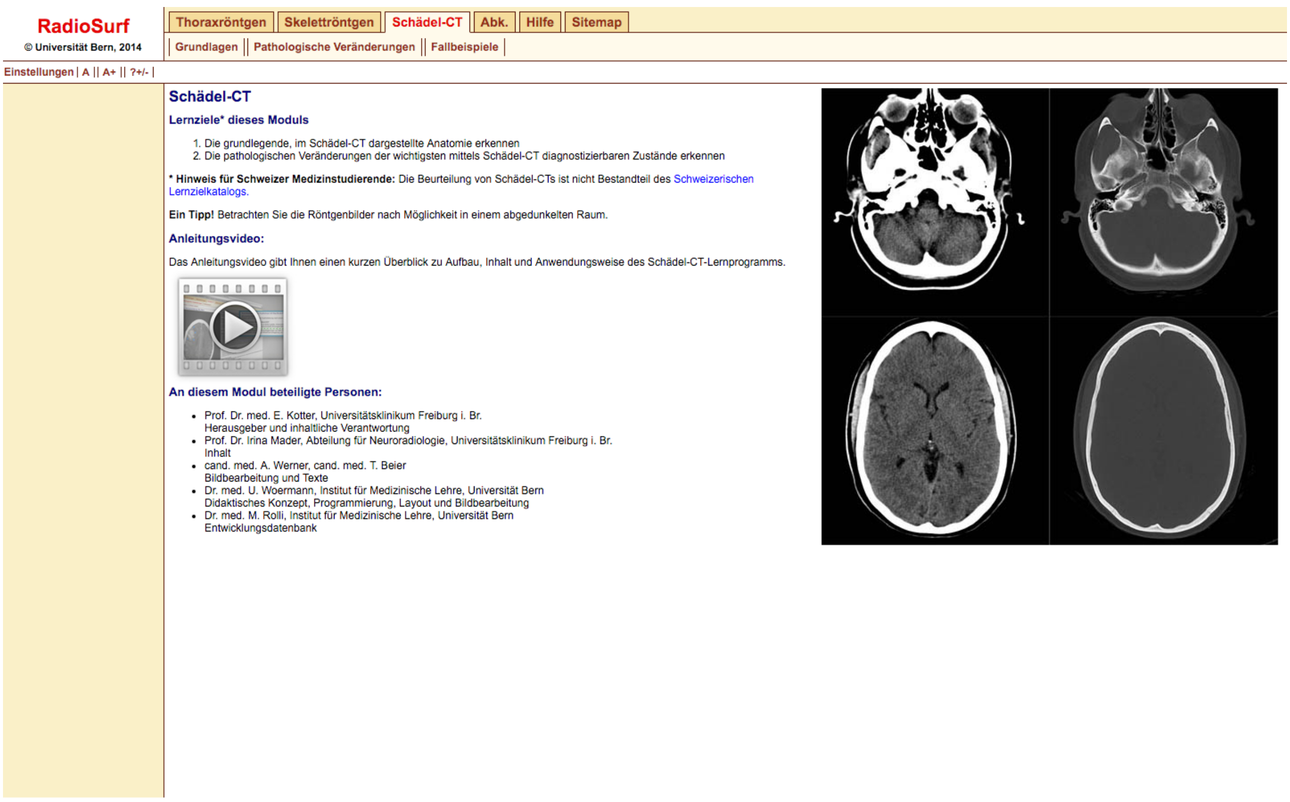


Fig. 5: Online platform[a].

[a] http://e-learning.studmed.unibe.ch/radiosurf/htmls/radskullct.html?radiosurf|radskullct accessed Apr/21/2016.

2.2.1.1. Extension of online education with simulation models

Since the web provides an effective and efficient distribution channel for educational material at any time, it can prepare the trainee with knowledge and/or skills in a timely and relevant manner. A few examples for the integration of online education in a medical curriculum are the Fundamentals of of Laparoscopic Surgery (FLS) and the "Surgical Conferencing with enHanced Opportunities for Online Learning (SCHOOL) for Surgeons". The FLS consists of two main components, the online e-learning platform and a technical skill component (see Fig. 6). First trainees acquire knowledge online and then perform laparoscopic tasks on the technical skill component within a technical skills training program which was validated in major parts by Gerry Fried (McGill University, University of Montreal). The other well known online education is SCHOOL for Surgeons which has been developed by the Royal College of Surgeons in Ireland (RCSI). Online tutors work with surgical trainees e.g. together in short courses, give and correct assignments, discuss literature and is mainly based on self-directed learning. The online education is coupled with a technical skills training program which takes place a certain number of days per year [46].

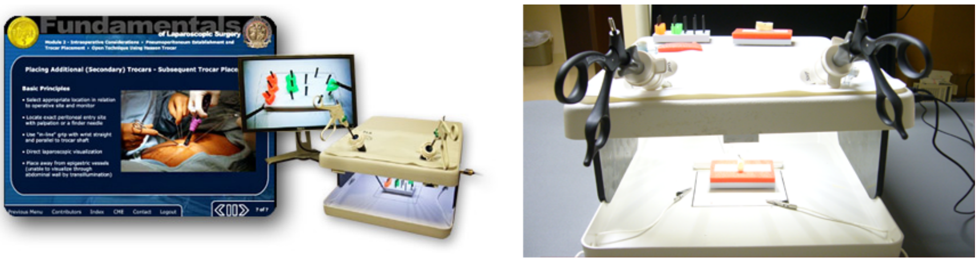


Fig. 6: From left to right: Online e-learning platform[a]; technical skill training and assessment component[b].

[a]Source: `http://sites.uci.edu/ucisurgicaleducation/fls-test-center/` accessed Feb/25/2016.
[b]Source: `http://laparoscopy.blogs.com/prevention_management_3/Tables_and_Figures/RST%20Figure%203.jpg` accessed Feb/25/2016.

2.2.2. Steps 4-5: Technical and procedural training in simulation labs

The simulation labs exist as a learning environment for trainees to acquire necessary technical, procedural and cognitive skills. Within these simulation labs various simulations can be used for practice and assessment of technical procedures. The purpose of low fidelity e.g. bench-top models is to train technical skills of instruments. Additionally medical imaging simulators e.g. Ultrasound simulators [12] are used to learn how to handle medical imaging devices. In the following section the range of the available simulations for technical and procedural training are depicted. Step 4 consists of using bench-top models and step 5 VR simulators for training.

Bench-top models

Animal tissue Animal tissue e.g. pieces of pork, chicken, bowel or liver (see Fig. 7) has already been in use for a few decades for [46] [27] training surgical skill. They are the most basic simulation materials for training surgical tasks e.g. suturing, makeing closure of incisions. The access to these types of models is easy as they can be bought in butcher shops and are rather inexpensive. The drawback is the hygiene and the technical performance while conducting the procedure step can be hardly assessed. Just the outcome can be assessed by the trainer objectively by e.g. measuring distances between aligned sutures.

Fig. 7: From left to right: chicken leg[a], pigs heart and liver[b], pigs trotter[c], pigs bowel[d].

[a]Source: http://www.glatt-organics.com/wp-content/uploads/2013/08/iStock_000012408317Large.jpg accessed Apr/21/2016.

[b]Source: http://3.bp.blogspot.com/-Fkh9cpzAjCg/Ur4gORH1DsI/AAAAAAAAl0Y/0MlG-Nb5wfc/s1600/heart+and+liver.jpg accessed Apr/21/2016.

[c]Source: https://colonelmustardinthekitchen.files.wordpress.com/2013/10/20131015_113034.jpg accessed Apr/21/2016.

[d]Source: https://classconnection.s3.amazonaws.com/978/flashcards/891978/jpg/pig_intestine1334116187297.jpg accessed Apr/21/2016.

Synthetic models Since the introduction of MIS e.g. Laparoscopy at the beginning of the 1990s the use of synthetic training models (see Fig. 8) increased, though these have already been used for a considerable period of time for surgical training [46]. They offer good face validity as they look like real anatomy and there is no issue on the hygiene aspect. However they have limited life time as they can only be used once and are sometimes quite expensive.

Fig. 8: From left to right: Open inguinal hernia trainer [a], open inguinal hernia close-up [b], gallbladder [c], saphenofemoral junction ligation [d], ingrowing toenail trainer [e].

[a]Source: https://www.firehousemedical.com/store3/manufacture/limbs-things.html?limit=20&order=name&dir=desc&p=2 accessed Apr/21/2016.

[b]Source: http://shoponline.cardiac-services.com/product_images/LIM60427_L.jpg accessed Apr/21/2016.

[c]Source: https://www.limbsandthings.com/global/our-products/details/gall-bladder-with-wide-common-bile-duct-stones-for-exploration accessed Apr/21/2016.

[d]Source: https://www.firehousemedical.com/store3/manufacture/limbs-things.html?limit=20&order=name&dir=desc&p=2 accessed Apr/21/2016.

[e]Source: http://assets.limbsandthings.com/products/80060_2.jpg accessed Apr/21/2016.

In 2009, the Minimal-invasive Interdisziplinäre Therapeutische Intervention (MITI) research group at the TUM developed the Endoscopic-Laparoscopic Interdisciplinary Training Entity (ELITE) simulator. ELITE is an ex vivo model designed to train conventional laparoscopic and endoscopic skills and to perform hybrid interventions. It includes realistic and functional replications of intestinal organs and is adapted to the requirements of laparoscopy, endoscopy and Natural Orifice Transluminal Endoscopic Surgery (NOTES). The latest initiative in the field of medical training is the innovative surgical training technologies (ISTT) project which has been started by the Hochschule für Technik, Wirtschaft und Kultur Leipzig (HTWK Leipzig) in 2010. As in the ELITE simulator, tactile feedback is simulated using replicated synthetic anatomy models.

VR simulations In contrary to reality, virtual reality (VR) exists. It is the representation of the reality in a real-time environment to create the impression of reality. The first approaches to create a learning/training environment for surgical skills using VR were started around 1990 [120] sponsored by the Defense Advanced Research Projects Agency (DARPA, US) by millions of dollars [46]. Two prototypes of a anastomosis simulator were developed, however they were hardly validated.

Part-task VR emulators The purpose of VR emulators is to imitate particular aspects of a task which have to be trained. A well-known example of a VR emulator is the MIST VR [146]. The development of this emulator was driven by the question "what skills are we trying to train and assess?" with the conduction of a laparoscopic cholecystectomy in mind. To date it remains the best validated VR emulator in surgery (see Fig. 9-left).

Part-task VR simulators In contrast to emulators, simulators aim at realistical reflection of as many as possible aspects of the tasks. LapSim of Surgical Science Sweden AB for example is a part-task VR simulator as the MIST VR described previously however the tasks look tissue-like (see Fig. 9-right).

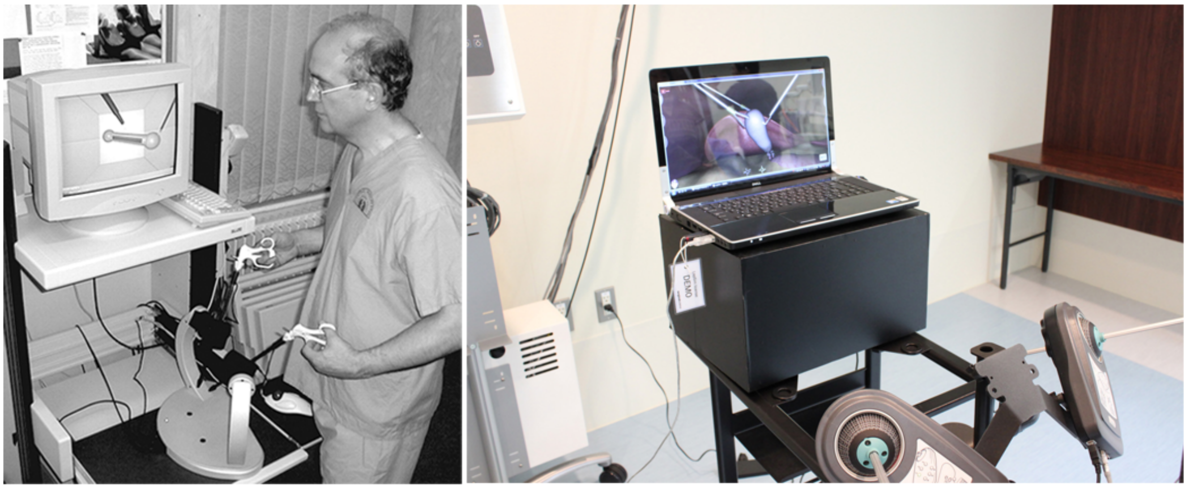


Fig. 9: From left to right: MIST VR[a], LapSim[b].

[a]Source: http://rstb.royalsocietypublishing.org/content/366/1562/276 accessed Apr/21/2016.
[b]Source: http://blog.sannoudai.or.jp/?eid=88110 accessed Apr/21/2016.

The advantages of such types of simulators like the MIST VR and the LapSim are cost-effectiveness, build-in metrics on task performance, rarelys require technical support, and can be used almost anywhere.

High fidelity VR simulators High fidelity VR simulators incorporate full procedures aiming at training procedure steps of full procedures. In particular, the widespread adoption of minimally invasive surgery procedures, which show a notably long and steep learning curve [136][117], led to the development of VR simulators for arthroscopic surgery [60], endoscopy, vascular interventions, orthopedics, ophthalmology, spinal surgery [18] and most recently neurosurgery [35]. Moreover, such simulators offer a cost-effective and efficient alternative to traditional training methods, e.g. cadaver training or animal models. Besides studies demonstrating content and construct validity, several studies showing the transfer of simulator-acquired procedural skills to the OR have been conducted [124] [51] [5].

Further companies providing high fidelity VR simulators for surgical training in particular MIS procedures are CAE Inc. (US), Simbionix Inc. (part of 3D Systems, US) and VirtaMed AG (Swiss) (see Fig. 10). These companies mainly produce endoscopy, laparoscopy simulators with haptic feedback and various training cases. CAE and Simbionix also offer cardiac/vascular simulators. The drawbacks of these simulators is according to [46] that the maintenance is cost-intensive and the most serious one is the teaching of bad habits [46] as dangerous behavior is not included in a summative assessment feedback.

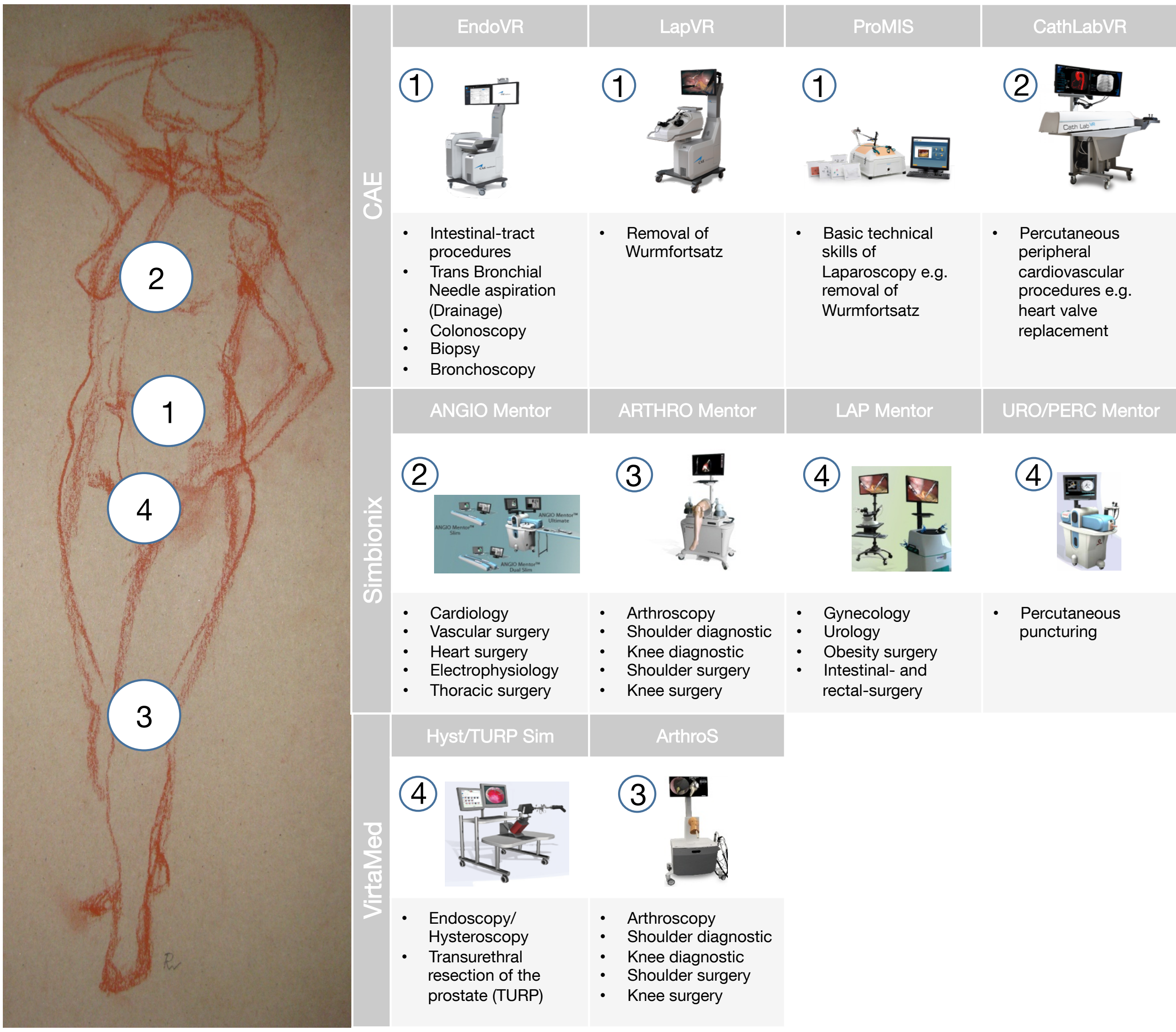


Fig. 10: Overview of VR simulators of CAE Inc.[a], Simbionix Ltd.[b] and VirtaMed AG related to operating site[c].

[a]Source: http://caehealthcare.com/eng/interventional-simulators accessed Apr/21/2016.
[b]Source: http://simbionix.com/simulators/ accessed Apr/21/2016.
[c]Source: http://www.virtamed.com/en/medical-training-simulators accessed Apr/21/2016.

Between 1986 and 2001 the Kinematic Simulation, Monitoring and Off-Line Programming Environment for Telerobotics (KISMET) has been developed at the Forschungszentrum Karlsruhe GmbH, Institut für Angewandte Informatik by the VR and Real-time Simulation Group. The VR simulator VEST System One (VSOne) was derived from KISMET in 2001. VSOne was designed with the objective to create a VR endoscopic surgery learning environment for a single user. It provides haptic feedback and includes realistic 3D anatomical models. Since 2005, developers, derived from the VOXEL-MAN research group, formed a separate entity within the University Medical Center Hamburg-Eppendorf, called the Medical simulation environment for multidisciplinary team assessment and training VOXEL-MAN Group. They created a VR simulator, named Tempo for training surgical access to complex and vulnerable structures of the middle ear in 2007. Furthermore they extended the VR simulator Tempo to endoscopic surgery in 2011. Both simulators (Tempo and Sinus) offer tactile feedback based on virtual models. Apart from the surgery simulators a dental simulator has been developed since 2009.

High fidelity/mannequins Standardized patients have been used to teach clinical skills. More specifically human patient simulators have been developed extensively and proven to be valid for cardiovascular examination and resuscitation training [119] [50]. Pioneered by David Gaba, computerized mannequin simulators (see Fig. 11) have been developed for training and performance assessment of anaesthetists [45] and have been in use for more than two decades. Today, mannequin simulators can be connected to medical ventilators and monitoring devices, physiologically respond to drug administration and show pathologic conditions [117] [43]. Several validation studies showing content and construct validity have been conducted [30], [29] and some studies regarding the transfer of skills to in vivo patient care have been described [20]. Since 2014 the training on those kind of simulators is mandatory for each anaesthetist in the US every ten years during his professional life. According to [46] the main drawbacks are that no standardized metrics exist for their usage, subjective assessment is conducted and moreover their usage is time-consuming.

Fig. 11: Computerized mannequin SimMan® 3G of Laerdal[a] with control unit[b].

[a]Source: `http://www.aimsi.org.au/site/laerdal+als+simulator+advanced+msc29_24.php` accessed Apr/21/2016.
[b]Source: `http://www.lmc.edu/faculty/communications/press-kit-may-school.htm` accessed Apr/21/2016.

High fidelity/full physics VR simulators Full physics VR simulators are the most realistic simulators of human patients. In real time they simulate the anatomical and physilogical behavior based on patient specific imaging data sets while using real medical instruments (see Fig. 12). New cases can be integrated however technical assistence is required.

Fig. 12: Vascular Intervention Simulation Trainer (VIST) of Mentice AB [a].

[a]Source: `http://www.mentice.com/vist-lab-with-vist-g5` accessed Apr/21/2016.

High fidelity/live tissue models Living tissues model e.g. pigs or sheep can be used for surgical training (see Fig. 13). They feature similiar behavior as a human patient as they breath and therefore move during the procedure. Additionally the instrument behavior is realistic e.g. a cautery instrument (instrument with electric charge) interacts with moist live tissue in proximity. However availability, ethical concerns and organisational overhead e.g. need of anaesthesist or veterinary and imaging devices e.g. mobile C-arms with technican are some limitations of the training on live tissue models.

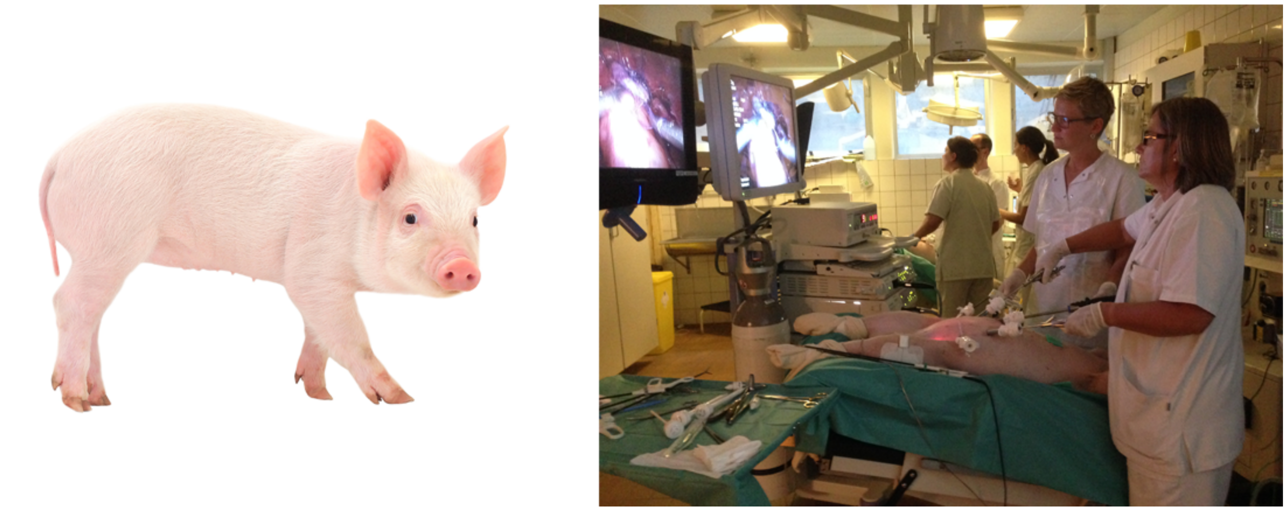


Fig. 13: From left to right: Pig [a], pig as operating model [b].

[a]Source: http://kids.nationalgeographic.com/animals/pig/ accessed Apr/21/2016.
[b]Source: http://www.miuc.dk/index.php?menu_id=13&content_id=59&lang=en accessed Apr/21/2016.

High fidelity/cadaver tissue models Training on cadavers is the most commonly used simulation in the medical domain (see Fig. 14). Medical instrument manufacturers e.g. Medtronic, Inc. or DepuySynthes Inc. use cadavers for training their sales staff or medical doctors, in particular when introducing a novel instrument into the market. In our days this kind of training is ethically questionable and similiar to training on live tissue requires organizational overhead. Other disadvantages are high costs, difference and variation to living anatomy and potential health hazard.

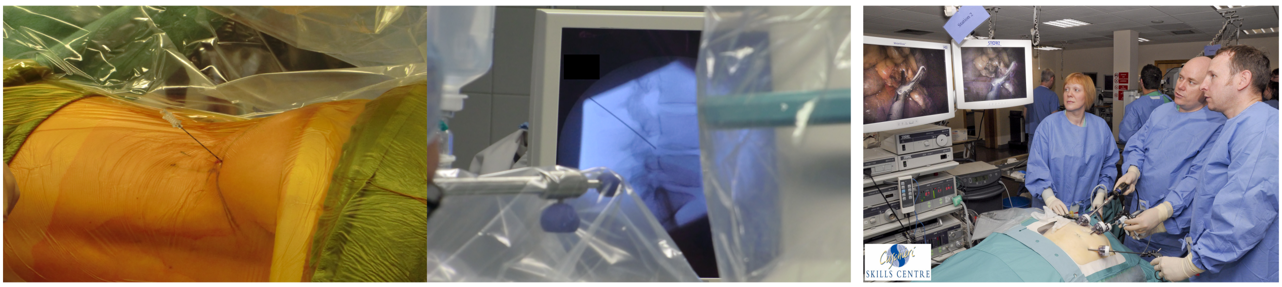


Fig. 14: From left to right: Needle inserted into the back of a cadaver; Fluoroscopic image illustrating needle inside cadaver's back; Cadaver course of instrument manufacturer [a].

[a]Source: http://www.eaes-eur.org/EAESCorporateWebsite/files/fa/fa20c3fe-0d49-4a5d-8a9a-488de163590f.jpg accessed Apr/21/2016.

High fidelity/live human (damaged) tissue models According to [46] in a far eastern country with a high population they use patients' tissue to conduct a training (see Fig. 15). Before an ischemic limb amputation is performed physicians learned and practiced their technical skill. However [46] states that this kind of training would probably not take foot in Western medicine.

Fig. 15: Live human (damaged) tissue model [a].

[a]Source: http://www.affcpodiatry.com/Photo%20Gallery.html accessed Apr/21/2016.

2.2.3. Steps 6-7: Real patients and widsom acquisition

Finally the trainee conducts procedures on patients under the supervision of a surgical mentor. Over time they gain more and more experience due to the exposure to various cases and crisis leading to surgical experience and proficiency.

2.3. The missing step 5c - Learning surgical skills in team-based training

Surgical skills are of a two-fold nature. First technical skill and procedural knowledge have to be acquired by the surgical trainee. The craftsmanship e.g. usage of instruments, procedural knowledge e.g. workflow and cognitive skills e.g. interpreting medical imaging are the learning goals. These are integrated in the PBP curriculum and can be achieved by using a variety of simulation models ranging from bench-top models e.g. suture pads, VR emulators/simulators, full-scale VR simulators, human cadavers to living animals e.g. pigs.
Second non-technical skill e.g. communication and decision making are learning objectives as they are essential for effective and efficient team work in the OR. A few decades ago the aviation industry started to intensively explore crew behavior in simulated crisis scenarios and created a basis for *crew resource management* (CRM) [43]. This approach was adopted later on by anaesthetist as *anesthesia crisis resource management* (ACRM) which used computerized mannequins to set up immersive scenarios for intradisciplinary team training [61] [45].
Those two facets of surgical skills form surgical excellence and are required to perform a successful operation [126]. Yet surgical excellence is merely associated with technical performance neglecting the importance of non-technical skill. Hence a simulated full-scale immersive OR theatre is proposed for conducting assessment and training of surgical skills within a multidisciplinary fully controlled environment. Within the PBP learning curriculum it would become step 5c (see Fig. 16).

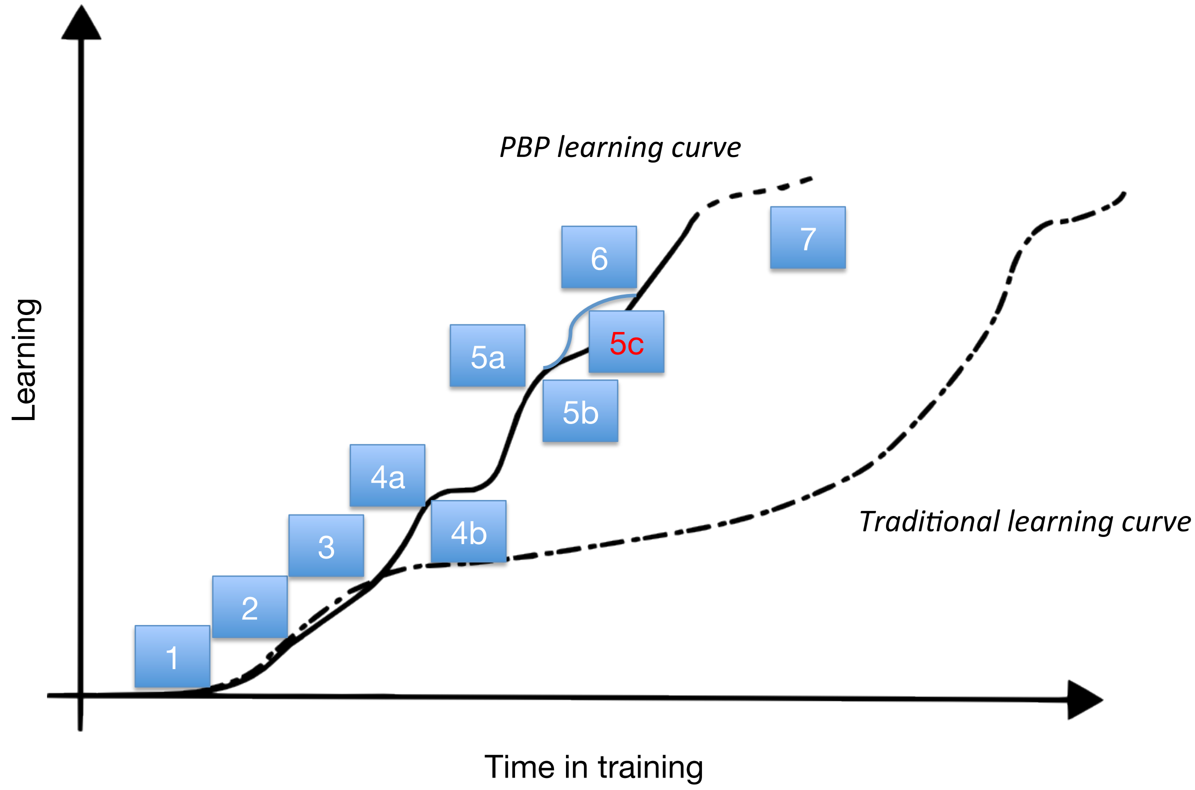


Fig. 16: PBP training extended with full simulated environment as step 5c.

2.3.1. Research on the missing step 5c: multidisciplinary team-based simulation environments

Today there still exist only a few examples of cross fertilization between both simulation approaches namely VR simulation and simulated patients (computerized mannequins) (see Section 2.3.) and only preliminary attempts to train surgeons and anaesthetists together [3], [95], [96], [101] Table 2.2. There are many concepts used exclusively in procedural or mannequin simulations that have the potential to effect a complementary cross fertilization in a joint effort. An example of such a concept is the *deliberate exposure to adverse events*. This concept was proposed by [46] as a solution to the common problem of missing experience of surgical students due to a limited exposure to actual patients experiencing low-frequency and high-risk situations [2]. The described assessment and training environments (see Table 2.2 consist at least of an anesthetic simulator (mannequin) most of the time in combination with synthetic models and offer deliberate exposure to adverse events. Only one study reported on the use of a VR simulator and a mannequin in simulated OR setting [100]. Paige et al. used in a pilot study with 10 participants a full-scale computerized mannequin and a VR cholecystectomy simulator (see Fig. 17). Three teams consisted of 3-4 team members. The crisis scenarios were embedded in the simulation program of the mannequin and forced the team members to manage the crisis. In particular the communication with the surgical resident who performed on the VR simulator was part of the training. A positive receptiveness of the participants was assessed by using questionnaires. However the 'integration' e.g. the correlation between physiological parameters of mannequin and VR simulator has not been truly made.

Fig. 17: A METI Human Patient Simulator (HPS) (left) combined with a VR cholecystectomy simulator of Simbionix Ltd. (image taken from Paige 2007).

Another example of Paige et al. combined a synthetic cholecystectomy model with a mannequin simulator into a mobile team training setup for crisis scenarios e.g. cardiac arrhythmias, malignant hyperthermia, anaphylactic allergic reaction to antibiotic and hypotension due to septic shock [101]. They essentially demonstrated face validity of their simulation setup. Moorthy et al. used a synthetic vein operation model placed over the right groin of a mannequin simulator to train surgeons in a team simulation setting. In one experiment, a hypoxia scenario was triggered by the trainer during the simulation session in order to assess the surgeons' awareness towards the anaesthetist and the patient [95]. In a second experiment, bleeding was started by the trainer at a standardized point in order to assess non-technical skills besides the ability to control the bleeding [96]. Both studies demonstrated the face validity of a novel full-team OR crisis simulation environment for technical and team skill training. Their studies could also establish construct validity of the simulation regarding technical ability of the surgeons, however they could not demonstrate it for non-technical skills. As a major limitation of their approach they see the lack of suitable high fidelity surgical simulators with a possible solution being VR simulators. [46] notes that "few, if any of the virtual reality simulations have the capacity for the trainer to control the introduction of an adverse event to the training scenario, although this is a common occurrence in anesthesia training" (see Table 2.2).

		Technology		
Author	Environment	Anesthesia	Surgical	Integration*
Helmreich [58]	Simulation center	Wilhelm Tell	Laparoscopic simulator	None
Sexton [122]	Simulation center	Wilhelm Tell	Laparascopic simulator	None
Aggarwal [3]	Simulation center	Mannequin	Synthetic model (Limbs & Things)	None
Moorthy [96]	Simulation center	SimMan (Laerdal)	Synthetic model (Limbs & Things)	Anatomic
Flanagan [37]	In situ	METI HPS	None	None
Kozmenko [73]	In situ	METI ECS with proprietary software	Simulab Torso Trainer (Simulab)	None
Paige [100]	Simulation center	METI HPS	Cholecystectomy model torso trainer (Simbionix)	None
Undre [135]	Simulation center	SimMan (Laerdal)	Saphenofemoral junction model (Limbs & Things)	Anatomic
Koutantji [72]	Simulation center	SimMan (Laerdal)	Saphenofemoral junction model (Limbs & Things)	Anatomic
Powers [107]	Simulation center	METI HPS	Synthetic model	None
Forsythe [41]	In situ	Unspecified combination of low and high-fidelity simulators		None
Paige [101]	In situ	METI ECS with proprietary software	Simulab Torso Trainer (Simulab)	None
O'Regan [99]	In situ	Basic CPR manikin	None	None
Volk [137]	In situ	SimMan (Laerdal)	None	None
Ziewacz [154]	Simulation center	SimMan 3G (Laerdal)	Tumor model and various Limbs & Things models	Anatomic
Stevens [129]	In situ	SimMan 3G (Laerdal)	Orpheus simulator (Ulco Technologies) and custom models	None
Lee [78]	Simulation center	SimMan 3G (Laerdal)	Synthetic model	None
Acero [1]	Simulation center	SimMan 3G (Laerdal)	None	None

Table 2.2: Adapted from Cumin 2012 and Tan 2014 highlighting the technology used in the studies which were conducted within the timeframe 1996-2012; excluded are the types of simulation "point-of-care" and "wet laboratories"; in situ means real OR environment; the term "integration" is used for the combination of mannequin and VR simulator.

Thus the missing link is a VR simulator which can be optimally integrated into a simulated OR theatre. The term 'optimally' is used in the sense that (i) the physiology of the VR simulator is coupled with the physiology of the mannequin e.g. cement leakage during a vertebroplasty provokes a lung embolism leading to increased heart beat of the simulated patient. And (ii) the hardware setup realism is not hindering immersion e.g. the VR simulator should be located at the position where you would normally expect the operating site of the surgical team. Additionally it should reflect the real operating site e.g. patient skin, appropriate drapping and real surgical instruments.
The community strongly agrees that the combination of both simulation technologies, mannequin and (VR) procedural simulator, would facilitate the integration of non-technical skills into the surgical curriculum and might even achieve the broadest potential of medical simulation: a full team training for all varieties of clinical teams and in particular among surgeons and anaesthetists [20], [3], [95], [96], [101], [2], [70], [121], [46]. This agreement is driven by the experience that the majority of errors in high reliability organizations such as aviation are due to non-technical errors [3].

Therefore the objective of the research work which is described within this thesis is to combine a computerized mannequin simulator with a novel VR surgical procedural simulator (see Fig. 1), as the key problems of simulated OR theatres with full mission scenarios for team-based assessment and training are:

Key problems of simulated OR theatres

- High fidelity VR simulators designed for team-based training does not exist referring to [23] [133]
- Little knowledge on correlating technical and non-technical performance indicators leading to differentiable factors which should be addressed for individual skill training (see Section 2.3.1.)

Research objectives

- The combination of a VR surgical procedural simulator and computerized mannequin is a novel training setup for medical assessment and training. Scientific evaluation and validation of this new simulation environment are conducted.
- Research and development of a surgical simulator reflecting complete workflows of surgical procedures, including intra-operative crisis scenarios which potentially result in adverse events. The capability to deliberately expose the trainee to realistic adverse events facilitates the research on robust error metrics for performance assessment in adverse or critical situations, as well as on strategies for augmentation and amplification of error perception and mediation of error recovery strategies, e.g. in deliberate practice or procedure rehearsal scenarios.
- Establishment of structured assessments of surgeons' and surgical team's technical and non-technical skills during simulation-based training that enhance impact on the acquisition, application, and retention of teamwork skills in healthcare.

According to the framework of [44] the developed simulation can be applied in health care in the following 11 dimensions (see Table 2.3). Furthermore the simulation environment comprises the surgeon-patient-machine triangle in Image-Guided Interventions (IGI) illustrated in [64] and focuses on the assessment level 3 though considering the study conditions "simulated clinical scenario, laboratory" of level 2.

Dimension 1: The purpose and aims of the simulation activity

		+	+	+
Education	Training	Performance assessment	Clinical rehearsal	Research (Human factors)

Dimension 2: The unit of participation in the simulation

+	+	+		
Individual	Crew	Team	Work unit	Organisation

Dimension 3: The experience level of simulation participants

		+	+	+
School Primary Secondary	College; university	Initial professional education	Residency or on-the-job training	Continuing education and training

Dimension 4: The health care domain in which the simulation is applied

			+	+
Imaging (Radiology Pathology)	Primary care; psychiatry	In-hospital ward based (Medicine/ Paediatrics)	Procedural (Surgery, OB/GYN)	Dynamic high hazard (OR, ICU, ED)

Dimension 5: The health care discipline of personnel participation in the simulation

			+		
Aids; clerks	Allied health; technicians	Nurses (Including advanced practice nurses)	Physicians	Managers; executives; trustees	Regulators; legislators

Dimension 6: The type of knowledge, skill, attitudes, or behaviour addressed in simulation

	+	+	+
Conceptual understanding *Knows*	Technical skills *Knows how* *Shows how* *Does*	Decision making skills *Meta-cognition* *Static Dynamic*	Attitudes and behaviours *Team work* *Professionalism*

Dimension 7: The age of the patient being simulated

			+	+
Neonates	Infants	Children; teens	Adults	Elderly

Dimension 8: The technology applicable or required for simulations

				+
Verbal Role playing	Standardised patients (Actor)	Part-task trainer *Physical; virtual reality*	Computer patient *Computer screen; screen based "virtual world"*	Electronic patient *Replica of clinical site; mannequin based; full virtual reality*

Dimension 9: The site of simulation participation

			+	
Home or office *Multimedia screen-only simulations*	School or library *Multimedia screen-only simulations*	Dedicated laboratory *Physical part-task trainers Virtual reality part-task trainers*	Replica clinical environment *Replica clinical sites Patient simulation systems Full video capture*	Actual work unit *"Insitu" simulation Mobile simulation*

Dimension 10: The extent of direct participation in simulation

			+	
Remote viewing only *No interaction*	Remote viewing with verbal interaction *Simulation based M & M conference*	Remote viewing with hands-on interaction *Remote haptic surgica trainer*	Direct on-site hands-on participation	Immersive participation

Dimension 11: The feedback method accompanying simulation

				+
None	Automatic critique by simulator *Real time delayed*	Instructor critique of records of prior simulation sessions	Real time critique *Pause/restart Real time mentoring*	Video based post-hoc debriefing *Individual/group*

Table 2.3: The 11 dimensions of simulation applications. In the figure the general application of a simulated OR is depicted (Gaba 2004).

2.4. The design, development and validation process of a multidisciplinary team-based simulation environment

To ensure a valuable simulation environment an iterative development approach Table 2.5 should be used to satisfy the miscellaneous requirements of different stakeholders. [121] recommends to involve specialists of various disciplines ranging from surgical experts (teachers), residents (learners), educationalists e.g. simulation experts (teaching the teachers), designers, engineers and psychologists.

Process step	Choice of procedure scenarios (see Section 3.3. and 3.4.)	Task analysis (see Section 3.5.)	Design and validation of assessment system (see Sections 3.6.1. and 4.3.)
Purpose/Outcome	Appropriate procedure scenarios with crisis	Assessment system/framework; Assessment construct e.g. surgical skill; Requirements for simulation	Accurate reflection and good predictor of surgical skill; Technology for assessing objectively performance; *Assessment of/for learning*
Methods/tools	Literature search strategy	Cognitive walkthrough, semi-structured interview; video-based analysis; Intraoperative documentation (e.g. observation)	Training environment with performance metrics
Sources	Books, publications, videos, medical expert interviews	Books, publications, videos, domain expert feedback and advice	Feedback/review sessions with domain experts

Table 2.5: Guideline: Iterative development process.

The following questions should be asked during the development:

Key questions and aspects

- Which procedure should be considered?
- Is appropriate simulation technology available for simulating the procedure?
- Can available simulation technology be optimally integrated into the team-based setup?
- Are validated performance metrics available for assessing technical and non-technical skills?

Additionally a few aspects need to be considered such as the fidelity and the choice of an appropriate scenario.

2.4.1. Fidelity of simulation equipment

Fidelity of medical simulation environments refers to the degree to which it reproduces the state and behavior of the real world [46]. The fidelity of the simulation materials or technology is highly dependent on the purpose of the simulation and is a barrier to conducting multidisciplinary team training in a simulated OR setting [23][133]. The key factors according to [133] are "the unit of participation, the experience level of simulation participants, the type of knowledge, skills and attitudes addressed in simulation and furthermore participants may perceive the simulation differently depending on their personal experience". Referring to [46] apart from the presentation and computer graphics the behavior of the instruments used during the simulation and the behavior of the tissue should be as realistic as possible. In order to achieve the greatest return on investment it is key to assess during the design and development phase how much fidelity is required.

2.4.2. Scenarios

The choice of the scenarios is key for enabling the simulation participants to immerse into the simulation. The scenarios should reflect crises in such a way that the participants believe that they are in a realistic situation and behave as they would in real life [133].

2.4.3. Validation study recommendations

According to [123] thoughts have to be spent on the following essential aspects.

Template and essentials for a study

- Deliberate selection of VR task(s).
- Selection of task difficulty levels.
- Selection of duration of training.
- Definition of reasonable performance objectives.
- Subjects (experience, norming e.g. via visuospatial reasoning test, multicenter).
- Task.
- Performance metric for task.
- Assessment tool and method (technical setup or at least two observers using checklists with rating scales => objective quantitative data).
- Hyphothesis.
- Groups (training vs. no training).
- Category of training.
- Environment (well-known conditions).
- Analysis.

The proposed simulation environment for team-based assessement and training has to be validated and requires appropriate methods to measure the surgical performance of each team member and the team performance. The next part of this thesis is targeted to provide the necessary background for the validation of simulation technology and further on illustrates currently used methods for assessing surgical technical and non-technical performance. These described methods were later on used within the studies to prove the validity of the simulation environment and additionally provide the first steps towards appropriate performance metrics for assessment of technical and non-technical performance.

2.5. Validation

In the case of assessment, validity refers to the degree to which a measurement instrument e.g. a simulated OR truly measures what it is supposed to measure. It is concerned with whether the right things are being assessed, in the right way, and with a positive influence of learning. Within this thesis the various validities referred to, consist of those mentioned in [9] and [46]:

Face validity Face validity can be described from the perspective of an interested lay observer. If he or she feels that the right things are being assessed in the right way, then the assessment has good face validity. Face validity is tested during the early development stages of a test construction.

Content validity An assessment has content validity if it contains the steps and reflects the abilities (knowledge, skills or behaviours) used in a procedure which it should measure.

Construct validity The extent to which the assessment and the individual components of the assessment test the professional constructs on which they are based. For instance, an assessment has construct validity if senior trainees achieve higher scores than junior trainees.

Criterion validity Agreement with other assessments intended to measure the same construct. For example, method A and method B both assess technical surgical skills and, despite being different assessment methods that are completed separately, could be expected to agree as they measure the same construct of surgical performance.

Apart from those validities previously described, which are proven within the studies in Chapter 3. and Chapter 4. the following validities are commonly used in discussions and publications:

Predictive validity This refers to the degree to which an assessment predicts expected outcomes. For example, a measure of attitudes (behaviour) towards preventive care should correlate significantly with preventive care behaviours.

Consequential validity Educational impact: This is an important aspect of the validity of assessment. It refers to the effect that an assessment has on learning, and in particular on what trainees learn and how they learn it. For example, they might omit certain aspects of a syllabus because they do not expect to be assessed on them, or they might commit large bodies of factual knowledge to memory without really understanding them in order to pass a test of factual recall, and then forget them soon afterwards. Both these behaviours would indicate that the assessment has poor educational impact because both lead to poor learning behaviours.

Concurrent validity If the pre-existing "gold standard" assessment tool can be replaced by a new assessment tool while both tools are measuring the same construct.

Discriminate validity It reflects the extent to which the scores generated by the assessment toll actually correlate with factors with which they should e.g. the assessment tool can discriminate the ability levels within a group with similar experience.

2.6. Assessment

The following section will give you an illustration of key concepts of assessment theory in order to provide a profound background for simulation-based assessment.

The term assessment stands for the process of measuring a trainee's knowledge, skills, judgement or professional behaviour against defined standards. It has to be as objective and reproducible as possible. The test should be reliable therefore it should produce the same or similar scores in two occasions or by two assessors. Furthermore the test should be valid which means that it measures what it sets out to measure and its educational impact. The focus of the thesis is the assessment of surgical skills. Hence the surgeons' skills are compared to a reference. There exist two ways for referencing assessment [9]:

- Criterion-referenced refers to an absolute standard, i.e. the trainee's performance against a benchmark. Such a benchmark might be the ability to perform a procedure competently without help from the assessor.
- Norm-referenced ranks a trainee's performance against all the others in the same cohort, i.e. satisfactory for that level of training. Norm-referenced assessments are inherently more difficult to determine and, whenever possible, should not be used.

2.6.1. Purpose of assessment

The main drivers for teaching and learning is assessment. In order to define the appropriate assessment methods and tools to apply the right metrics a global view on the purpose and level of assessment is given [9]. According to [22] an assessment has to be based on the purpose of the assessment and the following aspects should be considered while desiging an assessment:

- Choice of assessment methods.
- Selection of assessment tools.
- Way in which the above are combined.
- Number of assessments.
- Timing of assessments.
- Way in which outcomes are used to make decisions regarding progression or certification.

2.6.2. Levels of assessment

Models for describing the levels of assessment were proposed by [93] and [111].

Miller's model (see Fig. 18) provides the basis for guiding curricular design and selection of assessment methods to suitable levels of assessments. It consists of four levels of assessment which are related to surgical skill assessment methods.

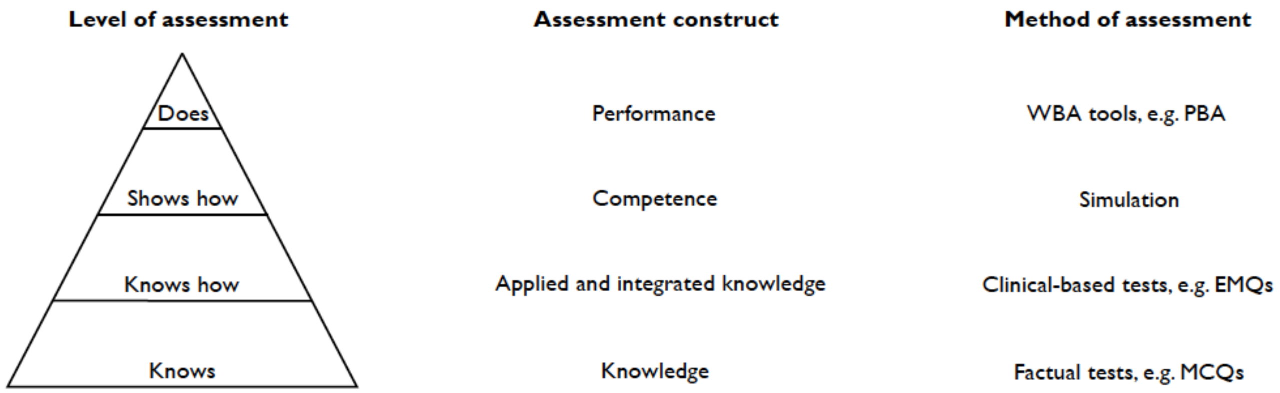


Fig. 18: Miller's pyramid linked to methods of assessment. EMQs: extended matching questions; MCQs: multiple choice questions (Beard 2011).

The Cambridge model of performance by [111] is derived from Miller's pyramid (see Fig. 19). It models performance as a window to competence. Each performance of a surgeon has to be seen 'case-related' (e.g. complexity or type of procedure). Furthermore the model integrates additional factors 'individual-related influences' (e.g. physical and mental health) which have an impact on the day-to-day performance of doctors and 'system-related influences' (e.g. time pressures).

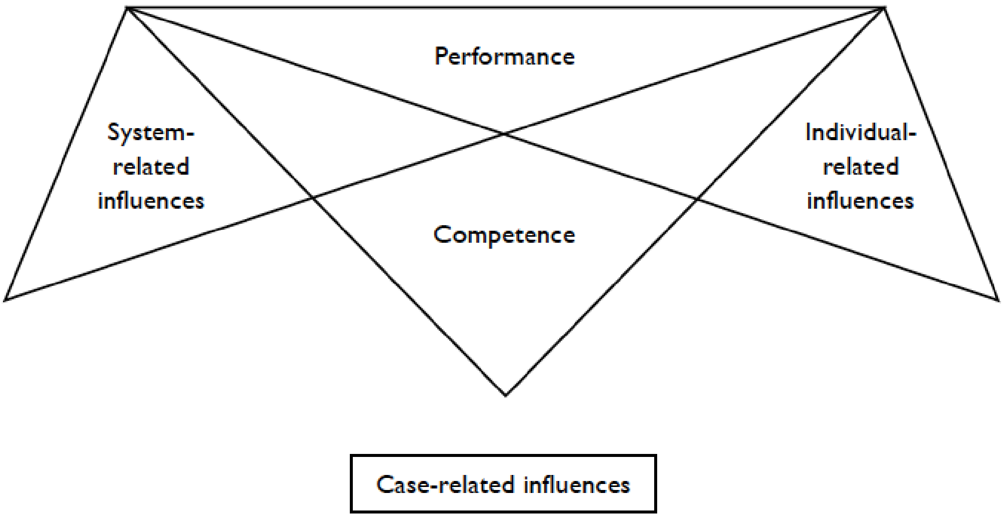


Fig. 19: The Cambridge model of performance (Beard 2011).

A more comprehensive account of the skill acquisition process is the skill model of Dreyfus (see Fig. 20 and Table 2.7) (1986/2000) [69] [46]. It depicts the process of building up expertise from novice, advanced beginner to competence and to proficiency and finally expert. All these different levels are related to performance characteristics (see Table 2.7). The progression of the trainee/learner is based on the talent, how similiar the new tasks are to the performance characteristics and the skills learned in previously performed tasks and the skill of the mentor.

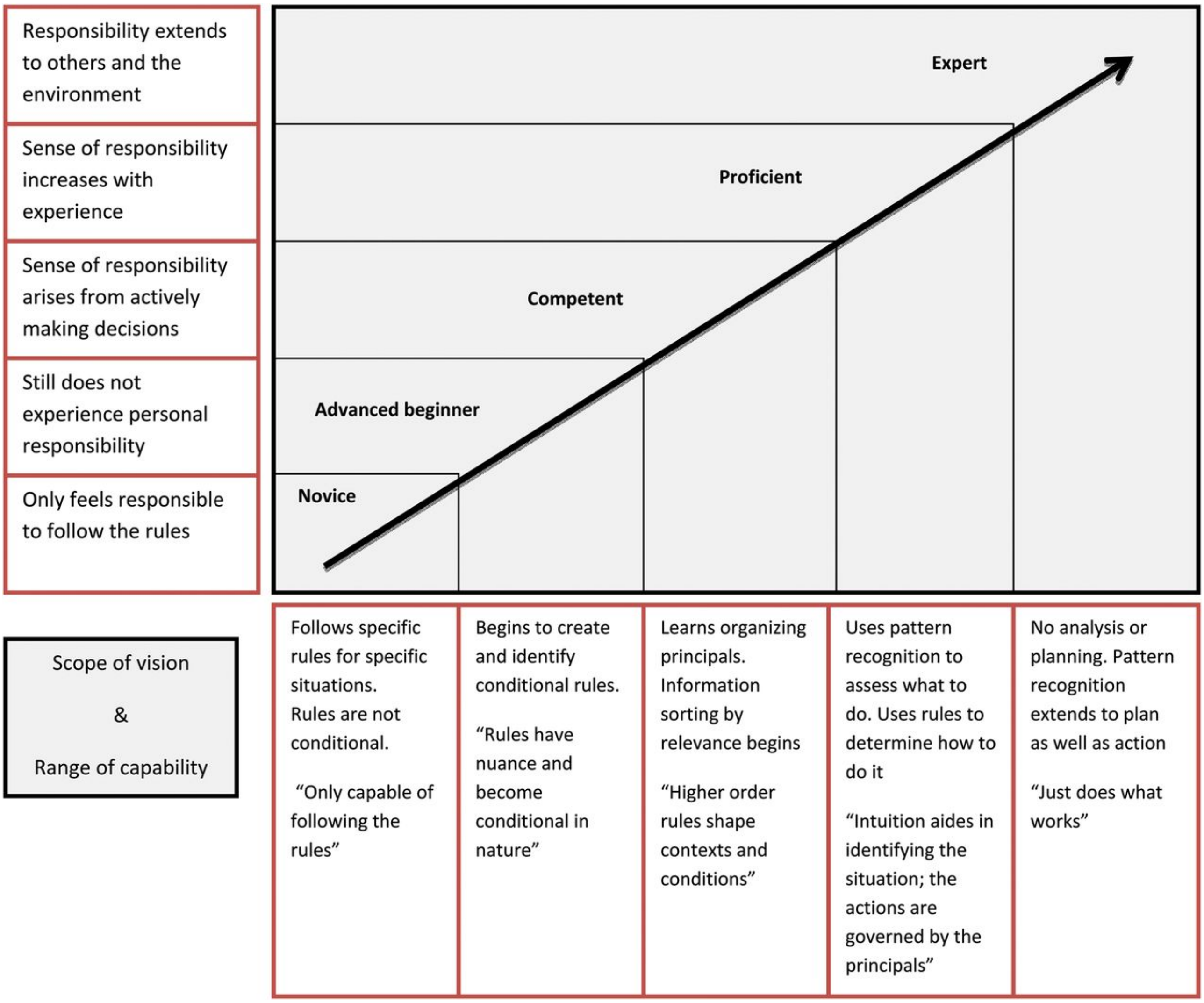


Fig. 20: The Dreyfus skill model (Kirkpatrick 2012).

Stage	Performance characteristics
Expert	• Source of knowledge and information of others • Continually looks for better methods • Work primarily from intuition • Being forced to follow rules degrades performance
Proficient	• Seeks to understand larger context • Frustrated by oversimplification • Can self-correct performance • Can learn from experience of others
Competent	• Can troubleshoot problems problems on his/her own • Seeks out expert user advice • Develops conceptual models
Advanced beginner	• Starts trying tasks on his/her own • Has difficulty troubleshooting • Begins to formulate principles, but without holistic understanding
Novice	• Has little or no previous experience • Is vulnerable to confusion • Does not know to respond to mistakes • Needs rules to function

Table 2.7: Characteristics of each stage of the Dreyfus skill development model from Gallagher 2012.

Assuming the skill model of Dreyfus as basis the interesting questions are: "to which level the learner should advance to and further on how do you benchmark it?" First, the trainee should reach the proficiency level related to a previously defined benchmark by experts. Meaning that the person is acting responsibly without the need of supervision and additional builds-up confidence/expertise based on experiencing problems/errors he can solve on his own. The defined benchmark derived from several trials of experts on the simulator (the number of trials depend on how fast the learning curve flattens out) which are recorded and assessed grounding on a validated assessment construct. This means that ultimately a performance score is derived to which trainees can be compared to.

2.6.3. Technical skill assessment

Various methods exist to assess the technical performance of surgeons e.g. OSATS, ICSAD and MISTELS (see Appendix C2 C2.), however up to date there does not exist a generic metric for MIS procedure performance characteristics. It seems that it has to be developed procedure-specific based on a task analysis [46]. Within the task analysis the behaviors are identified which should be measured (see Chapter 3. and Chapter 4.).

2.6.4. Non-technical skill assessment

SURG - Task Load Index (SURG-TLX) Within the second study which is described in detail in Chapter 4. the SURG-TLX is applied. [145] developed and validated the SURG-Task Load Index (SURG-TLX). Essentially it is a combination of the NASA-TLX which is a well validated instrument [67] and the Driving Activity Load Index (DALI) [104] considering the key intraoperative stressors of [143] for defining the dimensions approximating the demands the surgical operator faces. In total eight surgeons were consulted to give feedback on the proposed dimensions.

- Mental demands: How mentally fatiguing was the procedure?
- Physical demands: How physically fatiguing was the procedure?
- Temporal demands: How hurried or rushed was the pace of the procedure?
- Task complexity: How complex was the procedure?
- Situational stress: How anxious did you feel while performing the procedure?
- Distractions: How distracting was the operating environment?

In order to validate the SURG-TLX 30 novices performed a peg transfer task (see Fig. 21) which is a validated Fundamentals of Laparascopic Surgery (FLS) task [125]. The study showed that the SURG-TLX is sensitive to a several different surgical stressors including increased complexity, physical fatigue, multitasking and time pressure.

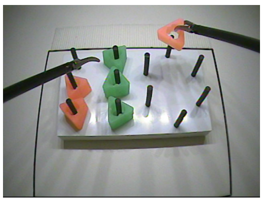


Fig. 21: Each object has to be transferred from one side to the opposite side of the pegboard. To perform this task, each object is grasped with the non-dominant hand, and then the object is transferred mid-air to the dominant hand. Afterwards the object is placed on a peg on the opposite side of the pegboard [a].

[a]Source: http://www.flsprogram.org/wp-content/uploads/2014/03/Revised-Manual-Skills-Guidelines-February-2014.pdf accessed Apr/21/2016.

3. Development and Procedural Evaluation of Immersive Medical Simulation Environments

This section is adapted from 'Development and procedural evaluation of immersive medical simulation environments' [148] and 'Task and Crisis Analysis during Surgical Training' [149] and it shares experiences in designing a complete VR simulator prototype for an immersive simulated OR theatre. The three conditions outlined in section 3.2. are accounted for through the following key research contributions:

- The combination of VR surgical procedural simulator and computerized mannequin in designing novel training setups for medical education.
- Based on a user-study, the quantitative evaluation through surgical workflow and crisis simulation for proving face validity of immersive medical training environments.

3.1. Simulator realism

To offer the same conditions to medical participants as if they were truly immersed in a live surgery, a simulator should be integrated with the same medical instrumentation used by surgeons. Further, the learning environment should address a broad spectrum of human sensory channels such as tactile, auditory and visual channels in real-time with the help of haptic feedback, CT imaging device simulation, and physiological monitoring.

3.2. Three conditions for an effective medical simulation learning environment

Many authors agree that the combination of mannequin technology and VR procedural simulators would facilitate the integration of non-technical skills into the surgical curriculum and might even achieve the largest potential of medical simulation: team assessment and training for all varieties of medical teams and in particular surgeons and anesthetists [75], [121]. To date there still exist only a few examples of cross fertilization of the above areas in team training and notably none uses a high fidelity mannequin simulator in combination with a VR simulator. This is probably due to VR simulators lacking trainer controllable interfaces important in CRM.

Conditions

1. Few, if any of the virtual reality simulations, have the capacity for the trainer to control the introduction of an adverse event to the training scenario, although this is a common occurrence in anesthesia training [75].
2. Second, many failed surgeries are directly linked to the surgeon's performance. The errors made can be distinguished into: (i) latent conditions which are inherent within the health care system e.g. time pressure, fatigue or unworkable procedures, and (ii) active failures which are of different type e.g. procedural violation, slips, and lapses. Thus, both surgeon and operating team should have situation awareness and experience with handling critical events which can endanger the patient. "The introduction of critical events into medical simulation learning environments helps to diminish the impact of disruptive unexpected events on the trainees' procedural skills. This enables the trainees to handle unfamiliar and unpredictable events" [121].
3. Third, there is ongoing discussion about the realism of simulators. For effective medical training the immersion into the environment is required [114]. For the setup of medical learning environments the utilization of real medical equipment is necessary. The learning environment should address a broad spectrum of human sensory channels such as tactile, auditory and visual channels in real-time.

3.3. Choice of a suitable procedure

We concentrate on vertebroplasty (see Fig. 24), a percutaneous image-guided minimally invasive surgery performed within orthopedic, trauma and radiology surgery rooms worldwide. Today percutaneous vertebroplasty is an assorted method that treats all types of vertebral fractures [74]. The main cause of vertebral compression fractures is osteoporosis, one of the most challenging diseases of the 21st century. Osteoporosis is a disease characterized by low bone density and mass, leading to deterioration of bone tissue, increased bone fragility, and risk of fracture [138]. Patients with osteoporosis have 23% higher rate of mortality compared to corresponding normal population [32], imposing a social and an economic concern. Osteoporosis affects 12% of population within the age range of 50-79 in Europe, and approximately 1.4 million Canadians [102] and 10 million Americans [82]. Every year about 1.4 million new vertebral compression fractures due to osteoporosis occur worldwide. The objective of vertebroplasty is to inject polymethylmethacrylate (PMMA) bone cement, under radiological image-guidance, into the collapsed vertebral body to stabilize it. However, the complication rate is markedly high and clinical adverse effects can be devastating if not treated immediately [106]. Intensive and accurate communication especially between surgeon and anesthetist is very important during the procedure to avoid such problems [42]. As such, we designed a simulation environment for the training of vertebroplasty procedures.

3.4. Adverse events and crisis simulation

The occurrence of adverse event(s) is crucial since understanding the impact of risk or danger on clinical judgement and skill is a vital element in becoming experienced [70]. During percutaneous vertebroplasty the most common complication is cement extravasation, i.e. cement leakage. When a leakage is not recognized during the procedure, a pulmonary embolism may develop as more PMMA is injected and increasingly migrates into the venous system. Pulmonary cement embolism is reported to occur in approximately 2-26% of procedures [139].

A reason for a surgeon's failure to recognize cement leakage is the lack of monitoring cement flow in caudal and cranial directions during (CT) guidance. As a result, an anesthesiologist aware of the procedure-related risks is present during surgery and can interpret clinical signs of a pulmonary embolism (i.e. sudden oxygen desaturation) and communicate it to the surgeon [42].

To ensure authenticity, our approach includes complete workflows of complex procedures with occurrence of adverse events, since understanding the impact of risk or danger on clinical judgment and skill is a crucial element in becoming an expert [70]. During percutaneous vertebroplasty, the most common complication is cement extravasation, i.e. cement leakage into (i) paravertebral soft tissue or the needle tract causing pain, (ii) intervertebral disc space possibly leading to new fractures of adjacent vertebrae (see Fig. 22), (iii) epidural or paravertebral veins potentially causing pulmonary embolism (see Fig. 23), and (iv) intervertebral foramen or the spinal canal which can lead to neuralgia or spinal cord compression respectively [28], [48].

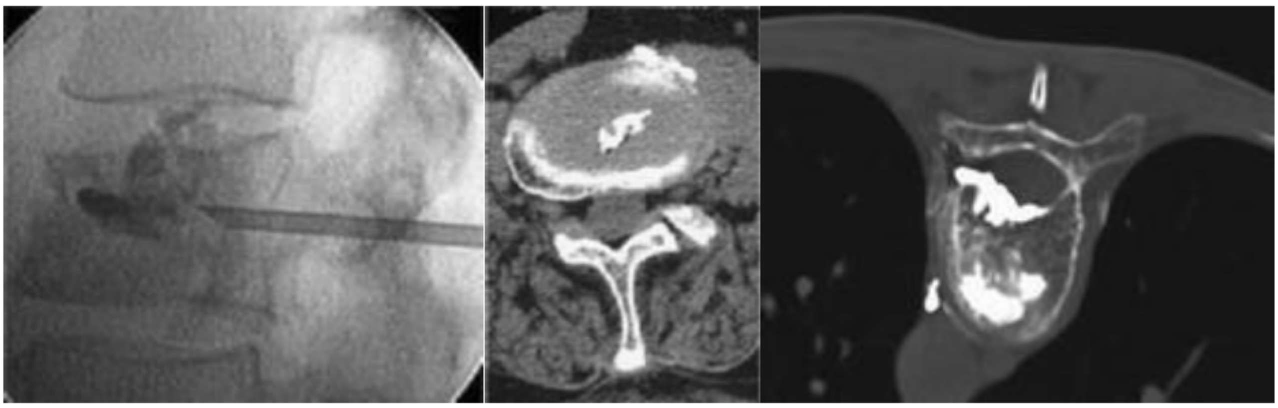


Fig. 22: (a) Fluoroscopy and (b) CT scan show cement leaks towards the intervertebral disc. (c) CT scan shows epidural space cement leak. From Gangi 2003.

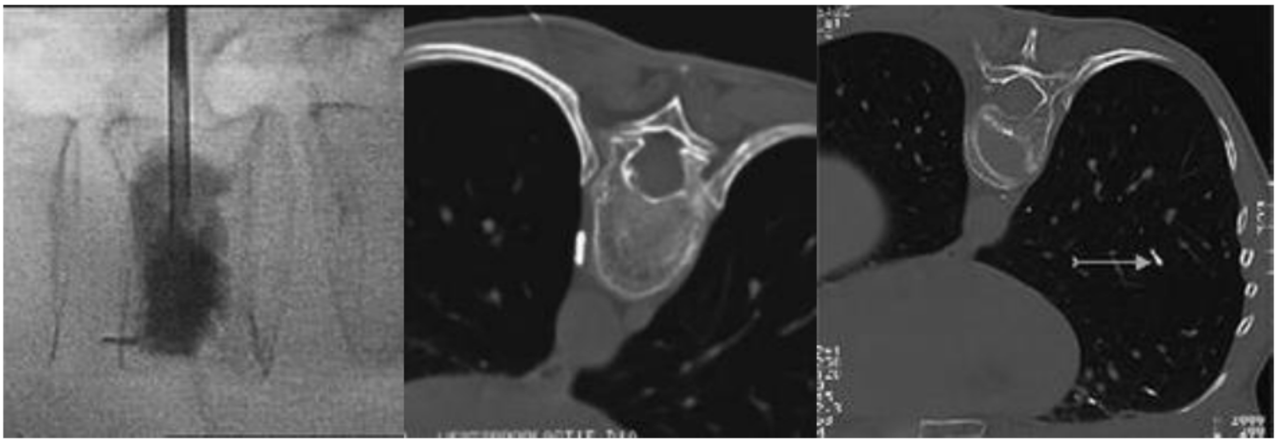


Fig. 23: (a) Fluoroscopy and (b) CT scan show venous leaks. (c) CT scan shows pulmonary cement embolism. From Gangi 2003.

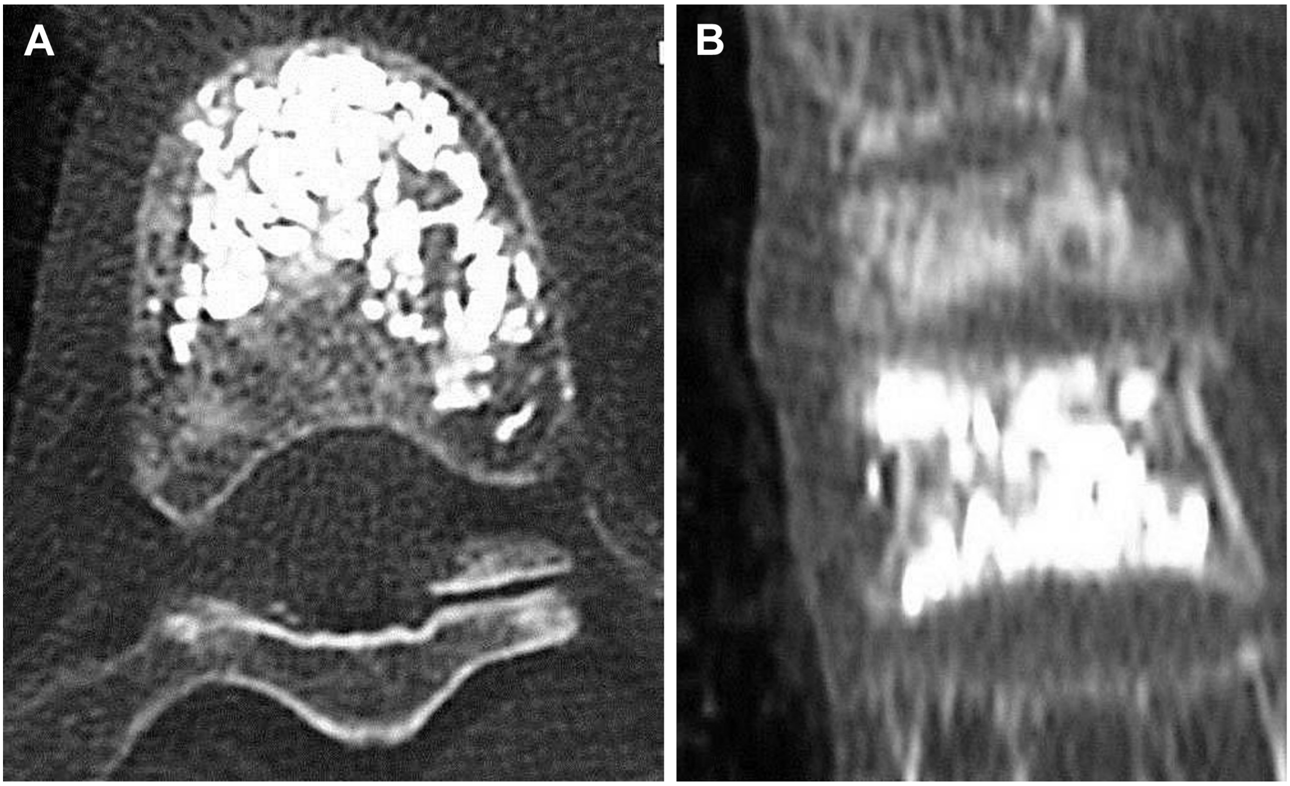


Fig. 24: Vertebroplasty is the injection of cement into fractured vertebrae. An (A) axial and a (B) sagittal view of a Computed Tomography (CT) scan shows no cement leakage outside of anatomy. [1]

When a venous leakage is not recognized early during the procedure, a pulmonary embolism may occur as more PMMA is injected and increasingly migrates into the venous system [42]. A reason for failure to recognize cement leakage on the surgical side is the lack of monitoring the cement in caudal and cranial directions. An anesthetist aware of the procedure related risks, however, can interpret clinical signs of a pulmonary embolism, e.g. sudden oxygen desaturation, arrhythmia or hypotension - which may in elderly osteoporotic patients be easily misinterpreted to a supposed cardiopulmonary comorbidity. In case of a suspected pulmonary embolism, the surgeon has to be informed to stop the injection immediately. Further examination and management of the situation (e.g. haemodynamic support, treatment with anticoagulants, or surgical intervention) have to be

[1]Source: `http://www.clinicaldensitometry.com/article/S1094-6950(15)00177-8/fulltext` accessed Apr/21/2016.

decided by the medical team [42].

3.5. The VR surgical procedural simulator for vertebroplasty

Our setup consists of a haptic device for instrument interaction (Fig. 25-1), a pad into which the instruments can be inserted (Fig. 25-2), a CT scanner mock-up including a positioning laser (Fig. 25-3), a foot switch triggering CT image acquisition (Fig. 25-4) and a monitor showing acquired CT images (Fig. 25-7). A computerized mannequin simulator is placed onto the operating room (OR) table (Fig. 25-5), the pad is fixed on the mannequin using a tension belt and the haptic device is attached to the table using a standard clamp. The computerized mannequin simulator is connected to the diagnostic devices (Fig. 25-8) and finally draped. Real surgical instruments (Fig. 25-6) can be attached to and detached from the haptic device using a clipping mechanism. CT imaging data is used to generate haptic feedback delivered to the instrument and visualize the patient's anatomy in combination with the simulated instrument on the CT monitor. The pad, essentially a box covered with synthetic skin, acts as housing for the instruments to avoid damage to the mannequin.

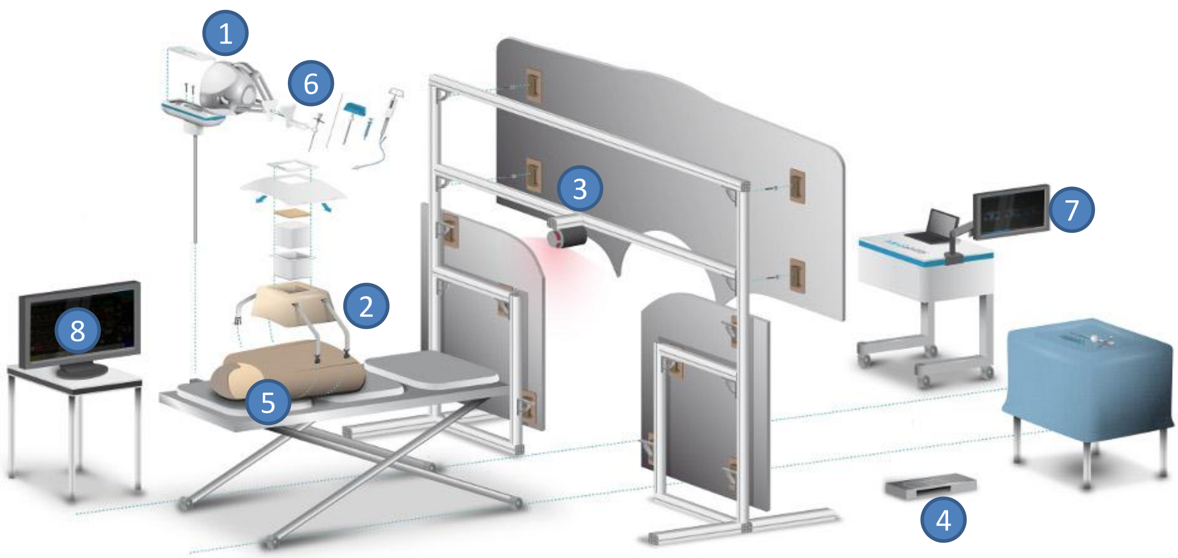


Fig. 25: Surgical training environment for vertebroplasty.

3.5.1. Surgical workflow steps and crisis simulation

The procedural steps were extracted from live surgery video recordings and literature [110] in conjunction with the feedback from expert surgeons. Through these surgical workflow steps the aim of our simulator is to realistically represent all subtasks of vertebroplasty up to cement injection and successful vertebral stabilization. Fig. 26 describes the tasks, instruments and learning objectives within three surgical workflow steps.

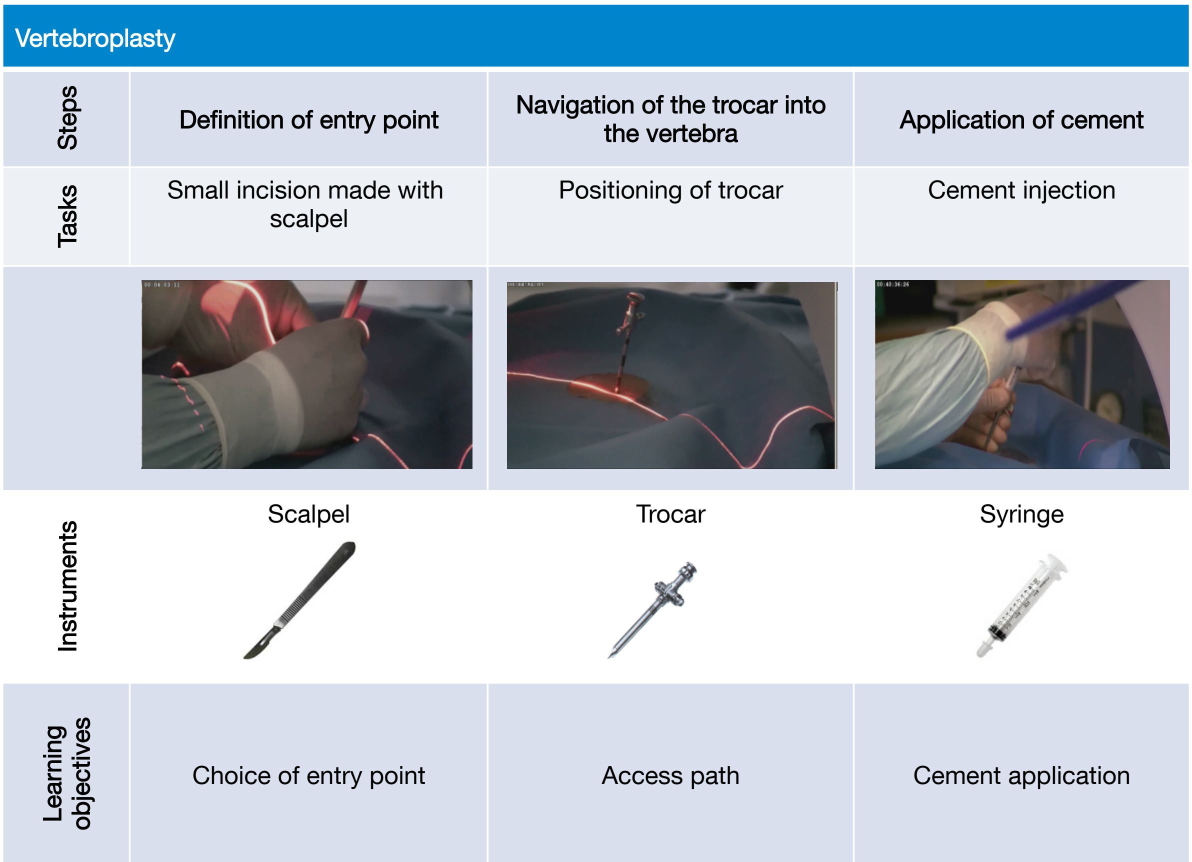


Fig. 26: Surgical workflow with corresponding instrumentation and learning objective.

Through a skin incision, the surgeon introduces a trocar into the virtual patient's body and advances it further through the pedicle into the vertebral body using CT guidance. Feedback generated by the haptic device gives the surgeon tactile information on the anatomy in contact with the instrument. Bone structures are discernible and clearly distinguishable from soft-tissue. When the desired position is obtained with the trocar inside the vertebral body, the surgeon injects bone cement using a syringe. A cement model is used to discern the amount injected and it is consequently augmented on the CT slice images. Crisis simulation: an 'unexpected event' is induced in terms of a cement extravasation into a perivertebral vein causing a lung embolism.
The aim here is to provoke communication between anesthesiologist and surgeon to relay proper response for this adverse event. For example, the surgeon is supposed to learn to better discern cement leakage in the CT image, before the pulmonary embolism occurs.

3.5.2. Technical details on instrumentation, haptic feedback and CT simulation

Instrumentation The instrument interface consists of a haptic device with a custom-made instrument connector and a pad, representing the patient's body, into which the instruments are inserted. The haptic device used is a Novint Falcon (Novint Technologies Inc., Albuquerque, NM, USA). It is a translation-only 3DOF variant of the delta-robot design which has the advantage of increased actuation stiffness [87]. The implications on the haptic feedback and the force reversal due to the fulcrum at the entry point are discussed in [84]. The end-effector of the Novint Falcon is detachable and can be replaced with custom attachments. Using rapid prototyping technology, we have developed an end-effector to which surgical instruments can be attached. The instruments are equipped with a plastic ball which is clipped to the end-effector socket in a ball-joint manner. To determine the amount of cement injected, we have developed a level-gauge model consisting of a level gauge with a USB interface installed in a syringe barrel (see Fig. 27). The cement injection syringe, filled with white colored water, is connected to it via a T-connector and standard syringe tubing. This T-connector makes it possible to attach the syringe to the trocar, creating the impression of injecting the cement into it, while in fact the liquid is channeled away into the measuring device.

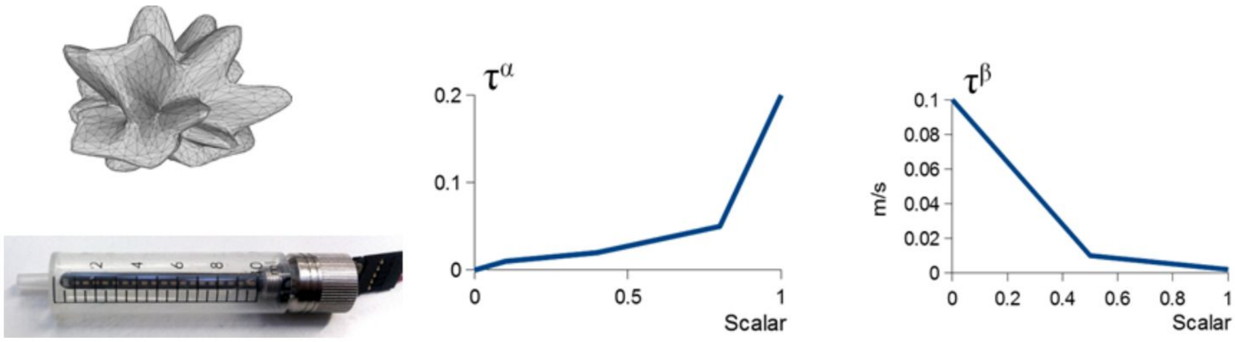


Fig. 27: From left to right: (Bottom) level gauge with USB connector. (Top) cement model; Transfer functions τ^α and τ^β.

Haptic simulation We use an approach similar to [109] to generate haptic feedback from CT imaging data. Specifically, two haptic primitives [84] described there, are used to generate the haptic feedback. The trocar path is modeled by a line primitive restraining the trocar from deviating. Trocar progression is controlled using a plane primitive exerting resistance as the trocar is advanced through the tissue. Instead of defining the strength of the line primitive as a function of depth, we use the radioopacity of the penetrated tissue as an influencing factor. It is defined by accumulating samples along the instrument path, from the entry point to the tip of the instrument, that are interpreted using a transfer function (see Fig. 27-middle). A second transfer function (see Fig. 27-right) is used to map strength to a maximum penetration speed which is enforced by the plane primitive as described in [109]. The transfer functions were experimentally defined with expert surgeons. During this process, it became apparent that the bone corticalis could not be clearly perceived by the user. Therefore, we added a proxy-based surface haptics rendering method [118] reflecting the distinct shape of the cortical bone using a surface mesh derived from a segmentation of the vertebrae. This has a high simulated stiffness and we simulate bone penetration by dropping the resistance if a particular force threshold is exceeded. The resistance also drops in reality as the bone corticalis is penetrated and the trocar advances into the brittle trabecular bone structures Fig. 28.

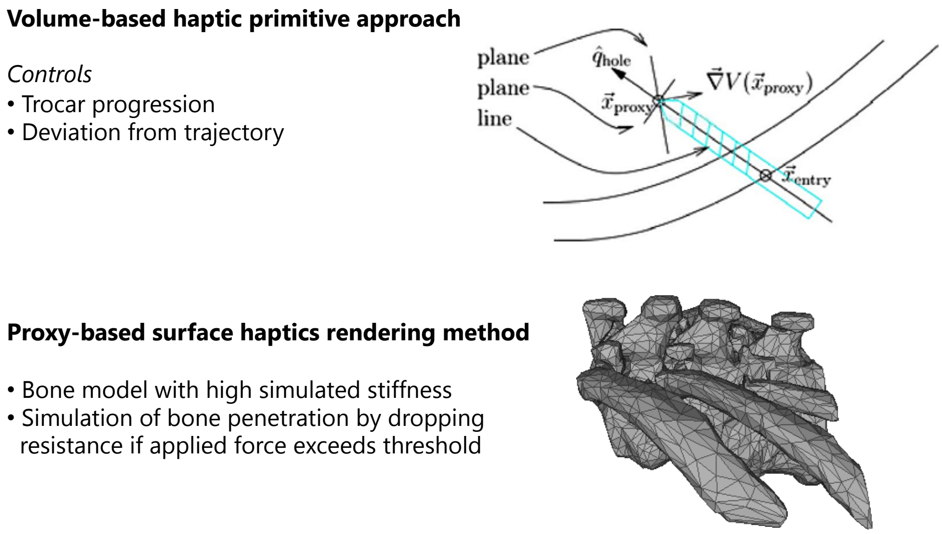


Fig. 28: Haptic feedback based on haptic primitives and proxy-based surface haptics.

CT simulation From Fig. 29-right, CT imaging is used in our setup to mimic the situation in the real OR which supports the surgeon in instrument navigation, verification of access paths, and injection and control of the distribution of the bone cement. A mockup consisting of printed Styrofoam plates mounted on an aluminum frame represents the CT scanner. A line-laser fixed to the frame marks the image acquisition plane on the patient and the instrument. It can be used to define an entry point and to check whether the instrument is in-plane. Using a footswitch, the operating surgeon acquires CT images, which are displayed on a monitor placed on the opposite side of the patient. The monitor shows three CT slice images with the central image's acquisition plane denoted by the laser line and the left and right images cranial and caudal respectively to the central image. The CT data used in this visualization originates from an anonymized dataset acquired in an actual vertebroplasty procedure. The instrument visualization is achieved by rendering 3D models of the instruments in a clipping plane capping approach developed in [91]. This process is repeated with different anteroposterior offsets for the clipping plane, blending the resulting images to simulate a slice thickness matching that of the original CT slices. The bone cement is modeled as a jagged sphere rendered using the same approach. The final images displayed on the monitor consist of the CT slice image superimposed with the instrument slice rendering.

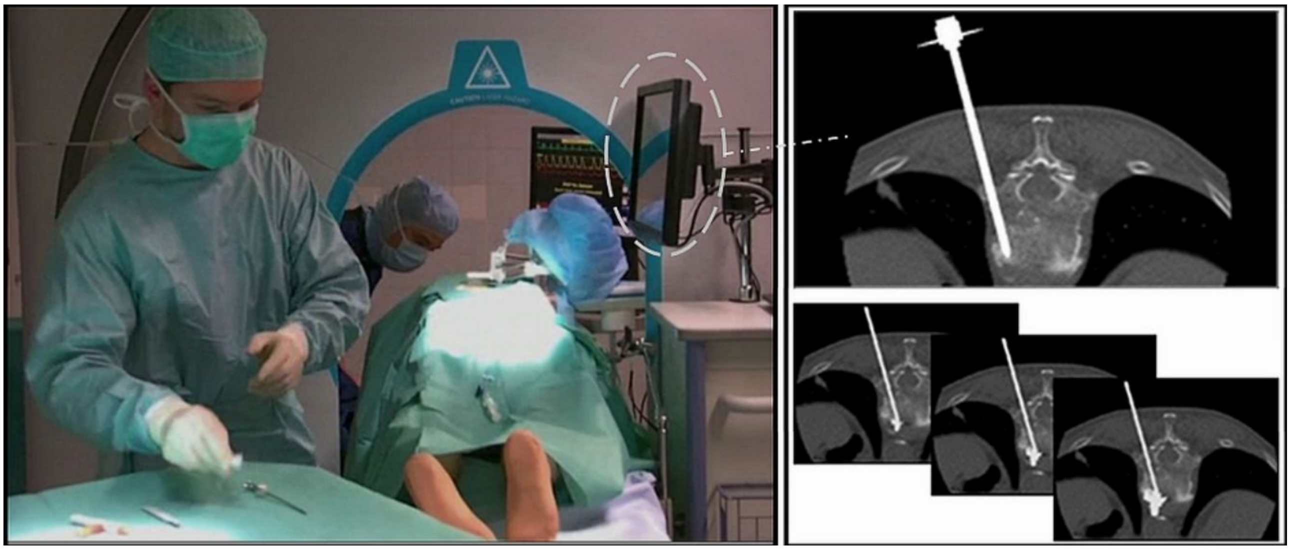


Fig. 29: (Left) A close-up of the operating site. (Right) The CT monitor shows the corresponding CT scans for (a) trocar insertion and (b) gradual cement injection.

3.6. Protocol

The immersion process: The surgeon entered the simulated operating theatre and was requested to put on medical gloves for single use. A short briefing about the patient was given: the patient's name: 'Mr. Huber', age: '79', bone structure: 'osteoporotic bone', the current level: 'oxygen saturation 98%', and that a local anesthesia was conducted, thus the 'patient is currently awake'. Then they were informed about the scenario and made familiar with the theatre environment. Afterwards, the independent anesthetist assumed his position on the other side of the CT scanner. The CT scanner and the patient monitor sound were turned on. The three surgical workflow steps were performed with real medical instruments and with the aid of VR, haptic, and multisensory feedback at specific instants of the procedure. During surgical workflow step 3, the simulation instructor introduced a visualization depicting cement extravasation into a perivertebral vein. Furthermore, the physiology of the computerized mannequin was influenced by the instructor simulating a lung embolism by gradually lowering the oxygen saturation from 98% to 80% beginning at a standardized point during the procedure. The simulation was stopped after the communication between the surgeon and the anesthetist occurred which determined their acknowledgment that an adverse event occurred.

3.6.1. Results and discussion

Four surgeons participated in a user-study involving the completion of the surgical workflow steps described in the previous section. The participants had varied experience: two senior experts (>150 executed vertebroplasties) and two junior experts (<150 executed vertebroplasties). Each participant was immersed individually in our VR surgical simulator in combination with a mannequin connected to the monitoring device. An independent person with knowledge of physiological responses and monitoring acted as the anesthesiologist. The surgeons were asked to give feedback using the Likert scale - a type of psychometric response and the most widely used scale in survey research. The subjects specified their level of agreement to a statement in our questionnaire. The 5-pt Likert scale format was: (1) Strongly disagree, (2) Disagree, (3) Neither agree nor disagree, (4) Agree, (5) Strongly agree. We assessed the face validity of the medical simulation environment, which is a subjective validation and usually used during the initial phase of test construction [46]. However the intent of the evaluation goes even beyond, trying to get answers related to obstacles hindering immersion into the simulation scenario and to disseminate these to the research community.

3.6.2. Survey results

3.1 provides details on the average scores for the survey. The scores were categorized as: workflow steps face validity, crisis simulation, face validity and simulation environment. There were consistently high levels of agreement for all the questions. The group of surgeons thought that the modeling of workflow step 1 is realistic. The majority found that the realism is high during workflow step 2. They considered the simulation of workflow step 3 and 4 realistic as well. The questions pertaining to the face validity of the simulation setup were answered with an overall average Likert score of 4.5 - signifying that the simulation is realistic.

Category	Statement: What do you think about the realism of ...	Score	Expert*
Workflow Step I: Definition of the entry point (1=not realistic, to 5=very realistic)	... the CT visualization of the needle	4.75 (0.25)	4
	... making the stab incision at the entry point using a scalpel	4.5 (1.00)	3
Workflow Step II: Navigation of the trocar into the vertebra (1=not realistic, to 5=very realistic)	... the CT visualization of the trocar	4.5 (0.33)	5
	... the haptic feedback (bone distinguishable from other tissue)	4.0 (0.66)	1
Workflow Step III: Application of cement (1=not realistic, to 5=very realistic)	... the CT visualization of the cement	3.75 (0.91)	4
	... the handling of the cement injection syringe	3.25 (0.91)	3
Crisis Simulation Complication cement leakage and lung embolism (1=not realistic, to 5=very realistic)	... the CT visualization of the cement leakage	4.75 (0.25)	4
	... the presentation of lung embolism on the patient monitor (signals seen by anesthetist, audio signals heard also by the surgeon)	4.5 (0.33)	4
	... the anesthetist communication	4.0 (1.33)	5
Face Validity (1=not realistic, to 5=very realistic)	... the appearance of the instruments	4.5 (0.33)	4
	... changing the instruments	4.25 (0.25)	1
	... the movement of the instruments	3.0 (0.66)	1
	... the function of the instruments	4.5 (0.33)	3
	... the workflow representation	4.75 (0.25)	4
	... the CT scanner monitor presentation	5.0 (0.00)	5
	... the CT scanner interface (footswitch)	5.0 (0.00)	5
	... the laser line representing the CT scanning plane	5.0 (0.00)	5
Simulation environment (1=strongly disagree, to 5=strongly agree)	The simulation environment is a realistic representation of a real OR.	4.75 (0.25)	-*
	I would behave in the same way even in real life.	4.5 (0.33)	-*
	The simulated procedure in the Simulated Operating Theater is a good method for training technical skills.	4.62 (0.29)	-*
	The simulated procedure in the Simulated Operating Theater is a good method for training team skills.	4.5 (0.67)	-*

Table 3.1: The mean values of the statements scored on 5-Point Likert Scale (variance in parentheses). *The expert was excluded from the overall score, as this participant performs vertebroplasty, in contrast to the simulated CT-based environment, under fluoroscopic guidance only.

3.6.3. Limitations

The lowest score was assigned during workflow step 3 related to the usage of the syringe and visualization of the cement in CT. Here, surgeons differed in response claiming that the manual pressure they had to apply on pushing the stamp of the syringe was either too low or too high. A major complaint of the surgeons was that the movement of the trocar used in workflow step 2 was not sufficiently limited by the bone tissue. After the surgeons placed the trocar inside the vertebra they could still move it laterally. This aspect does not reflect the surgeons experience with this instrument behavior during real surgeries and therefore it decreased the level of realism.

3.6.4. Overall assessment

The complete simulation environment was ranked with an average Likert score greater than 4.5 when assessing all aspects of the realism of the simulation environment, specifically on whether it is suitable for the training of technical skills team training.

3.6.5. Synopsis

The goal for the modern learner is to arrive at the bedside of a real patient with proficiency already demonstrated in the requisite skills. In this process, the most expensive and scarce resource is the experienced clinical instructor. In this area, the synergy between computer-assistance and real medical instrumentation can make invaluable contributions by enabling focused and deliberate practice to further motivate the trainee. Thus, clinical education specialists need a customizable medical simulation environment to experiment with new learning models and training regimens. In this paper, we outlined some key aspects that we believe should characterize a customizable simulation environment. We have designed a procedural VR simulator, in combination with mannequin technology, into an OR training and assessment environment. The simulator is capable of representing the entire surgical workflow including a medical imaging device simulation with the capacity to use patient-specific data, thus allowing the representation of a broad range of anatomical and pathological variety. Real surgical tools and instruments are augmented with realistic haptic feedback. Inherently, we also addressed a broad spectrum of human sensory channels such as tactile, auditory and visual channels in real time. To our knowledge, this is the first VR simulator with the capacity to control the introduction of adverse events or complication yielding a wide spectrum of highly adjustable crisis simulation scenarios. Moreover, this is the first study that combines a VR simulator with a computerized mannequin simulator in an OR crisis simulation scenario. Future work will involve the improvement of: (i) haptics feedback, in particular limiting the lateral movement of the trocar inside bone tissue and (ii) CT scanner being substituted with intraoperative C-arm fluoroscopy. NOTE: we will add (the possibility to use) fluoroscopy as a second imaging modality for guidance.

3.7. Conclusion

This study has demonstrated the face validity or realism of our medical training environment. Our conclusions validate the importance of incorporating surgical workflow analysis together with VR, human multisensory responses, and the inclusion of real surgical instruments when considering the design of a simulation environment for medical education. The proposed training environment for individuals can be certainly extended to training medical teams.

4. Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis

In this chapter derived from 'Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis' [147] and [141] a unique approach is presented to orthopedic and trauma surgery training which evaluates both surgeon technical and non-technical (cognitive) skills during their immersion in a complete medical simulation. The simulation environment combines a VR surgical procedural simulator and computerized mannequin in a novel training setup, and also addresses a broad spectrum of human sensory channels, such as tactile, auditory and visual, in real-time. Included are two crisis scenarios which allow for the evaluation of the effect on the low-level surgical skills. Training for these mixed-mode scenarios can be evaluated on the platform, allowing for improved assessment and a stronger foundation for credentialing, with the potential to reduce the occurrence of adverse events in the OR. The three conditions outlined in 3.2. are accounted for through the following key research contributions:

- Scientific evaluation and validation of the combination of a VR surgical procedural simulator and computerized mannequin is a novel training setup for medical assessment and training.
- Research and development of a surgical simulator reflecting complete workflows of surgical procedures, including intra-operative crisis scenarios which potentially result in adverse events. The capability to deliberately expose the trainee to realistic adverse events facilitates the research on robust error metrics for performance assessment in adverse or critical situations, as well as on strategies for augmentation and amplification of error perception and mediation of error recovery strategies, e.g. in deliberate practice or procedure rehearsal scenarios.
- Establishment of structured assessments of surgeons' and surgical team's technical and non-technical skills during simulation-based training that enhance impact on the acquisition, application, and retention of teamwork skills in healthcare.

4.1. Aim

Scientific evaluation and validation of our work was conducted together with 19 junior surgeons in order to achieve the following goals: (i) to provide a qualitative measure of usability, (ii) to assess vertebroplasty technical performance of the surgeon, and (iii) to explore the relationship between mental workload and surgical performance during crisis.

4.2. Methodology

4.2.1. Visualization

Unlike previous works [148][149], in which a CT simulator was used, a C-arm fluoroscope is integrated into the simulation, since it is the intraoperative modality of choice during vertebroplasty. The visualization is implemented on a GPU-accelerated system, using a digitally-reconstructed radiograph (DRR) to simulate fluoroscopy from CT data [105]. Using a footswitch, the participating surgeon acquires AP and LAT C-arm images, which are displayed on a monitor placed on the opposite side of the patient. The DRRs are generated by CT data originating from an anonymized dataset acquired in an actual vertebroplasty procedure. A model of the surgical instrument (trocar) is rendered into the generated image using OpenGL.

4.2.2. Haptics

The haptic device uses only translational force feedback and attached to it is a needle which is introduced through a hole into the mannequin body. This hole is used as a pivot point to establish force feedback similar to the approach described in [148]. Haptic feedback is generated using the H3D API [53]. A VR-based anatomical surface-based model allows for the calculation of feedback forces from previously segmented CT patient data and provides a haptic rendering percept that clearly distinguishes between soft tissue and bone. A mesh of the vertebra with a high simulated stiffness is used to render the hard cortical bone structure, dropping the resistance if a particular force threshold is exceeded. In order to constrain the instrument movement in lateral directions, we use a Magnetic Geometry Effect [39] [40] as described below. Fig. 30 visualizes the haptic model of a trocar insertion towards the vertebra. In Fig. 30a, no force is applied to the haptic device and the trocar is rendered in the simulated fluoroscopic images at the location reported by the haptic device giving the user the possibility to adjust the instrument trajectory. As soon as the trocar reaches an insertion depth greater than 1cm, a line is fixed at the current trajectory of the instrument. The line is outfitted with a Magnetic Geometry Effect (spring constant = 800) constraining the movement of the trocar in directions perpendicular to the trajectory (see Fig. 30b) simulating soft tissue. Then, at a position where the trocar tip touches the corticalis of the bone, an additional force preventing the trocar from easily penetrating the bone is exerted through a Frictional Surface. The stiffness parameter was set to $k = 11N/mm$, the highest value still generating stable haptic feedback. The dynamic friction coefficient was set to $\mu = 0.6$ constraining the trocar movement on the bone surface (see Fig. 30c). Lastly, on penetration of the bone, defined by the haptic device reported force exertion greater than $30N$, the Frictional Surface is deactivated, leading to a fall through of the trocar into the bone. The Magnetic Geometry Effect spring constant is set to $k = 3500$, the highest value still generating stable haptic feedback, constraining the lateral movements of the trocar. At the moment of penetration, the fixed trajectory is set to the current trajectory, thus making an adjustment of the penetration direction possible. After the penetration, the trocar is rendered at the projection of the haptic device location onto the fixed trajectory. This creates the impression that the trocar is not moving in directions perpendicular to the trajectory (see Fig. 30d).

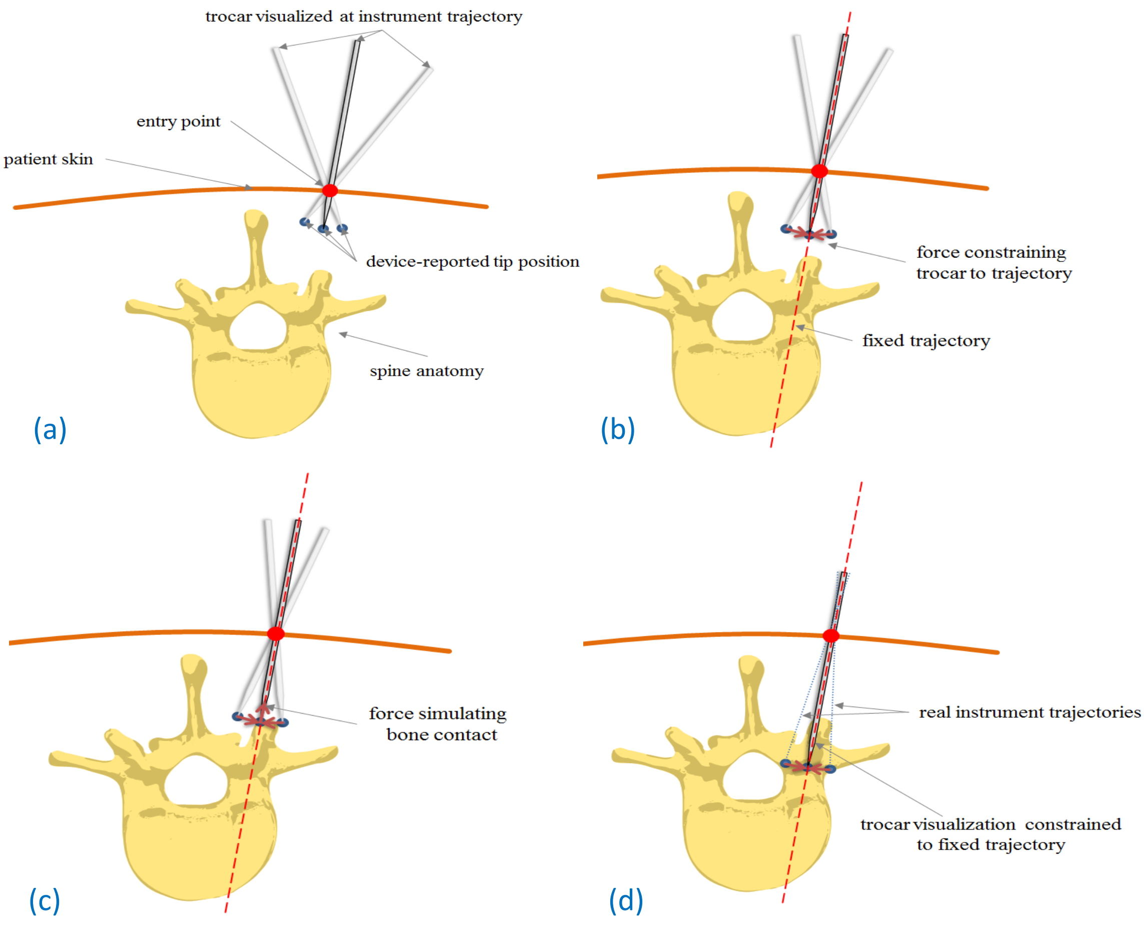


Fig. 30: Virtual haptic rendering and force feedback model during insertion of a trocar towards vertebrae.

4.2.3. Cognitive workload and crisis scenarios

Surgeons' workload is a major adverse determinant of surgical performance in the OR [145]. Workload is generally defined as "the cost incurred by a human operator to achieve a particular level of performance" that "emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, and perceptions of the operator" [55]. Often, workload assessment is subjective. Nevertheless, performance for subjects with different cognitive skills may be affected differently; the same task performance may be observed for higher workload in an experienced surgeon, but the same objective performance could only be attained with lower cognitive workload in a novice [55]. Specifically within surgery, subjective workload has been proposed as an essential link between work demands in the OR and resulting surgical performance [145]. To ensure authenticity in our study, we evaluate the workload for a surgeon under complete workflows of vertebroplasty with occurrence of adverse events, since understanding the impact of risk or danger on clinical judgment and skill is a crucial element of expert-level performance [70].

4.2.4. Protocol

Infrastructure The Institute for Emergency Medicine and Management in Medicine (INM) provided the necessary infrastructure for conducting the validation studies. A replication of a real operating theater with an adjacent control room was available for conducting this study (see Fig. 31). The simulated operating theater (SOT) is equipped with a complete audio-video surveillance system which we used for recording the simulation sessions. Thus, logfiles and video streams for the validation studies could be gathered for offline viewing and interpretation.

Fig. 31: The infrastructure at the Institute for Emergency Medicine and Management in Medicine (INM), Munich, Germany. Our vertebroplasty simulation environment consisted of mannequin and haptic devices with real surgical instruments, real-time C-arm fluoroscopy visualization (bottom left), and an independent control room varying mannequin physiology and crisis scenarios (bottom right).

Equipment In addition to the standard operating theater equipment, one of the primary features of the SOT is a VR surgical procedural simulator. A mobile C-arm fluoroscope was integrated into the SOT. Two monitors were integrated into the SOT for displaying LAT and AP views, as determined by two expert surgeons. The mobile C-arm was fixed relative to the OR table. Apart from the VR surgical procedural simulator,

Fig. 32: (Left) Surgeon inserting a trocar and anesthetist. (Right) Surgeon having finished trocar insertion, scrub nurse in background.

a computerized mannequin simulator manufactured by Gaumard (HAL S3000) was embedded into the full-scale simulations. This simulator is a mid-fidelity mannequin, which allows manipulation of the mannequin's physiologic parameters through a control interface located in the control room. The VR surgical procedural simulator was placed between the upper and the lower body part of the mannequin simulator. Both simulators together represent a virtual and physical patient. This setup was then enveloped with surgical drapes.

Workflow The traditional workflow consists of (i) entry point definition on patient skin, (ii) navigation of the trocar inside vertebrae, and (iii) cement injection. For the purpose of this study, we focused solely on phase (ii) of the workflow (see Fig. 32).

Subjects 19 junior surgeons with no prior experience in vertebroplasty participated in this study. These subjects were divided into two groups with equal distribution of male/female participants, and underwent two different crisis scenarios in the OR. The focus was to assess low-level task performance in the face of adverse events, hence the participation of only junior surgeons.

Briefing The senior surgeon overseeing the study explained the vertebroplasty procedure and workflow to the 19 subjects. He also described various crisis scenarios which may occur during an actual procedure. The duration of the briefing was 15 minutes.

Study Protocol The physiology of the mannequin was turned on and the VR surgical simulator was restarted to initialize the simulated X-ray imaging. The participating surgeon was informed about the study protocol and made familiar with the theater environment. The training phase consisted of the following sessions:

- Informal training: During this stage, subjects could get familiar with the trocar and the medical imaging.
- Training run 1: In the first run, subjects had to insert the trocar correctly. As guidance, a green 3D cylinder representing the correct insertion path was integrated into the medical imaging. The position of the 3D cylinder was defined by two medical experts.
- Training run 2: The subjects had to insert the trocar without the guidance from Run 1 into the simulated patient and navigate it towards the target vertebra.
- Crisis scenario: A standardized theater team consisting of an anesthetist and a scrub nurse assumed their positions. The subject entered the SOT through an OR door with gloves on, surgical mask and cover. A short briefing about the patient was given: the patient's name: 'Mr. Max Müller', age: '79', bone structure: 'osteoporotic bone' and that a local anesthesia was conducted, thus the 'patient is

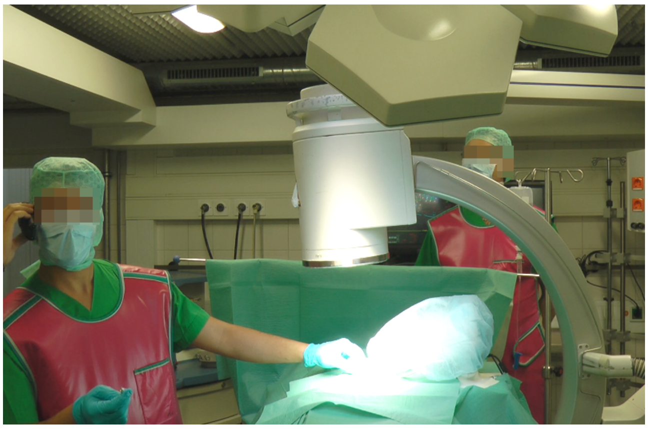


Fig. 33: 'Phone call' crisis scenario. The nurse holds the telephone.

currently awake', Scenario: 'the head surgeon defined the entry point where to insert the trocar'; 'he had to leave and you have to continue the procedure'. The participant had to put on a lead vest given by the scrub nurse and then start inserting the trocar. The surgical workflow step was performed with real medical instruments and with the aid of VR, haptic, and multisensory feedback at specific instants of the procedure. During this surgical workflow step, the simulation instructor introduced either the crisis 'phone call' or 'patient discomfort' on a standardized point during the procedure namely when the trocar reached a specific depth (0.5 of total depth) within the patient's body:

- Crisis 'phone call': The phone rang. The scrub nurse responded to the phone call. The caller was the head of the surgical department and wanted to speak to the surgeon about an urgent issue. The crisis ended when the nurse ended the phone call (see Fig. 33).
- Crisis 'patient discomfort': The patient started to voice discomfort. Additionally the heart rate went up and the anesthetist reported to the surgeon that the patient feels pain. The surgeon had to inject more local anesthetics to solve the crisis. The crisis ended when the patient said: "I am feeling better now". During the evaluation, participants were divided into two groups during crisis: Group I: 3 women and 7 men participated in crisis scenario 'phone call'; Group II: 3 women and 6 men participated in crisis scenario 'patient discomfort'.

4.2.5. Measured metrics

1. The surgeon technical skills were calculated by the location of the trocar and the time elapsed being recorded to compute the accuracy and speed of the participants at each stage of the experiment. Trocar positioning was measured using a gold standard patient CT segmentation in which the start and end point of the trocar trajectory was indicated by two expert surgeons. The time taken to deal with the interruptions, i.e. speaking over the phone or administrating anesthesia, was excluded in our analysis. In addition, number of X-ray shots and their duration were recorded to measure the overall administered radiation dose.
2. The surgeon cognitive workload was assessed by posing the following questions:
   (a) What is the association between mental workload and surgical performance during intra-operative crisis?
   (b) What is the difference between mental workload during the training runs and intra-operative crisis?
   (c) What is the difference between the two crisis scenarios?
3. The face validity and training value of the simulation environment was evaluated via 13 questions on a 5-point Likert scale (1 "not realistic" to 5 "very realistic").

4.3. Results

Surgeon technical skills The analysis of performance is shown in Table 4.1. It should be noted that in case of crises scenarios, position information was not properly recorded for 3 participants due to technical difficulties and therefore were excluded from the performance analysis.

	n	RMSD error (mm)	X-ray exposure (sec)	Completion time (sec)
Training run 1	19	4.29 ± 2.22	16.34 ± 14.13	126.64 ± 70.85
Training run 2	19	4.60 ± 2.85	9.94 ± 13.26	81.85 ± 59.25
Crises 1	8	6.69 ± 3.17	12.7 ± 13.0	305.5 ± 208.4
Crises 2	8	5.92 ± 2.82	11.8 ± 15.1	352.5 ± 256.4

Table 4.1: Technical skill results.

A significant difference was observed between groups in terms of task completion time (Kruskal-Wallis: Chi-Square = 26.9. $p < 0.001$). Post-hoc analysis using Tukey-HSD test reveals that the significant difference was between the two training and the immersive runs ($p < 0.05$). Furthermore, no significant difference was observed between different phases in terms of error or X-ray exposure time. Nevertheless, the following observations can be made:

Time Although training subjects resulted in a decrease in time (run 3 vs. run 2), subjects performed significantly slower when placed in an immersive environment and faced with the crises. It should be noted that the time taken to deal with the interruptions was excluded in our analysis.

Accuracy After the training, subjects' performance with no visual aid reached the same level of accuracy as with a visual aid. Even though not significant, subjects tended to be less accurate during the crises. X-ray exposure: Similar to task completion time, there is a trend that training decreases, and facing with crises increases, the X-ray exposure time.

Precision The avg. pedicle diameter for upper lumbar spine is 8 mm, and as such, the deviation from the ideal trajectory should be approximately an average of 4 mm to guarantee precise intra-pedicular placement [10]. Interestingly, our trainees achieve this prior to crisis events which accentuates the obligation to train young surgeons to handle other unexpected complications (such as cement leakage in [148] [149]). An example of the force plots by participant #4 and #6 are depicted in Fig. 34. We observe that the exerted force by participant #6 was greater in magnitude over the course of the experiments. Also, the time taken to complete the experiment was longer compared to participant #4. In fact, the subjective workload, defined previously, was 0.5 for participant #4 and 4.4 for participant #6. Further analysis demonstrated that all other participants had a lower workload than participant #6, while only three other participants had a lower workload than participant #4. Fig. 35 demonstrates visually the precision of trocar insertions for these two participants compared to ground-truth. The RMSD were 2.84 mm and 8.87 mm for participants #4 and #6 respectively.

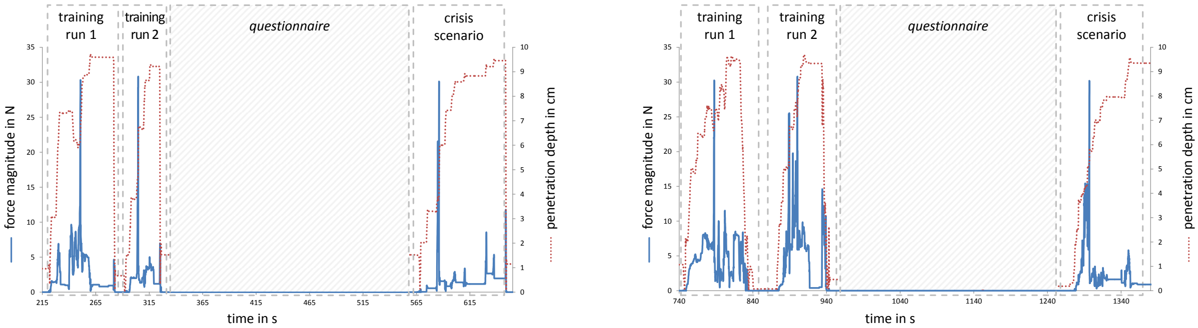


Fig. 34: The figure shows the exerted force magnitude in N as reported by the haptic device (blue line) as well as the insertion depth of the trocar in cm (dotted red line) during the training runs and the crisis scenario. Participant #4 and #6 sequences are shown on top and bottom respectively.

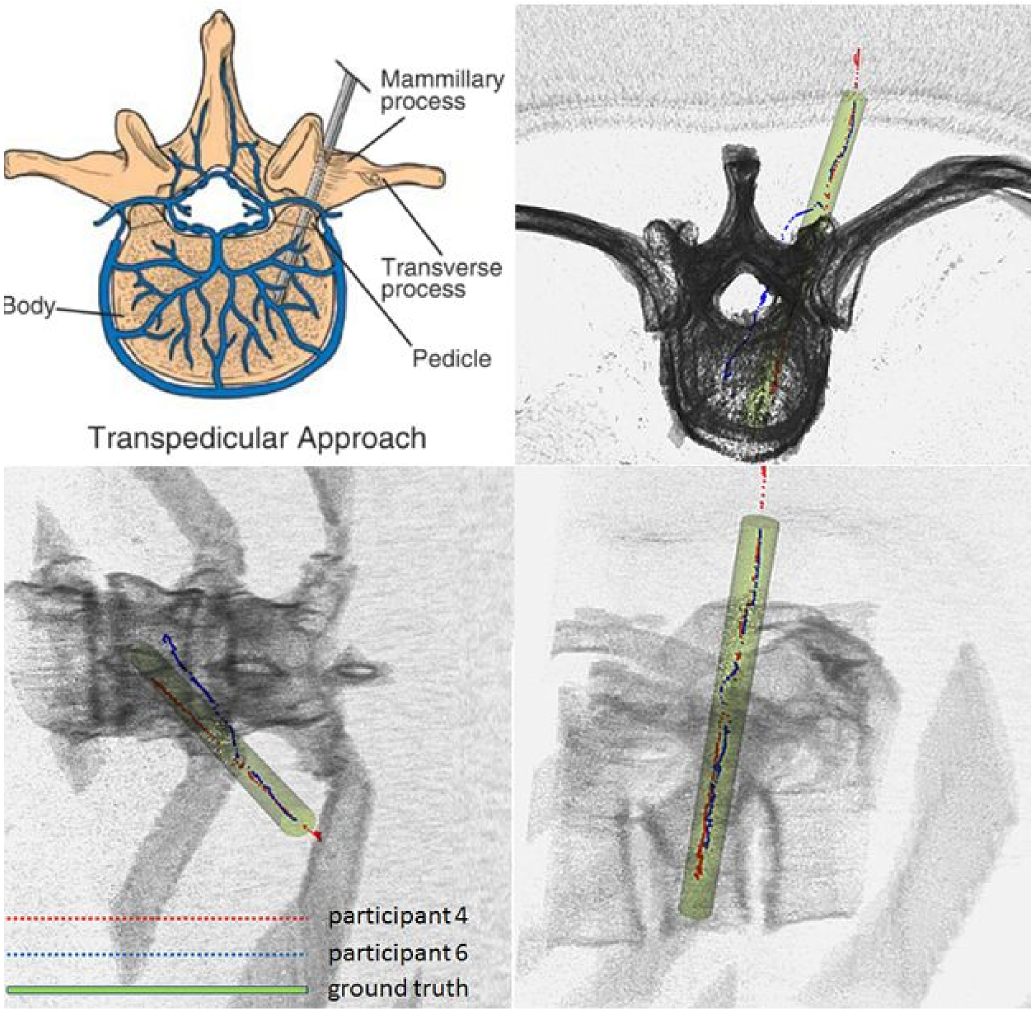


Fig. 35: Drawing of the transpedicular approach and DRRs from different viewpoints and two surgeons' trocar insertions versus ground truth trajectories. Transpedicular approach with a needle traversing the pedicle, from [88] (top-left) and DRR for cranio-caudal view (top-right). AP (bottom-left) and LAT (bottom-right) viewpoints as used in study. Upper row: arbitrary viewpoint (left) Deviation from ground-truth for participant #4 is 2.84 mm (RMSD), while for participant #6 is 8.87 mm (RMSD).

Surgeon non-technical (cognitive) workload Cognitive load was measured using the Surgery Task Load Index (SURG-TLX) [145]: this surgery-specific, multi-dimensional workload measure enables subjective assessments of load relevant to a specific task, distinguishes between different task complexities, and indicates objective performance. The surgeons communicated their workload directly after study completion based on five items: mental demand, physical demand, temporal demands, complexity, and situational stress). Surgeons received the following instruction: "Please rate your average workload during the procedure you just completed" on a scale range: 0 = very low. 20 = very high. All item responses are summed up to a score indicating overall subjective workload during the task execution. Some measures below are reported as (mean ± standard deviation).
Results for research question (1): Association between mental workload and surgical performance during intra-operative crisis?

- Mental workload during intra-operative crisis: 44.89 (CI 36.65 − 53.13).
- Deviation from gold standard: 5.79 ± 2.21 mm (Range 2.13 − 8.87).
- Partial correlation analyses (controlled for professional tenure): The association between inaccuracy and mental workload was $r = 0.52$, $p = 0.049$ (df = 13).

Results for research question (2): Difference between mental workload during the training runs and intra-operative crisis?

- Mental workload during training runs (reported after training runs): 35.64 ± 13.11.
- Mental workload during crisis (reported after intra-operative crisis): 44.63 ± 17.28.
- Test for mean difference (t-test for paired samples): $T = -3.81$ (df = 18), $p < 0.01$.

Results for research question (3): Subjective reports in regard to experienced level of distraction during the crisis, i.e. difference between the two crisis scenarios (telephone call vs. patient discomfort)?

- Outcome: Perceived interruptions during the crisis (item 6 of SURG-TLX index)
- Group - phone call ($n = 8$) : 8.83 ± 5.21.
- Group - patient discomfort ($n = 8$) : 4.83 ± 2.18.
- Test for mean difference (t-test for independent samples): $T = 2.13$ (df = 17), $p < 0.05$.

Face validity and training value Face validity: In Table 4.2, results are reported as mean and variance. The training phase was scored as 4.41 ± 0.41; the crisis simulation was scored as 4.22 ± 0.67; the phone call crisis was scored as 4.38 ± 0.55; and the patient discomfort crisis was scored as 4.16 ± 0.37.
Training value: In Table 4.3, the overall training value of the simulation environment was given a score of 4.68 ± 0.48.

4.4. Discussion

Hierarchal Task Analysis There are three main concurrent sub-tasks involved in a formal vertebroplasty procedure: targeting, controlling the view-point, and maintaining the patient safety. The viewpoint control involves applying short bursts of X-ray to view the trocar and anatomical structures. Maintaining patient safety is a competing but over-arching task which requires minimizing the amount of X-ray radiation. In our case study, in addition to these sub-tasks, there are emergent interruptions that also compete for the attention of the operator. As a general principle of multi-task performance analysis, when faced with an interruption, subjects may perform poorly if the interruption provides no useful information, consuming cognitive resources needed to perform the task. Such irrelevant distractions impose a type of cognitive burden known as extraneous load, which is associated with unnecessary information and should be reduced or eliminated. Failure to do so

Face validity (1=not realistic, to 5=very realistic)	Statement	Average score
Training Phase	How realistic is the visualization of fluoroscopy?	4.41 (0.32)
	How realistic is the haptic feedback (difference: Bone vs. tissue)?	3.83 (0.76)
	Is the visual representation of the surgical trajectory helpful to get familiar with the simulation environment?	4.78 (0.19)
	Is the visual representation of the surgical trajectory helpful to get familiar with the surgical workflow step?	4.62 (0.35)
Crisis simulation	How realistic is the simulation environment?	4.33 (0.44)
	Would you act the same way in real life?	3.83 (1.21)
	How realistic were the medical instruments?	4.47 (0.48)
	How realistic was the instrument movement?	3.85 (0.86)
	How realistic was the instrument functioning?	4.28 (0.63)
	How realistic is the fluoroscopy simulation (foot pedal)?	4.58 (0.43)
Crisis I	How realistic was the scenario phone call?	4.38 (0.55)
Crisis II	How realistic is the scenario "patient in pain"?	4.01 (0.25)
	How realistic was the communication with the anesthetist?	4.31 (0.48)

Table 4.2: Feedback of the 19 study participants on face validity; The mean values of the statements scored on 5-point Likert scales; The values in parentheses are the (variance).

Training value of simulation (1=strongly disagree, to 5=strongly agree)	Statement	Average score
	Is the simulation environment suitable for OR team training?	4.82 (0.25)
	Does the medical training environment represent a real OR?	4.46 (0.69)
	The simulation environment is suitable for education and training?	4.77 (0.51)

Table 4.3: Feedback of the 19 study participants on training value; The mean values of the statements scored on 5-point Likert scales; The values in parentheses are the (variance).

may lead to cognitive overload, which in turn, severely disturbs the efficiency and effectiveness of information transfer [131]. In this study, this phenomenon was verified empirically, as our data shows that introducing interruptions significantly increased the task completion time. Although not significant, surgeons tended to perform less accurately when faced with interruptions.

Secondly, performing a targeting task requires navigating the trocar to the target while depressing the X-ray button for the time required receiving more accurate visual information about the relative position of the tool and the target. While obtaining visual-spatial information is necessary to succeed in the navigation task, it is nevertheless subjective as some junior surgeons have developed higher skills of perception and spatial reasoning than others during their training. Hence, some may need to obtain more X-ray images to attain the same level of perceptual accuracy or to form a spatial representation of the perceptual-motor task space than others. This is manifested in the form of large standard deviations across subjects as illustrated in Table 4.1.

Thirdly, increasing the number of X-ray images may, on one hand, improve the navigation performance, while on the other hand, compromise the safety of the patient and of the surgical team. Novice users tended to apply more x-ray dose, in order to maintain the same perceptual-motor accuracy, in the 'interruption' conditions. Apart from training the surgeons targeting skill which includes the appropriate placement of the image intensifier, the positioning of personnel relative to the fluoroscope could be a training goal within the fully-immersive simulation environment. It could be organized within an appropriate distance from radiation source, and therefore adjunctively the occupational exposure could be reduced for all those involved in spine operations [127] [97]. The ability to practice a unique skill, such as successfully performing an image-guided targeting task without substantially compromising surgical team safety, is one of the key benefits of our fully-immersive simulation environment. Further, a natural byproduct of our environment is its capacity to allow evaluation of novel AR technologies [92] [17], visualization methods [144] [33], and pose estimation tools [103].

Cognitive workload The more workload the surgeons experienced, the poorer was the surgical performance. Further, the intra-operative crisis imposed, significantly increased workload upon the participants compared to training runs. Finally, there is preliminary support for the construct validity of the different crisis scenarios and their inherent level of interruptiveness. As such, the group that was disturbed through the telephone call perceived increased interruptiveness compared to the group with patient discomfort. However, this difference is not significant (due to the small sample size) but tends to be meaningful.

4.5. Conclusion

We have designed and implemented a surgical simulation platform according to the requirements and technological necessities for creating a low-cost, yet highly realistic VR surgical procedural simulator for multidisciplinary medical training. Through a comprehensive study involving 19 junior surgeons, we assessed both technical and non-technical skills in a realistic simulation environment, detailing our design methodology and outlining an evaluation framework as in [31] that is sensitive to multiple sub-task performance criteria and the ability to distinguish operating conditions.

4.6. Limitations

One of the major limitations is the small sample size, though the increased mental workload was evident based on our results. In order to better investigate the impact of interruptions on patient safety and surgical accuracy a larger sample size would be required. Another limitation was the lack of time of the study participants. Hence the validation was focusing on one workflow step rather than the entire surgical workflow.

5. Visuospatial Performance on C-Arm Image Guided Vertebroplasty: VR Surgical Simulator Evaluation

This part of the thesis reports on the design and evaluation of a performance metric for a vertebroplasty VR simulator. Based on a hierarchical task analysis of the procedure well-posed performance metrics for each phase of the procedure were derived. A comprehensive study involving 14 surgical residents was conducted under the direction of three osteosurgical experts with the aim to demonstrate a training advantage that accrues to surgeons with higher spatial reasoning skills, in terms of their improved performance and rate of skill acquisition.

5.1. Introduction

Current surgical practice depends upon several factors that include visuospatial and technical ability. However, while some surgeons believe that only resident trainees with appropriate levels of innate ability should be selected for training [24], this policy is currently not in place; although accepted long ago within other professions with similar visuospatial requirements (e.g. aviation flight training). Assessment of competency during surgical training is also far less rigorous than in aviation [90], and trainees might have to drop out of the surgical program late in their training if they are not able to perform. Today's surgical apprenticeships allow senior surgeons to assess trainees and observe their skillsets informally, however, the availability of proven, reproducible techniques to assess individual surgical aptitude reliably remain limited. Key parameters such as technical and visuospatial ability may improve the selection process of resident trainees which might avoid allowing individuals to proceed to higher training without basic skills competency.

Psychometric research into individual differences has enabled a differentiation between several cognitive abilities, and to arrange those in hierarchical models of general to specific cognitive abilities [16]. The top tier is represented by a single factor - general intelligence. It is defined as the shared factor loadings of a number of second tier factors such as reasoning ability, visuospatial ability, and memory. Of these, visuospatial ability refers to the human cognitive ability to form, retrieve, and manipulate mental models of a visual and spatial nature [81]. Visuospatial ability has been successfully linked to a variety of surgical and medical skills [57].

Two key reasons make visuospatial ability interesting over other cognitive abilities. First, contrary to other cognitive ability factors, visuospatial ability has not been selected as an assessment metric during today's academic phase of the medical curriculum. As such, it may largely account for differences in performance in surgeries with a spatial emphasis [85]. Second, whereas general cognitive ability is especially important during early medical learning, visuospatial ability remains important throughout training, due to task specific aspects of the majority of surgeries [68].

John Carroll identifies five third-tier factors that together form visuospatial ability [16]. These are Visualization, Spatial relations, Speed of closure, Flexibility of closure, and Perceptual speed. Below, these factors are discussed in light of surgical training, leading to an outline of the current work. Visualization is the ability to manipulate complex mental representations of a visuospatial nature. Mental manipulations required in visualization tests can be quite elaborate, and require reasoning about the rotation of complex spatial shapes, perspective taking, or others. Spatial relations indicate the ability to quickly manipulate simple mental representations of a visuospatial nature. Tests for spatial relations generally correlate positively with surgical simulator performance [113]. Speed of closure is defined as the ability to identify partly obscured spatial forms. Tests for speed of closure are used in [112] [140] which determined significant correlation of task-on-time on simulator dexterity drills. Flexibility of closure indicates the ability to identify spatial forms that are specified to the learner in advance in a cluttered visual environment. Tests in [128] for this factor show positive correlations with surgical performance in some research. Perceptual speed of recall is the ability to quickly identify a given shape from a number of alternatives, which can be implicated as a factor in performance [85].

In orthopedic and trauma surgery, mobile C-arm fluoroscopes are used frequently to assess the complexity of fractures, guide the surgical procedure, and verify the results of surgical repair. Interestingly, according to the Surgical Skills Task Force of the American Board of Orthopaedic Surgery (ABOS), in collaboration with the American Academy of Orthopaedic Surgeons (AAOS), resident surgeons do not have any experience and skillset with C-arm imaging when they begin residency program. They are acquired at different rates by different learners. Many of the skills are not intuitive and/or are not incorporated into previous training prior to entering residency. Even more critical is the capacity for the resident trainee to perform 3D spatial recognition from 2D images, which is termed "mental mapping", which becomes the focal point of any successful surgeon. The use of computer-based simulation seems to be an especially important factor in the success of X-ray imaging training efforts [13], yet no objective metrics for surgical performance level assessment for X-ray imaging has been disseminated yet. A preliminary study with 14 surgical residents was conducted on a vertebroplasty simulator. The participants ran through a study protocol and their performance was recorded with objective metrics in both the spatial reasoning test (time, accuracy) and the surgical simulation (duration of x-ray exposure, C-arm movement and instrument handling). The results corroborate that we were able to quantify the training advantage that accrues to residents with higher visuospatial reasoning skills, in terms of their improved performance and rate of skill acquisition.

5.2. Methods

5.2.1. Medical procedure and surgical tasks

A vertebroplasty is performed, for patients with vertebral fractures, in the following general steps, which have mostly been reproduced with our simulator. The patient is placed on a radiology table in a prone position, and the back is exposed. Fluoroscopy with the C-arm is used to verify the level, and then adjusted to obtain the best view in an anterior-posterior (AP) position, and identify the pedicles. The skin is then incised, which was also simulated. Then a trocar is inserted aiming for the lateral aspect of the pedicle, and then advanced within the pedicle towards the center of the vertebral body (see Fig. 36. The C-arm needs to be adjusted from an AP position to a lateral view to see if the trocar is still in the pedicle or has reached the vertebral body. The AP view is needed to advance the trocar in the pedicle, while the lateral view is necessary once the needle has reached the vertebral body to target the middle of the vertebra, or slightly anteriorly. The injection of cement for vertebroplasty can then be safely performed.

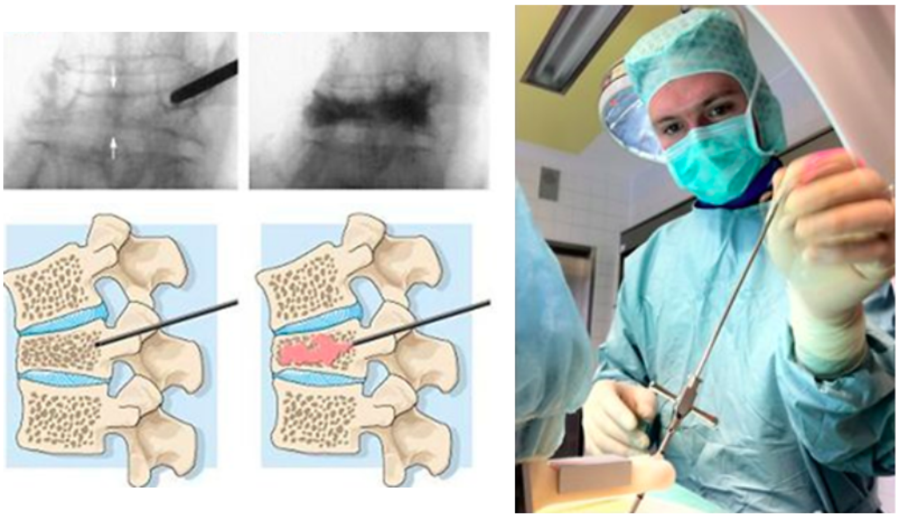


Fig. 36: Vertebroplasty procedure. Using C-arm fluoroscopy, surgeons navigate a trocar inside the vertebral cavity to inject cement in an effort to stabilize the fracture.

5.2.2. Technical components

A simulator platform (see Fig. 37) was used in which real surgical instruments were integrated as in [147]. Onto a haptic device (Novint Falcon, Novint Technologies Inc.) a haptic end-effector is attached to enable the usage of different surgical instruments. These can then be tracked in 3D and their orientation is calculated by using a fixed entry on a surgical suture pad, which mimics the patient's back.

Fig. 37: (Left) Our stand-alone VR simulator with C-arm control and haptic feedback. (Right) Our simulator in a controlled medical simulation setting at the Institut für Notfallmedizin und Medizinmanagement, Munich, Germany.

Surgical instruments To set the fixed entry point on the skin pad as a first workflow step, a custom made instrument was designed together with medical experts. After having this instrument positioned the subject needs to step on a pedal to give feedback that the entry point was chosen. In the second surgical workflow step the subject attaches a trocar to the haptic end-effector to insert it into the vertebral body.

Visualization A monitor shows a virtual mobile C-arm, two numbers, the orbital and the angular angle, X-ray imaging and a radiation symbol, when an X-ray shot is triggered. The X-ray imaging is simulated using a digitally-reconstructed radiograph (DRR) from CT data [147]. It is computed on a graphics processing unit (GPU) and the instrument model is rendered into the generated image. For acquiring an updated X-ray image the subject needs to step on a tagged foot pedal.

Haptics The development of the haptic feedback is as [147]. However, we improved the author's concept by the addition of a haptic interaction between the instrument and the cancellous bone inside of the vertebrae (see Fig. 38). The simulation of this interaction is realized using the notion of haptic primitives [83]. A plane primitive is located at the proxy position, the internal representation of the instrument tip. The orientation of the primitive is set according to the actual instrument trajectory, defined by the instrument tip, as measured by the haptic device, and the pivot point. The strength of the primitive is controlled by the volume intensity at the visualized instrument's tip (insertion depth corresponding to proxy position in direction of the fixed instrument trajectory) via a transfer function, and the insertion depth, i.e. the distance between bone entry and instrument tip. Furthermore, we altered the parameters used according to feedback from two medical doctors: a) we set the spring constant of the Magnetic Geometry Effect corresponding to insertion depth, i.e. $k = 100$ at 15 mm, $k = 200$ at 20 mm and $k = 300$ at 25 mm, and after bone penetration to $k = 1200$, b) we reduced the stiffness of the Frictional Surface to $k = 9N/mm$ to improve the stability of the haptic feedback, and c) accordingly reduced the force threshold deactivating the Frictional Surface, leading to a fall-through of the instrument into the bone, to 20 N.

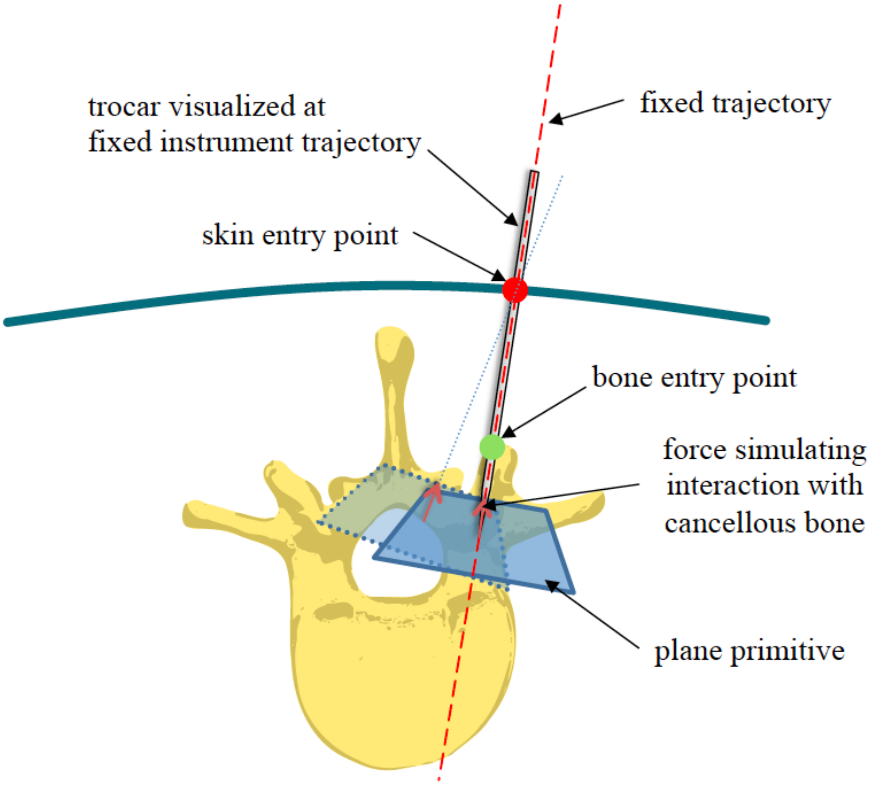


Fig. 38: Depiction of the haptic feedback and primitives.

C-arm control The virtual mobile C-arm and the corresponding perspective of the X-ray visualization are controlled in two angles (orbital, angular) and translation along the superior-inferior axis, with three hardware sliders. Further, the X-ray visualization can be aligned horizontally or vertically by pressing a switch. In order to denote the end of a scenario in the log file, i.e. once the placement of the trocar inside of the vertebral body, a switch is pressed. Fluoroscopy event timestamps are also logged.

5.3. Protocol

Infrastructure The Klinik für Allgemeine, Unfall-, Hand- und Plastische Chirurgie provided the necessary infrastructure for conducting the validation studies. Two rooms were used. One room was equipped with three personal computers for performing the spatial reasoning and the Santa Barbara evaluations. In the second room three VR surgical simulators were available for conducting the vertebroplasty procedure. During the study the simulation sessions were recorded with video cameras. Thus, log files and video streams for the validation studies could be gathered for offline viewing and interpretation.

Equipment Each simulator enveloped with surgical drapes was spatially separated by walls and was placed on a table. On the right hand side a monitor was showing fluoroscopic imaging, a 3D model of a patient and mobile C-arm. Further on two numbers were displayed to show the angles the surgeon could adjust with a C-arm hardware controller in order to get an appropriate view of the operating site. To acquire a new fluoroscopic image the surgeon had to step on a pedal.

Workflow The traditional workflow consists of (i) entry point definition on patient skin, (ii) navigation of the trocar inside vertebrae, and (iii) cement injection. For the purpose of this user study, we focused on phase (i) and (ii) of the workflow.

Subjects $N = 14$ surgical residents, $N = 9$ male (64.2%); Procedure specific experience: 7 subjects observed vertebroplasties (≤ 50). 5 surgeons performed a vertebroplasty (≤ 25). 3 subjects did not have any experience.

The study participants were divided into 2 groups. Each group got a briefing before working on the VR surgical simulator.

Briefing The senior surgeon overseeing the study explained the vertebroplasty procedure and workflow to the 14 subjects on the VR surgical procedural simulator. The duration of the briefing was 10 minutes. Study Protocol: At the beginning of each scenario the subject was asked for their participation number to be able to log the data of each simulation and further on to identify the subject in the video. They were informed about the study protocol and were made familiar with the simulation environment. The simulator study consisted of 5 scenarios of varying difficulty levels. These were defined by three expert surgeons.

- Scenario 1: During the first scenario subjects could get familiar with the entry point marking tool, the trocar and the medical imaging. For all the subjects the same data set (difficulty level - easy) was used.
- Scenario 2-5: In those scenarios the subjects had to go through the workflow steps i and ii on 4 randomly chosen data sets of varying difficulty levels (Dataset 2 - hard (but the axial cut does not show the pedicles too well); Dataset 3 - easy; Dataset 4 - moderate (but the axial cut does not show the pedicles too well); Dataset 5 - moderate). After each scenario the subject's performance was reviewed by an expert surgeon to give the subject direct feedback. This feedback was given to the subject based on a 'control view' consisting of a transversal rendering of the vertebral body and the inserted trocar.

Evaluation The approach taken for our first main analysis was to establish a ranking of the subjects on a non-clinical visuospatial reasoning task. And in addition, to rank them on a visuospatial reasoning task testing their abilities to reason about volumes and orthogonal 2D views (using Shepard and Metzler blocks matching task, and the Santa Barbara (SB) Visuospatial Test questionnaire [19]). The Santa Barbara test consisted of determining the shape of the cross section of intersecting 3D objects. This test was shown to verify different visuospatial abilities than the mental rotation test. Their performance was assessed in terms of speed and accuracy.
We then analyzed the trajectories of the needle tip within the context of the 2D views taken from the perspective of the C-arm, for each clinical dataset. The data sets are rich with trajectory information, and so our first main investigation was to consider the subjects' ability to target positions in 3D. We examine the speed and accuracy of their trajectories to target clinical sites, and examine the improvements over each trial as a function of their spatial abilities and their clinical expertise.
Finally, we examined their control of the C-arm and the use of X-ray to acquire views to enable the navigation and targeting. Our hypothesis was that participants with higher spatial reasoning abilities would require lower X-ray dose to achieve performance in the navigation and targeting of the clinical workspace and vertebral targets. In addition, over the course of the training sessions, subjects with lesser experience on the procedure would show the largest improvement in optimizing this trade-off, as compared with participants who had already established these visuospatial skills in the context of the overarching clinical tasks and the low-level visual-motor subtasks.

5.4. Results

When analyzing the whole group (14 surgical residents), the overall amount of X-ray used decreased with training (83 ms to 52 ms), independently of the difficulty of the scenario. The overall time for the whole procedure also decreased for all participants, however, that measure was more variable, and influenced by the difficulty of the scenario. There is however a trend for all participants to perform faster in the last trials than at the beginning. There was a negative correlation between overall amount of X-ray used across procedure and the participant's performance score for the Santa Barbara Test (speed*accuracy): -0.45; while there was a negative correlation between the average duration of each trial and the participant's performance score of the Santa Barbara Test: -0.65. The same results were obtained when analyzing the mental rotation test with the subjects performing poorly on the simulator having poor visuospatial performance (i.e. participants 5 and 7 were poor in visuospatial ability and were slow and used lots of X-rays, whereas participant 14 and 17 had good visuospatial scores; and were fast and minimized the use of X-rays).

ID	hit time	reject time	hit	miss	FALSE	reject	F1	acc
1	6118.4	7073.5	13	0	0	14	1.00	1.00
12	15082.2	18932.2	13	0	0	14	1.00	1.00
16	16647.4	21754.6	13	0	0	14	1.00	1.00
7	9716.9	9145.7	13	0	1	13	0.96	0.96
6	5919.1	7683.4	12	1	0	14	0.96	0.96
18	5439.7	6263.7	11	2	1	13	0.88	0.89
17	8073.2	8163.0	13	0	4	10	0.87	0.85
3	5480.3	4548.8	12	1	3	11	0.86	0.85
8	7365.5	9984.2	11	2	2	12	0.85	0.85
4	9899.9	10699.0	10	3	1	13	0.83	0.85
5	5713.5	5568.0	13	0	7	7	0.79	0.74
14	2170.6	2517.0	9	4	2	12	0.75	0.78
9	3613.9	5408.3	8	5	2	12	0.70	0.74
2	5334.8	3104.2	8	5	3	11	0.67	0.70

Table 5.1: Standard Contingency Table Analysis of Participant visuospatial abilities on Shepard and Metzler 3D blocks matching task, sorted by their F1 scores.

From Table 5.1, we can clarify the quantitative results by including a few descriptive observations. Subjects 12 and 16 are ”very slow and very accurate” on both the Blocks test and SB test. Subjects 1, 6, 7, 17, 18 are ”faster with good accuracy” on Blocks reasoning, but only subject 18 scored well on SB. The other four were poor on SB. Subject 3 is reasonably fast with reasonable accuracy on both tests of spatial reasoning. Subject 14 is very fast with poor accuracy (perhaps guessing often) while being the best on SB. Subject 5 is not so fast, and poor accuracy (lowest performer). These data will support the following main comparisons: to test the correlation between visuospatial ability, speed and accuracy for subjects 1, 3, 7, 16 to perform their final targeting phase in the first and last scenario in their simulator training.

When we examine the participant performance on the navigation phase on the simulator, and compare to their visuospatial reasoning abilities, we confirm the correlation between visuospatial reasoning abilities and the performance in the first trials on the simulator. The participants with higher visuospatial reasoning abilities perform better on the first trials run on the simulator. This is demonstrated most clearly by considering the cases in which the randomized presentation of data sets was presented as the common first trial. In other words, for the subjects who were tested on dataset ’study12’ as their first trial after the practice run, the navigation phases of the trajectories can be isolated. In Fig. 39 we observe the (x,y,z) trajectories over time. The speed and accuracy of these targeting phases are summarized in Table 5.2.

	Navigation time (sec)	Navigation error (mm)	Spatial task accuracy	Spatial task time hit:miss
Participant 1	104	2.49	100	6.1:7.1
Participant 3	172	5.35	85	5.5:6.3
Participant 7	182	9.68	96	9.7:9.1
Participant 16	243	8.64	100	16.6:21.7

Table 5.2: Speed and accuracy of a representative block: the four participants whose randomized trials involved presentation of dataset ’study12’ as their first scenario viewed after the common initial practice trial.

Participant 1 has the highest accuracy on the simulator, and the highest accuracy in the spatial reasoning tasks, and is the fastest on the vertebroplasty navigation while fast in spatial reasoning.

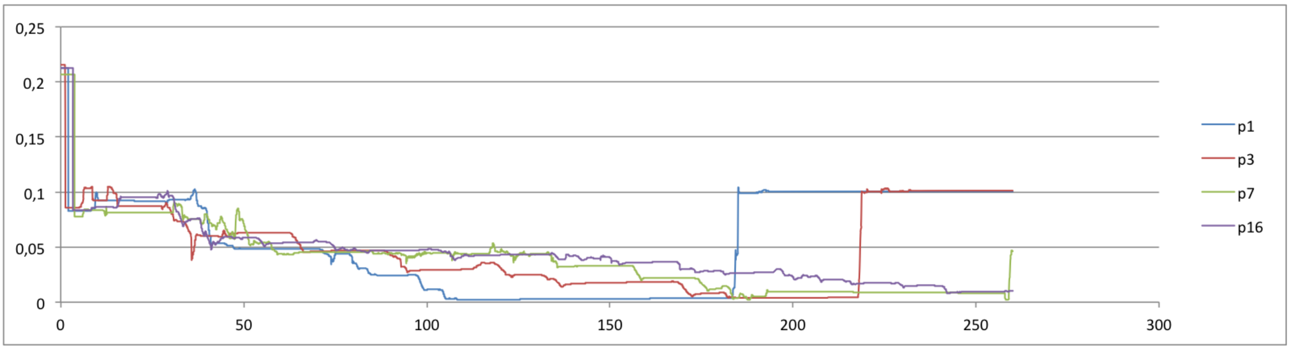
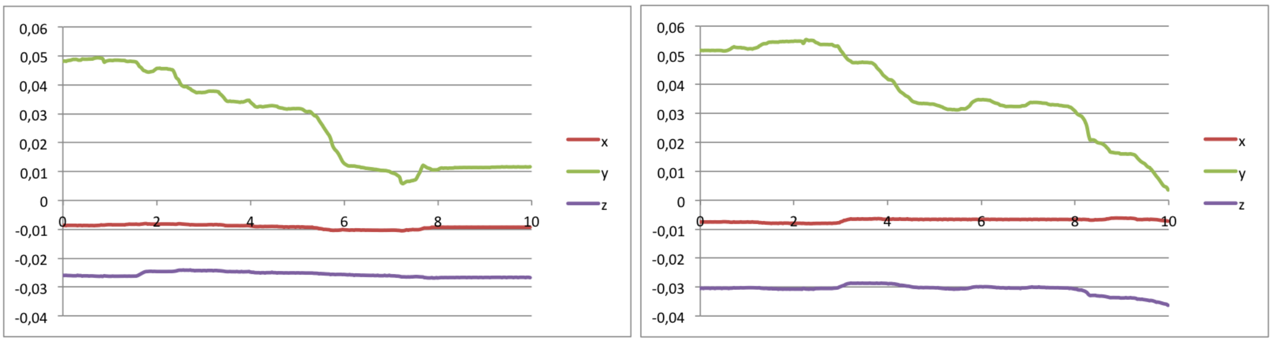


Fig. 39: Distance from Target versus time in seconds in navigation phase for subjects 1, 3, 7, 16 while running study 12 dataset as their first trial after practice session.

Speed and accuracy of a representative block: the four participants whose randomized trials involved presentation of dataset 12 as their first scenario viewed after the common initial practice trial. Participant 16 is the slowest on the vertebroplasty navigation phase and slowest in spatial reasoning tasks, with highest accuracy in spatial reasoning and median simulator accuracy. Participant 3 is more accurate than participant 7 while performing vertebroplasty navigation at the same speed, while participant 3 is less accurate than participant 7 while performing spatial reasoning at a faster speed. These two subjects have chosen 'conservative' versus 'liberal' response strategies at different operating points on the clinical task simulator as compared to the non-clinical task. After using the simulator for 4 training sessions, the final scenarios exhibit responses from each of the trainees that have improved in performance for both speed and accuracy of the navigation and targeting phases. Two plots from typical subjects P1 (see Fig. 40a) and P7 (see Fig. 40b) are included here to illustrate the decrease in time and convergence to the target. Fig. 40 shows improved performance after four training scenarios.



(a) Fifth trial of subject P1 exhibits improved speed and accuracy.

(b) Fifth trial of subject P7 exhibits improved speed and accuracy.

Fig. 40: Convergence to target within a 10 second time window is illustrated in plots.

5.5. Conclusion

The vertebroplasty simulator provides capabilities and functionality to allow subjects to practice realistic scenarios, but from an evaluative perspective, the simulator allows each interaction to be logged – and thereby we may analyse these correlations between pre-test aptitudes, abilities, and skill level, and correlate with performance across learning trials in the clinical domain. Our correlations provide strong evidence of criterion validity for the metrics derived from the simulator and logfiles. We were able to quantify the training advantage that accrues to residents with higher visuospatial reasoning skills, in terms of their improved performance and rate of skill acquisition, as well as to quantify the performance improvements incrementally with each set of simulator scenario blocks.

5.6. Limitations

During the study technical issues occured therefore a restart of the scenarios was sometimes necessary. Further the sample size of the study is small. Additionally a few participants received comments of experts or colleagues who were passing by the simulation stations hence they got distracted.

6. Summary

A new era of surgical education has begun. Today medical education is facing the challenge of objectively assessing competence of technical, non-technical skills and knowledge for ensuring high surgical competence and providing high patienty safety. Hence the integration of simulation techniques is progressing and the need for validated assessment and training methods is increasing with the aim to standardize the medical education. The main drivers for this disruptive change are the (i) ethical concerns about the learning of a surgical craft on patients, the (ii) working hour restrictions, (iii) increased application of minimal-invasive techniques with a steep learning curve, (iv) a significant increase in aging population worldwide, and (v) financial constraints. The training paradigm "proficiency-based progression" proposed by Gallagher et al. in 2012 addresses these circumstances and claims a paradigm shift in surgical education from the traditional surgical curriculum towards a supportive training of surgical skills. First technical skills namely the craftsmanship e.g. usage of instruments, procedural knowledge e.g. workflow and cognitive skills e.g. interpreting medical imaging are the learning goals. These are acquired during practice on a variety of simulation models ranging from bench-top models e.g. suture pads, VR emulators/simulators, full-scale VR simulators, human cadavers to living animals e.g. pigs. Surgical skills are of a twofold nature. Apart from the technical and procedural skills it is a fact that failures in non-technical skill are often root causes for harm and near-misses in the operating room. E.g. communication and decision making should be learning objectives as they are essential for effective and efficient team work in the OR. Thus novel approaches for multidisciplinary team training are key to ensure surgical excellence successfully. Full-scale simulated environments for training seem to perfectly bridge the gap between the necessary skill acquisition of technical and non-technical skill forming future surgeons and therefore enable surgical excellence. In this thesis a simulated full-scale immersive OR theatre is proposed and investigated for conducting assessment and training of surgical skills within a multi-disciplinary fully controlled environment. To this end, a surgical VR simulator is developed and validated for team training and a first assessment framework for measuring technical and non-technical performance was applied with appropriate bench marks and criteria for competency based performance assessment.

Chaper 3 We introduced a novel mixed reality simulation environment consisting of a computerized mannequin coupled with a surgical VR simulator in a realistic setting. Key aspects were investigated and expert recommendations were considered while designing a surgical training environment such as task and crisis analysis of the surgical workflow and simulation realism. Real surgical instruments that are augmented with realistic haptic feedback and VR capabilities, (ii) human sensory channels such as tactile, auditory and visual in real-time, and (iii) the ability to facilitate deliberate exposure to adverse events enabling mediation of error recovery strategies were integrated in the simulation environment. Surgeons were immersed in the medical simulation environment through task and crisis scenarios of a typical vertebroplasty workflow. In the user study the face validity of the simulation environment was confirmed by investigating surgeon behavior and workflow response. The result of the conducted user-study corroborated the unique medical simulation concept of combining VR and human multisensory responses into surgical workflow.

Chapter 4 We proposed a first step towards a robust assessment framework. A unique simulator-based methodology was presented for assessing both technical and non-technical (cognitive) skills for surgical trainees while immersed in a complete medical simulation environment. Further, two crisis scenarios were included which allowed for the evaluation of the effect of cognitive strategy selection on the low-level surgical skills. Training these mixed-mode scenarios can thereby be evaluated on the platform, allowing for improved assessment and a stronger foundation for credentialing, with the potential to reduce the occurrence of adverse events in the OR. Scientific evaluation and validation of the work was conducted together with 19 junior surgeons in order to achieve the following goals: (i) to provide a qualitative measure of usability, (ii) to assess vertebroplasty technical performance of the surgeon, and (iii) to explore the relationship between mental workload and surgical performance during crisis. The results indicated that: (i) the surgeons scored the face validity of our modeled simulation environment very highly, (ii) surgeon training enabled completion of tasks more quickly, and (iii) the introduction of crisis scenarios negatively affected the surgeons' objective performance. Taken together, the results underscored the need to develop realistic simulation environments that prepare young residents to respond to emergent events in the operating room.

Chapter 5 The visuospatial reasoning of surgeons was assessed using the virtual reality surgical simulator component of the multidisciplinary team training environment. The users performed a vertebroplasty procedure under simulated X-ray control while manipulating the simulated mobile C-arm, making use of a simulated needle that is tracked using a Novint Falcon haptic device. A hierarchical task analysis of the procedure was performed and thereby well-posed metrics of performance for each phase of the procedure could be formulated. A comprehensive study involving 14 surgical residents was conducted under the direction of three osteosurgical experts. The speed and accuracy for each phase of the procedure was established and the tradeoff faced by each trainee in terms of the dual competing task they faced, which was to minimize the amount of X-ray dose, was analyzed. Lastly, the training advantage that accrues to surgeons with higher spatial reasoning skills, in terms of their improved performance and rate of skill acquisition was quantified.

Overall, the outcome of this thesis can be summarized as follows:

- For the first time a VR surgical procedural simulator and computerized mannequin are fully integrated and function as one simulation setup. It provides a novel setup for medical assessment and training.
- A surgical simulator reflecting complete workflows of surgical procedures, including intra-operative crisis scenarios which potentially result in adverse events with the capability to deliberately expose the trainee to realistic adverse events.
- First step towards a robust assessment framework consisting of objective performance metrics for the assessments of surgeons' and surgical team's technical and non-technical skills during simulation-based training that enhance impact on the acquisition, application, and retention of teamwork skills in health-care.
- Extension of the PBP curriculum for surgical education.

The first steps have been made to create an assessment construct for surgical skill consisting of the measurement of technical and non-technical performance within a simulated OR theatre. Those findings will enable other national/international research and simulation centers to develop more sophisticated training environments and the hope exists that this work will trigger a new generation of simulators which will eventually improve medical education of multidisciplinary teams and hence improve patient safety and raise the quality of patient care. "The potential of team training is only starting to be realized and its cooperation into surgical training is sure to grow as evidence and reliability and validity of the training and assessment of CRM are further established." [27]

6.1. Future work

6.1.1. Technical development

The next steps for improving the technical setup are (i) a real-time CT simulation with CT-typical streak artifacts, (ii) the development of realistic computational models for instrument-tissue interaction, (iii) the development of a control interface for procedure and scenario management and (iv) the integration of instruments e.g. automatic detection. During this development it is necessary (i) to avoid provoking "a false sense of security" due to oversimplification, (ii) the isolated learning of technical skills [14] and (iii) to provide direct feedback about procedural errors to the surgical trainee since it is key for successful training. E.g. the trainee can touch critical structures without proximate formative feedback resulting in an "unwarranted self-belief in their skills" [46].

6.1.2. Research studies

Future studies should involve investigating the entire surgical workflow, conducting longitudinal training for both technical and non-technical skills - individually and combined, and studying the construct validity of the VR simulator. It should be able to distinguish whether a novice or an expert is performing the procedure based on a performance metric. Furthermore serial tests of lab and clinical performance should be performed with the aim to reveal cause-effect relationships between technical and non-technical parameters though keeping in mind that no single parameter measured can by itself demonstrate proficiency. Then a custom-made curriculum can be designed which addresses the individual grades of surgical expertise according to the Dreyfuss model (see Fig. 20 and Table 2.7).
Moreover studies should consist in multicenter trials with standardised approaches and with sufficient participants. The skills being evaluated should be part of a standard surgical skills training course, not just stand-alone technical skills. Additionally, once efficacy has been determined cost-benefit analyses could be attempted. [130]. Finally in order to form a more robust and reliable assessment construct in the next studies existing assessment methods should be comprised to measure team performance. In particular, further investigations should be conducted on the non-technical assessment methods [62]. In Chapter 4 the SURG-TLX was used for showing the impact of surgical load on surgical performance. Other assessment methods exist and could be used and even validated in the simulated OR theatre such as e.g. the NOn-Technical Skills for Surgeons (NOTSS).

NOn-Technical Skills for Surgeons (NOTSS) The NOn-Technical Skills for Surgeons (NOTSS) was developed by Yule et al. using several methods of task analysis with consultant surgeons [150]. Yule et al. revised it in 2008 and removed the tasks managment category from the initial five categories of the non-technical skill set: Situation Awareness, Decision Making, Communication and Teamwork, and Leadership (see Table 6.1). With subject matter experts the content validity of the NOTSS system was derived.

Category	Element
Situation Awareness	Gathering information
	Understanding information
	Projecting and anticipating future state
Decision Making	Considering options
	Selecting and communicating option
	Implementing and reviewing decisions
Communication and Teamwork	Exchanging information
	Establishing a shared understanding
	Co-ordinating team
Leadership	Setting and maintaining standards
	Supporting others
	Coping with pressure

Table 6.1: NOn-Technical Skills for Surgeons (NOTSS) skills taxonomy v1.2 by Yule in 2008

Further assessment methods which could be applied and tested are the observational Teamwork Assessment for Surgery (OTAS) [142], the Anesthetists' Non-Technical Skills (ANTS) [151], the Oxford Non-Technical Skills (NOTECHS) [94], the Revised NOTECHS [11], the Line Operations Safety Audit Checklist (LOSA; selected elements) [95], the State Trait Anxiety Inventory (STAI) [142], the Imperial Stress Assessment Tool (ISAT) [6], the SVF78 Stress-Coping Questionnaire [56], the Communication-based Objective Structured Clinical Examination (OSCE) [76], the Utterance frequency [95] and the Behrenz Fatigue Questionnaire [66].

6.2. Open research questions

The purpose of the proposed simulated immersive OR theatre is to facilitate investigation of key research topics recommended by the Association for Surgical Education (ASE) Simulation Committee [65].

- What are the best methods to assess technical and non-technical performance on simulators?
- What are the performance criteria that surgical trainees need to achieve to be competent/proficient based on their training level?
- How do we train and assess teams effectively using simulation?
- How can we use simulation to teach and assess judgement and decision making in routine and crises situations?
- Role in surgical safety?
- How many hours or number of procedures are required to train a surgeon? [9]
- Investigation of most effective and efficient ways (regarding nature and duration of training) in which simulation-based training can be integrated into the surgical curriculum.
- Optimal stage of training at which trainees receive maximum skill transfer benefits from different forms of simulation.
- Effect of mentoring intensity during the training period.

6.3. Outlook

The ultimate goals are (i) increasing patient saftey and (ii) improving patient outcome. With a validated team-based training environment including appropriate performance metrics which are assessed with robust assessment methods and tools, a certification of the OR team members can be conducted which means nothing less than the standardization of surgical competence. This is the basis for ensuring excellence in surgical education. Moreover novel technology such as robots could be evaluated in the proposed fully controlled environment. The challenge is the finding of factors which can be influenced positively in respect to quality of patient outcome, safety and work load of medical personnel and economical factors. These have to be addressed during the development of novel surgical techniques and OR technologies. Furthermore the introduction of all kinds of innovations into the OR can then be accompanied with an appropriate training strategy/concept. In future it is reasonable to establish a national/international society controlling the admission of simulators for training ensuring standards of medical simulations to guarantee high quality training.

Overview of key opportunities using a simulated multidisciplinary team training environment

- Standardization of surgical competence.
- Quality control based on standardized surgical performance metrics measuring surgical competence.
- Selection in early career stages.
- Standards: Certification and retention of skills.
- Usability testing of medical devices before deployment in medical care units and processes.

6.3.1. Barriers for simulation-based training of surgeons

In the last century the training model consisted of teaching scientific principles and the knowledge transfer during surgeries. In the past decade the topics *evidence-based medicine, patient safety* and *practice efficiencies* became relevant within the society. These are tightly coupled with medical education [152]. Today medical education needs to develop competence of cognitive, technical, non-technical skills and knowledge of the trainees and furthermore to objectively assess their performance.
Simulation mimicking the real scenario provides a suitable environment for objective performance assessment and for deliberate practice, which is the key for the acquisition and retention of skills. Simulation itself is not a novel approach for deliberate practice in medicine since it has been in use in primitive forms for centuries. Anatomical models were created long before advanced synthetic models or computers were available. With these novel techniques including computer simulated virtual environments new ways of training are made possible however it takes time to understand, accept and integrate them. In particular three major reasons for the slow advancement are mentioned by [117] "skepticism, lack of communication, and the burden of proof". Moreover appropriate curricula incorporating simulation techniques for ensuring surgical excellence do not exist and resources are missing in terms of qualified staff for education [153]. According to [98] "Identifying qualified faculty skilled in simulation use and debriefing is another significant barrier. This includes availability in terms of protected time as well as those faculty trained to teach. This technology does not obviate the need for faculty trained in solid educational principles and teaching techniques. In other words, simulators do not replace good educators" [98]. Further, the expenses for running simulation centers comprise simulators ($6000-$250000), equipment maintenance, space for educational labs, and personnel are high [98]. Another barrier mentioned by [98] are faculty time constraints. Additionally validity and fidelity are still hindering a break through in integration into medical curricula [117]. Finally Cumin et al. highlighted in 2013 the key challenges of conducting simulations for multidisciplinary team training. They mainly consist of (i) the recruitment of subjects with various experience levels, (ii) model realism is highly dependent on training purpose and not well determined yet and (iii) the financial costs of simulations.
Until now only a few studies were conducted to validate VR simulators [121] [26] [47] mainly comprising endoscopic and laparascopic procedures. The challenging part of the validation is that many factors have the potential to influence the transfer of skills, including those relating to simulator design and functionality and the way that it is used as a training device. Moreover, prelearning, the nature and type of formative and summative feedback, and opportunities for reinforcement of learning have an impact. Consequently, the evidence for skill transference reported in studies has to be taken with care according to [26]. Further [121] claims

that "validity research is hampered by a paucity of widely accepted definitions and measurement methods of validity". Additionally the results of these studies are based only on a (i) small number of participants mainly surgical residents, normally the study was conducted within (ii) only one institution and the (iii) surgical procedure was low-risk (laparoscopic cholecystectomy) [47]. Most studies exhibit a drug trial-like design consisting of randomization of subjects to VR training and control group. The randomization depended on the number of conducted procedures to define the proficiency level for the seperation of studied subset groups. However this approach is not being meaningful in terms of treatment quality since it can only serve as one single indicator of proficiency out of many. Reviews of available reports on skills transfer from VR training to real-world performance criticise the lack of accepted norms for control groups. Meaning that a required baseline does not exist. For example a control group which had no training most probably performs differently than a control group which had traditional training. As mentioned previously the proficiency level of subjects related to the number of performed treatments is used as basis for the distinction into to study groups. Furthermore the training conditions differ widely e.g. intensity and duration of mentoring given to participants [26]. Moreover the different complexity levels of the assessed procedures prevent a valid comparison of the conducted studies. Other key flaws mentioned are "ill-defined parameters measured during the assessment" and "usage of the tool for training and in a later stage for performance assessment allowing the trainee to gain familiarity with the testing procedure or device". Finally [26] critized that "assessors were not blinded to the training status of participants and there may have been assessor bias."

The next steps for decreasing the barriers in order to include simulation-based training into the medical curriculum are the conduction of additional studies on training transfer of more procedures with robust metrics, decrease of costs of simulation techniques, establishment of key opinion leaders in politics and clinics and finally more money.

Studies like [47] illustrate the training transfer of VR simulators to real-world tasks for enhancing the real-world performance are valuable. In two studies the transfer of training (ToT) and the training effectiveness ratio (TER) with two subject groups, (a) 195 experienced surgeons (MIS > 50) and (b) 30 novices (MIS $= 30$) was measured. In both studies the subjects were distributed to two groups, a VR training group (condition 1) and a control group (condition 2). Both groups performed a laparoscopic task mimicking a real-world task on a box-trainer at the end. In the first study with (a) the subset VR-trained group made significantly more correct incisions. In study 2 with (b) the group trained on the VR simulator needed significantly less time and made fewer errors on the real-world task. Then [121] recommended "it would be helpful to those considering the use of simulators in training programs if there were consensus on guidelines for validating surgical simulators and the development of training programs." Further multidisciplinary simulation centers can be created to reduce the costs for the different disciplines by sharing the expenses [98]. Moreover solutions addressing the issues of Cumin et al. 2013 could be the integration of key opinion leaders of the surgical discipline to draw attention to the increased need for simulation-based training and to allow time for surgical trainees to practice. Arguments for these leaders could be (i) a clear statement on benefits and relative cost of surgical simulation-based would result from consistency in training and assessment methods across studies and secondly (ii) the return on investment: Money saved in patient care and legal proceedings vs. investment in surgical education. Finally [4] requests: "it is absolutely the time for physicians, hospital managers, policy makers, patients, and the public alike to demand resources..." in order to get resources for creating and running medical curricula with simulation techniques. Indicators for an increasing acceptence already exist since the acceptance of the use of novel simulation techniques is growing [117].

A List of abbreviations

ED Emergency department

GYN Gynecology

ICSAD Imperial College Surgical Assessment Device

ICU Intensive care unit

IPPI Integrated procedural performance instrument

MIS Minimal invasive surgery

MISTELS McGill Inanimate System for Training and Evaluation of Laparoscopic Skills

OB Obstetrics

OR Operating room

OSATS Objective Structured Assessment of Technical Skills

OSCE Objective structured clinical examination

OTAS Observational Teamwork Assessment for Surgery

PBP Proficiency-based progression

RCT Randomised controlled trial

SCIM Structured clinical instructional modules

VR Virtual reality

B Glossary

Definitions are mainly adopted from [46] and [9].

Assessment Assessment can have different and multiple purposes, including determining a level of competence, aiding learning through constructive feedback, measuring progress over time or certifying competence. Assessments can be categorised as for or of learning, although there is a continuum between these two poles.

Assessment for learning Is primarily aimed at aiding learning through constructive feedback that identifies areas for development. Alternative terms are formative or low-stakes assessment. Lower reliability is acceptable for individual assessments as they can and should be repeated frequently. This increases their reliability and helps to document progress. Such assessments are ideally undertaken in the workplace.

Assessment of learning Is primarily aimed at determining a level of competence to permit progression of training or certification. Such assessments are undertaken infrequently (e.g. examinations) and must have high reliability as they often form the basis of pass/fail decisions. Alternative terms are summative or high-stakes assessment.

Assessment system An assessment system (or assessment programme) is designed to ensure that trainees learn the knowledge, skills, judgement and professional behaviours required by a training syllabus. The combination of an assessment system and a syllabus are the key components that specifically address assessment practice within a curriculum. Contemporary best practice favours assessment systems that are multifaceted and assess an appropriate spectrum of a syllabus in a reliable way. This is done through a blueprint.

Certification The process by which governmental, non-governmental or professional organisations or other statutory bodies grant recognition to a trainee who has met certain predetermined standards specified by the organisation and who voluntarily seeks such recognition.

Competence A trainee's ability to perform a particular activity to the required standard (i.e. that required for patient safety), while being observed in the workplace or in a controlled representation of the workplace (e.g. in simulation). Competence comes from experience combined with constructive feedback and reflective practice (self-assessment/insight).
Competence is a prerequisite for satisfactory performance in real life, although many doctors progress to a higher level of excellence during their career. A competent doctor may perform poorly for many reasons including tiredness, stress, illness or a lack of resources. Competencies A set of abilities that includes knowledge, skills, judgement and professional behaviours.

Construct A construct is an attribute, proficiency, ability or skill that exists in theory and has been observed to exist in practice, such as 'surgical skill'. Constructs are vital within assessment theory as they provide the underpinning framework for establishing assessment design and validity.

Curriculum A curriculum is a statement of the aims and intended learning outcomes of an educational programme. It states the rationale, content, organisation, processes and methods of teaching, learning, assessment, supervision and feedback. If appropriate, it will also stipulate the entry criteria and duration of the programme.

Formative assessment See Assessment of learning.

High-stakes assessment See Assessment of learning.

Learning outcomes The competencies to be acquired by the end of a period of training.

Low-stakes assessment See Assessment for learning.

Performance The application of competence in real life. In the case of medicine, it denotes what a trainee actually does in his or her encounters with patients, their relatives and carers, colleagues, team members, other members of staff, etc. Performance is not the same as knowing or being able to do everything. On the contrary, it may well be about knowing what you do not or even cannot know - in other words, knowing your own limitations.

Reliability Expresses a trust in the accuracy or provision of the correct results. In the case of assessments, it is an expression of precision and discrimination. There are several important dimensions of reliability. These include:

- Equivalence or alternate-form reliability is the degree to which alternate forms of the kind of assessment produce congruent results.
- Homogeneity is the extent to which various items in an assessment legitimately link together to measure a single characteristic.
- Inter-rater reliability refers to the extent to which different assessors give similar ratings for similar performances.
- Intra-rater reliability is concerned with the extent to which a single assessor would give similar marks for almost identical performance.

Skill The ability to perform a task to at least a competent level. A skill is best (most efficiently) gained through regular practice (experience) combined with reflective practice (self assessment/ insight) and constructive feedback.

Standards In medical education standards may be defined as 'a model design or formulation related to different aspects of medical education and presented in such way to make possible assessment of graduates' performance in compliance with generally accepted professional requirements'. Thus, a standard is both a goal (what should be done) and a measure of progress towards that goal (how well it was done).

Summative assessment See Assessment of learning.

Trainee Any doctor participating in an educationally approved postgraduate medical training programme (foundation or specialty).

Trainer A senior doctor who provides educational support for a more junior doctor (trainee). Trainers include clinical and educational supervisors. All trainers require training in teaching and assessment methods, including giving constructive feedback. Educational supervisors require additional training in appraisal and career guidance.

Training The ongoing, workplace-based process by which experience is obtained, constructive feedback provided and key competencies achieved.

C Appendix

C1. Overview of multidisciplinary team-based simulation approaches

The list was derived from [23] and [133] including in-situ and simulated OR environments excluding wet labs.

Year	Group	Author	Country
1996	University of Basel	Helmreich	Switzerland
1998	University of Basel	Sexton	Switzerland
2004	Imperial College London	Aggarwal	UK
2005	Imperial College London	Moorthy	UK
2006	Imperial College London	Moorthy	UK
2007	Louisiana State University	Paige	USA
2007	Southern Health Simulation and Skills Centre	Flanagan	Australia
2007	Imperial College	Undre	UK
2008	Imperial College	Koutantji	UK
2008	Havard Medical School	Powers	USA
2008	Louisiana State University	Kozmenko	USA
2009	Louisiana State University	Paige	USA
2009	Oklahoma University Medical Center	Forsythe	USA
2010	Memorial University of Newfoundland	O'Regan	Canada
2011	Children's Hospital of Boston	Volk	USA
2011	Havard Medical School	Ziewacz	USA
2012	University Hospital of Montreal	Stevens	USA
2012	St. Michael's Hospital, Toronto	Lee	Canada
2012	University of Pennsylvania	Acero	USA

Table C1: Adapted from Cumin 2012 and Tan 2014 depicting all the studies which were conducted within the timeframe 1996-2012. Excluded is the type of simulation "wet laboratories".

Author	Study aim	# participants	Assessment	Results
Helmreich	Assess participant perceptions of OR-MDT training course	not specified	Post-course questionnaire	Participants assessed the training favourably
Sexton +Musson2004?	Assess participant of OR-MDT training course	291?	Post-course questionnaire	Participants assessed the training favourably
Aggarwal	Assess participants perceptions of OR-MDT training course	25 surgeons and one surgical team	Post-course questionnaire; Video-based, blinded assessment of technical skills; non-technical skills measured by two expert observers on global rating scale	
Moorthy	Proof of face validity; Assessment of surgical technical and non-technical skills	27 surgical trainees	ICEPS for procedure-specific technical skill; Post-course questionnaire for face validity; modified LOSA checklist; Utterance Frequency (UF) for communication between surgical team	90% of subjects felt immersed; 50% considered synthetic model to be a realistic representation
Moorthy	Proof of face validity; Assessment of technical and non-technical skills	20 surgeons plus standarized theater team	Post-course questionnaire; Procedure-specific rating scale (5 vascular surgeons assessed); NOTECHS [95]	Participants assessed the trainign favourably; Construct validity of bleeding crisis;
Paige	Assess participant perceptions of OR-MDT training course	10	Post-course questionnaire	Participants assessed the training favourably
Flanagan	Improve attitudes toward safety and teamwork	59	Participants assessed the training favourably	
Undre	Create an OR-MDT course and explore nontechnical skills across professions	80	Post- Course questionnaire	Participants assessed the training favourably
Koutantji	Promote understanding of teamwork and safety	34	Post-course questionnaires	Participants assessed favorably simulation scenarios as a training method

Table C2: Continued: Adapted from Cumin 2012 and Tan 2014 depicting all the studies which were conducted within the timeframe 1996-2012. Excluded is the type of simulation "wet laboratories".

Author	Study aim	# participants	Assessment	Results
Powers	Assess participant perceptions of an OR-MDT training scenario	20	Rating scales on technical and non-technical performance; Objective outcome measures	Face and construct validity; Technical and non-technical performance discrimination is observed between novices and experts
Kozmenko	Assess participant perceptions of an OR-MDT training course	39	Team performances were assessed by both direct observation and team's self-assessment where each team member assessed his or her own performance as well as the performance of all other team members (*360 degree assessment*)	?
Paige	Assess participant perceptisons of an OR-MDT training course	56	Pre- and post-course questionnaires on participant's self-efficacy for performing targeted teamwork competencies	Improvement in self-reporting efficacy after module 1 and further improvement in self-efficacy were confirmed by direct observation
Forsythe	Improve team communication	62	Pparticipants' narratives in the form of collaborative recall and reflection to standardize task, process, and language	Participants identified and changed multiple tasks, process, and language item. Positive changes for task and efficiencies, team interactions, and overall functionality of the team
O'Regan	Identify latent weaknesses in Malignant hyperthermia management processes	6	Post-course questionnaire	Participants assessed the training favourably.
Volk	Teach teamwork and crisis resource management skills	59	Post-course questionnaire	Participants assessed the training favourably.
Ziewacz	Improve adherence to critical management steps	11	Post-course questionnaire	Participants assessed the training favourably.

Table C3: Continued: Adapted from Cumin 2012 and Tan 2014 depicting all the studies which were conducted within the timeframe 1996-2012. Excluded is the type of simulation "wet laboratories".

Author	Study aim	# participants	Assessment	Results
Stevens	Assess participant perceptions of an OR-MDT training course	27	Pre- and post-course (6 months after trainig) questionnaires (operating room teamwork survey) Interviews	Participants assessed the training favourably. Only one aspect of team communication improved between the two surveys, that is, improved teamwork concepts. Perceived relevance to current practice and feasible to change teamwork behaviour in cardiac surgery.
Lee	Evaluation of scenario and assessment of technical and nontechnical performance of urology and anesthesiology residents	16	Post-course questionnaire; Technical skills: (1) Task-specific checklist. (2) 5-point Likert style Global rating scale. (3) Amount of blood loss. Non-technical skills: (1) NOTSS (2) ANTS	Urology and anaesthetist resident level of training did not correlate with non-technical performance rated by expert faculty. Participants assessed the training favourably. Urology junior technical performance was significantly worse than senior resident performance. Non-technical performance self-assessment by urology residents was consistently higher than faculty rating.
Acero	Assess participant perceptions of an OR-MDT training scenario	171	Post-course questionnaire; Number of essential emergency steps and the time to complete steps during the simulation; Pre- and post-training knowledge procedural-based questions	Participants assessed the training favourably. All OR teams completed more essentials steps and in less time in 'warm' versus 'cold' scenarios. Participants reported feeling more comfortable with their role in an OR emergency post-training. Increase in correct responses in the knowledge test after training.

Table C4: Continued: Adapted from Cumin 2012 and Tan 2014 depicting all the studies which where conducted within the timeframe 1996-2012. Excluded are the types of simulation "point-of-care" and "wet laboratories".

C2. Technical skills assessment methods

In the following section the most common technical assessment methods are described.

Logbooks Logbooks have been used for a long time as tools for assessment of procedural skills. [134] mentions that 52.3% of Canadian and 47.8% of American case or procedure logbooks are used as a form of assessment in postgraduate programs. It was reported that logbooks were used at least two times more commonly than simulation. Maybe this popularity is due, to the ease with which this form of assessment can be implemented. Though, a logbook is a tool for monitoring the experience bandwidth and for documenting the trainee's progress. It does not ensure direct observation nor verifies the level of performance of the trainee [134].

MISTELS and ICSAD Assessment tools to measure proficiency in procedural skills have been created relying on methods such as the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) and the Imperial College Surgical Assessment Device (ICSAD) [21,22].
The MISTELS is designed to assess technical steps in laparoscopic surgery while the ICSAD has been specifically used for hand motion analysis for performance of technical skills. Those are useful for highly technical procedural skills, these tools address training and assessment of surgical skills.

OSCE Another potential tool for training and assessing residents in procedural skills is the objective structured clinical examination (OSCE). The OSCE has demonstrated validity and reliability in multiple settings [23]. Using multiple stations in an OSCE format was the framework used for the development of the Objective Structured Assessment of Procedural Skills (OSATS) [24,25].

OSATS The OSATS is used to assess technical skills through both a procedure-specific checklist and a global rating scale of operative performance. Multiple studies have demonstrated high internal consistency and inter-rater reliability of the OSATS in laboratory multi-station settings and in the operating room [26]. A similar format was also shown to be useful for the assessment of minor surgical skills for clinical clerks and for family medicine residents [27,28].

SCIM Structured clinical instructional modules (SCIM) were developed as teaching OSCEs to compensate for difficulty in accessing relevant clinical experience and to provide opportunities to learn about clinical situations that are infrequently encountered by trainees [29,30].

IPPI Additionally, the integrated procedural performance instrument (IPPI) was developed to assess a candidate's ability to not only demonstrate the technical aspect of a procedural skill, but also the non-technical aspects such as communication, collaboration and professionalism [31]. These stations are logistically more complex and may require more time if included in a more traditional OSCE. Although OSCEs and IPPIs provide opportunities for training and assessment they are expensive and labour intensive. As this brief review of the literature reveals, research has clearly established that laboratory-based simulation and OSCE frameworks hold great promise as settings for the teaching and assessment of procedural skills. However, while considerable critical attention has focused on evaluating assessment tools, instructional modules, and different teaching and assessment settings, a similar depth of inquiry has yet to delve into the experiences of the trainees who actually face the challenge acquiring procedural skills. Indeed, when it comes to the teaching and assessment of procedural skills, research has carefully examined the delivery side of the question, but has paid less heed to the recipient side. Specifically, little is known about Medicine residents' individual experiences of 1) learning and acquiring procedural skills proficiency, 2) of practicing these techniques, or 3) of being assessed on their proficiency.

C3. Stakeholder of standardization and certification

The development and validation of assessment methods and their use are aiming at four different stakeholders according to [9].

- For the trainee
  - Provide feedback about strengths and weaknesses to guide future learning
  - Foster habits of self-reflection and self-remediation
- For the curriculum
  - Respond to lack of demonstrated competence (targeted training)
  - Certify progression in training over time
  - Certify achievement of curricular outcomes
  - Foster curricular change
  - Create curricular coherence
  - Cross-validate other methods of assessment in the curriculum
  - Establish standards of competence for trainees at different levels
- For the institution
  - Discriminate among trainees for progresion in training or access to subspecialty training
  - Guide a process of institutional self-reflection and self-remediation
  - Develop shared educational values among a diverse community of educators
  - Promote faculty development
  - Provide data for educational research
- For the public
  - Certify competence of doctors in training
  - Identify unsafe or poorly performing doctors

List of Figures

List of Tables

Bibliography

[1] Acero, Natalia Martinez, Motuk, Gregory, Luba, Josef, Murphy, Michael, McKelvey, Susan, Kolb, Gretchen, Dumon, Kristoffel R and Resnick, Andrew S [2012]: Managing a surgical exsanguination emergency in the operating room through simulation: an interdisciplinary approach, Journal of surgical education **69**(6): 759–765.

[2] Aggarwal, R., Mytton, O. T., Derbrew, M., Hananel, D., Heydenburg, M., Issenberg, B., MacAulay, C., Mancini, M. E., Morimoto, T., Soper, N., Ziv, A. and Reznick, R. [2010]: Training and simulation for patient safety, Quality and Safety in Health Care **19**(Suppl 2): i34–i43.

[3] Aggarwal, R., Undre, S., Moorthy, K., Vincent, C. and Darzi, A. [2004]: The simulated operating theatre: comprehensive training for surgical teams, Quality & Safety in Health Care **13 Suppl 1**: i27–32.

[4] Aggarwal, Rajesh and Darzi, Ara [2011]: Simulation to enhance patient safety: why aren't we there yet?, CHEST Journal **140**(4): 854–858.

[5] Ahlberg, Gunnar, Enochsson, Lars, Gallagher, Anthony G., Hedman, Leif, Hogman, Christian, McClusky, David A., Ramel, Stig, Smith, C. Daniel and Arvidsson, Dag [2007]: Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies, American Journal of Surgery **193**(6): 797–804.

[6] Arora, Sonal, Sevdalis, Nick, Aggarwal, Rajesh, Sirimanna, Pramudith, Darzi, Ara and Kneebone, Roger [2010]: Stress impairs psychomotor performance in novice laparoscopic surgeons, Surgical endoscopy **24**(10): 2588–2593: read.

[7] Atesok, Kivanc, Mabrey, Jay D, Jazrawi, Laith M and Egol, Kenneth A [2012]: Surgical simulation in orthopaedic skills training, Journal of the American Academy of Orthopaedic Surgeons **20**(7): 410–422.

[8] Barry Issenberg, S, Mcgaghie, William C, Petrusa, Emil R, Lee Gordon, David and Scalese, Ross J [2005]: Features and uses of high-fidelity medical simulations that lead to effective learning: a beme systematic review, Medical teacher **27**(1): 10–28.

[9] Beard, Jonathan D, Marriott, J, Purdie, H and Crossley, J [2011]: Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology, Clinical Governance: An International Journal **16**(3).

[10] Bernard Jr, Thomas N and Seibert, Charles E [1992]: Pedicle diameter determined by computed tomography: Its relevance to pedicle screw fixation in the lumbar spine., Spine **17**(6): S160–S163.

[11] Black, SA, Nestel, DF, Kneebone, RL and Wolfe, JHN [2010]: Assessment of surgical competence at carotid endarterectomy under local anaesthesia in a simulated operating theatre, British Journal of Surgery **97**(4): 511–516.

[12] Blum, Tobias, Rieger, Andreas, Navab, Nassir, Friess, Helmut and Martignoni, Marc [2013]: A review of computer-based simulators for ultrasound training., Simulation in healthcare : journal of the Society for Simulation in Healthcare **8**(2): 98–108.

[13] Bott, Oliver Johannes, Dresing, Klaus, Wagner, Markus, Raab, Björn-Werner and Teistler, Michael [2011]: Informatics in radiology: use of a c-arm fluoroscopy simulator to support training in intraoperative radiography, Radiographics **31**(3): E65–E75.

[14] Brindley, Peter G, Jones, Daniel B and Christopher de Gara MB, MS [2012]: Canadian association of university surgeons' annual symposium. surgical simulation: The solution to safe training or a promise unfulfilled? victoria, bc, sept. 10, 2009, Canadian Journal of Surgery **55**(4 Suppl. 2): S200.

[15] Buchholz, Joseph, Vollmer, Charles M, Miyasaka, Kiyoyuki W, Lamarra, Denise and Aggarwal, Rajesh [2015]: Design, development and implementation of a surgical simulation pathway curriculum for biliary disease, Surgical endoscopy **29**(1): 68–76.

[16] Carroll, John B [1993]: Human cognitive abilities: A survey of factor-analytic studies, Cambridge University Press.

[17] Chen, Xin, Wang, Lejing, Fallavollita, Pascal and Navab, Nassir [2013]: Precise x-ray and video overlay for augmented reality fluoroscopy, International journal of computer assisted radiology and surgery **8**(1): 29–38.

[18] Chui, C.-K., Ong, J. S. K., Lian, Z.-Y., Wang, Z., Teo, J., Zhang, J., Yan, C.-H., Ong, S.-H., Wang, S.-C., Wong, H.-K., Teo, C.-L. and Teoh, S.-H. [2006]: Haptics in computer-mediated simulation: Training in vertebroplasty surgery, Simulation & Gaming **37**(4): 438–451.

[19] Cohen, Cheryl A and Hegarty, Mary [2012]: Inferring cross sections of 3d objects: A new spatial thinking test, Learning and Individual Differences **22**(6): 868–874.

[20] Cooper, JB and Taqueti, VR [2004]: A brief history of the development of mannequin simulators for clinical education and training, Quality and Safety in Health Care **13**(suppl 1): i11–i18.

[21] Cosby, Karen S, Roberts, Rebecca, Palivos, Lisa, Ross, Christopher, Schaider, Jeffrey, Sherman, Scott, Nasr, Isam, Couture, Eileen, Lee, Moses, Schabowski, Shari et al. [2008]: Characteristics of patient care management problems identified in emergency department morbidity and mortality investigations during 15 years, Annals of emergency medicine **51**(3): 251–261.

[22] Crossley, Jim, Humphris, Gerry and Jolly, Brian [2002]: Assessing health professionals, Medical education **36**(9): 800–804.

[23] Cumin, David, Boyd, Matt J, Webster, Craig S and Weller, Jennifer M [2013]: A systematic review of simulation for multidisciplinary team training in operating rooms, Simulation in Healthcare **8**(3): 171–179.

[24] Cuschieri, Alfred, Francis, Nader, Crosby, Joy and Hanna, George B [2001]: What do master surgeons think of surgical competence and revalidation?, The American journal of surgery **182**(2): 110–116.

[25] Datta, Vivek, Mackay, Sean, Mandalia, Mirren and Darzi, Ara [2001]: The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model, Journal of the American College of Surgeons **193**(5): 479–485.

[26] Dawe, Susan R., Windsor, John A., Broeders, Joris A.J.L., Cregan, Patrick C., Hewett, Peter J. and Maddern, Guy J. [2014]: A Systematic Review of Surgical Skills Transfer After Simulation-Based Training: Laparoscopic Cholecystectomy and Endoscopy, Annals of Surgery **259**(2): 236–248.

[27] de Montbrun, Sandra L and MacRae, Helen [2012]: Simulation in surgical education, Clinics in colon and rectal surgery **25**(3): 156.

[28] Denaro, Luca, Longo, Umile Giuseppe and Denaro, Vincenzo [2009]: Vertebroplasty and kyphoplasty: reasons for concern?, Orthopedic Clinics of North America **40**(4): 465–471.

[29] Devitt, J Hugh, Kurrek, Matt M, Cohen, Marsha M and Cleave-Hogg, Doreen [2001]: The validity of performance assessments using simulation., Anesthesiology **95**(1): 36–42.

[30] Devitt, J Hugh, Kurrek, Matt M, Cohen, Marsha M, Fish, Kevin, Fish, Pamela, Noel, Alva G and Szalai, John-Paul [1998]: Testing internal consistency and construct validity during evaluation of performance in a patient simulator, Anesthesia & Analgesia **86**(6): 1160–1164.

[31] Eagleson, Roy, de Ribaupierre, Sandrine, King, Sharla and Stroulia, Eleni [2011]: Medical education through virtual worlds: the hlthsim project., Stud Health Technol Inform **163**: 180–184.

[32] Ensrud, Kristine E, Thompson, Desmond E, Cauley, Jane A, Nevitt, Michael C, Kado, Deborah M, Hochberg, Marc C, Santora, Arthur C and Black, Dennis M [2000]: Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass, Journal of the American Geriatrics Society **48**(3): 241–249.

[33] Erat, Okan, Pauly, Olivier, Weidert, Simon, Thaller, Peter, Euler, Ekkehard, Mutschler, Wolf, Navab, Nassir and Fallavollita, Pascal [2013]: How a surgeon becomes superman by visualization of intelligently fused multi-modalities, SPIE medical imaging, International Society for Optics and Photonics, S. 86710L–86710L.

[34] Eubanks, Thomas R, Clements, Ronald H, Pohl, Dieter, Williams, Noel, Schaad, Douglas C, Horgan, Santiago and Pellegrini, Carlos [1999]: An objective scoring system for laparoscopic cholecystectomy, Journal of the American College of Surgeons **189**(6): 566–574.

[35] Fargen, Kyle Michael, Siddiqui, Adnan H, Veznedaroglu, Erol, Turner, Raymond D, Ringer, Andrew J and Mocco, J [2012]: Simulator based angiography education in neurosurgery: results of a pilot educational program, Journal of neurointerventional surgery **4**(6): 438–441.

[36] Farnworth, Lance R, Lemay, David E, Wooldridge, Todd, Mabrey, Jay D, Blaschak, MJ, DeCoster, Thomas A, Wascher, Daniel C and Schenck, RC [2001]: A comparison of operative times in arthroscopic acl reconstruction between orthopaedic faculty and residents: the financial impact of orthopaedic surgical training in the operating room, Iowa Orthopaedic Journal **21**: 31–35.

[37] Flanagan, Brendan, Joseph, Michele, Bujor, Michael and Marshall, Stuart [2007]: Attitudes to safety and teamwork in the operating theatre, and the effects of a program of simulation-based team training, Multimodal Safety Management and Human Factors. Ashgate Publishing Limited, Aldershot, UK S. 211–220.

[38] Flin, Rhona, O'Connor, Paul and Mearns, Kathryn [2002]: Crew resource management: improving team work in high reliability industries, Team Performance Management: An International Journal **8**(3/4): 68–78.

[39] Force Effects [n.d.].
URL: *http://www.h3dapi.org/modules/mediawiki/index.php/Force_effects/*

[40] Force Effects - Surfaces [n.d.].
URL: *http://www.h3dapi.org/modules/mediawiki/index.php/Surfaces*

[41] Forsythe, Lydia [2009]: Action research, simulation, team communication, and bringing the tacit into voice society for simulation in healthcare, Simulation in healthcare **4**(3): 143–148.

[42] Freitag, M., Gottschalk, A., Schuster, M., Wenk, W., Wiesner, L. and Standl, T. G. [2006]: Pulmonary embolism caused by polymethylmethacrylate during percutaneous vertebroplasty in orthopaedic surgery, Acta Anaesthesiologica Scandinavica **50**(2): 248–251.

[43] Fritz, Peter Z, Gray, Tim and Flanagan, Brendan [2008]: Review of mannequin-based high-fidelity simulation in emergency medicine, Emergency Medicine Australasia **20**(1): 1–9.

[44] Gaba, David M [2004]: The future vision of simulation in health care, Quality and safety in Health care **13**(suppl 1): i2–i10.

[45] Gaba, David M and DeAnda, Abe [1988]: A comprehensive anesthesia simulation environment: re-creating the operating room for research and training., Anesthesiology **69**(3): 387–394.

[46] Gallagher, Anthony G. and O'Sullivan, Gerald C. [2012]: Fundamentals of surgical simulation: principles and practices, Improving medical outcome–zero tolerance, Springer, London.

[47] Gallagher, Anthony G, Seymour, Neal E, Jordan-Black, Julie-Anne, Bunting, Brendan P, McGlade, Kieran and Satava, Richard Martin [2013]: Prospective, randomized assessment of transfer of training (tot) and transfer effectiveness ratio (ter) of virtual reality simulation training for laparoscopic skill acquisition, Annals of surgery **257**(6): 1025–1031.

[48] Gangi, Afshin, Guth, Stephane, Imbert, Jean Pierre, Marin, Horia and Dietemann, Jean-Louis [2003]: Percutaneous vertebroplasty: Indications, technique, and results 1, Radiographics **23**(2): e10–e10.

[49] Gawande, Atul A, Zinner, Michael J, Studdert, David M and Brennan, Troyen A [2003]: Analysis of errors reported by surgeons at three teaching hospitals, Surgery **133**(6): 614–621.

[50] Gordon, Michael S [1974]: Cardiology patient simulator, American Journal of Cardiology **34**(3): 350–355.

[51] Grantcharov, TEODOR P, Kristiansen, VB, Bendix, Jørgen, Bardram, L, Rosenberg, J and Funch-Jensen, Peter [2004]: Randomized clinical trial of virtual reality simulation for laparoscopic skills training, British Journal of Surgery **91**(2): 146–150.

[52] Greenberg, Caprice C, Regenbogen, Scott E, Studdert, David M, Lipsitz, Stuart R, Rogers, Selwyn O, Zinner, Michael J and Gawande, Atul A [2007]: Patterns of communication breakdowns resulting in injury to surgical patients, Journal of the American College of Surgeons **204**(4): 533–540.

[53] H3D API [n.d.].
URL: *http://www.h3dapi.org*

[54] Halverson, Amy L, Casey, Jessica T, Andersson, Jennifer, Anderson, Karen, Park, Christine, Rademaker, Alfred W and Moorman, Don [2011]: Communication failure in the operating room, Surgery **149**(3): 305–310.

[55] Hart, Sandra G and Staveland, Lowell E [1988]: Development of nasa-tlx (task load index): Results of empirical and theoretical research, Advances in psychology **52**: 139–183.

[56] Hassan, I, Weyers, P, Maschuw, K, Dick, B, Gerdes, B, Rothmund, M and Zielke, A [2006]: Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance, British Journal of Surgery **93**(12): 1554–1559.

[57] Hegarty, Mary, Keehner, Madeleine, Cohen, Cheryl, Montello, Daniel R and Lippa, Yvonne [2007]: The role of spatial cognition in medicine: Applications for selecting and training professionals, Applied spatial cognition S. 285–315.

[58] Helmreich, Robert L and Davies, Jan M [1996]: 3 human factors in the operating room: interpersonal determinants of safety, efficiency and morale, Baillière's Clinical Anaesthesiology **10**(2): 277–295.

[59] Helmreich, Robert L, Merritt, Ashleigh C and Wilhelm, John A [1999]: The evolution of crew resource management training in commercial aviation, The international journal of aviation psychology **9**(1): 19–32.

[60] Heng, Pheng-Ann, Cheng, Chun-Yiu, Wong, Tien-Tsin, Yangsheng, X, Chui, Yim-Pan, Chan, Kai-Ming and Tso, Shiu Kit [2003]: Virtual reality based system for training on knee arthroscopic surgery., Studies in health technology and informatics **98**: 130–136.

[61] Howard, Steven K, Gaba, David M, Fish, Kevin J, Yang, George and Sarnquist, Frank H [1992]: Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents., Aviation, space, and environmental medicine **63**(9): 763–770.

[62] Hull, Louise, Arora, Sonal, Aggarwal, Rajesh, Darzi, Ara, Vincent, Charles and Sevdalis, Nick [2012]: The impact of nontechnical skills on technical performance in surgery: a systematic review, Journal of the American College of Surgeons **214**(2): 214–230.

[63] James, John T [2013]: A new, evidence-based estimate of patient harms associated with hospital care, Journal of patient safety **9**(3): 122–128.

[64] Jannin, Pierre and Korb, Werner [2008]: Assessment of image-guided interventions, Image-Guided Interventions, Springer, S. 531–549.

[65] Johnston, Maximilian J., Paige, John T., Aggarwal, Rajesh, Stefanidis, Dimitrios, Tsuda, Shawn, Khajuria, Ankur and Arora, Sonal [2015]: An overview of research priorities in surgical simulation: what the literature shows has been achieved during the 21st century and what remains, The American Journal of Surgery **211**(1): 214–225.

[66] Kahol, Kanav, Leyba, Mario J, Deka, Mary, Deka, Vikram, Mayes, Stephanie, Smith, Marshall, Ferrara, John J and Panchanathan, Sethuraman [2008]: Effect of fatigue on psychomotor and cognitive skills, The American Journal of Surgery **195**(2): 195–204.

[67] Karavidas, Maria Katsamanis, Lehrer, Paul M, Lu, Shou-En, Vaschillo, Evgeny, Vaschillo, Bronya and Cheng, Andrew [2010]: The effects of workload on respiratory variables in simulated flight: a preliminary study, Biological psychology **84**(1): 157–160.

[68] Keehner, Madeleine, Lippa, Yvonne, Montello, Daniel R, Tendick, Frank and Hegarty, Mary [2006]: Learning a spatial skill for surgery: How the contributions of abilities change with practice, Applied cognitive psychology **20**(4): 487–504.

[69] Kirkpatrick, Katherine and MacKinnon, Ralph James [2012]: Technology-enhanced learning in anaesthesia and educational theory, Continuing Education in Anaesthesia, Critical Care & Pain **12**(5): 263–267.

[70] Kneebone, Roger [2010]: Simulation, safety and surgery, Quality and Safety in Health Care **19**(Suppl 3): i47–i52.

[71] Kohn, Linda T, Corrigan, Janet M, Donaldson, Molla S et al. [2000]: To err is human:: building a Safer Health System, Bd. 6, National Academies Press.

[72] Koutantji, Maria, McCulloch, Peter, Undre, Shabnam, Gautama, Sanjay, Cunniffe, Simon, Sevdalis, Nick, Davis, Rachel, Thomas, Piers, Vincent, Charles and Darzi, Ara [2008]: Is team training in briefings for surgical teams feasible in simulation?, Cognition, Technology & Work **10**(4): 275–285.

[73] Kozmenko, Valeriy, Paige, John and Chauvin, Sheila [2007]: Initial implementation of mixed reality simulation targeting teamwork and patient safety., Studies in health technology and informatics **132**: 216–221.

[74] Krueger, Antonio, Bliemel, Christopher, Zettl, Ralph and Ruchholtz, Steffen [2009]: Management of pulmonary cement embolism after percutaneous vertebroplasty and kyphoplasty: a systematic review of the literature, European Spine Journal **18**(9): 1257.

[75] Lateef, Fatimah [2010]: Simulation-based learning: Just like the real thing, Journal of Emergencies, Trauma and Shock **3**(4): 348.

[76] LeBlanc, Vicki R, Tabak, Diana, Kneebone, Roger, Nestel, Debra, MacRae, Helen and Moulton, Carol-Anne [2009]: Psychometric properties of an integrated assessment of technical and communication skills, The American Journal of Surgery **197**(1): 96–101.

[77] LeBlanc, VR, Bould, MD, McNaughton, N, Brydges, R, Piquette, D and Sharma, B [2011]: Simulation in postgraduate medical education, Members of the FMEC PG consortium .

[78] Lee, Jason Y, Mucksavage, Phillip, Canales, Cecilia, McDougall, Elspeth M and Lin, Sharon [2012]: High fidelity simulation based team training in urology: a preliminary interdisciplinary study of technical and nontechnical skills in laparoscopic complications management, The Journal of urology **187**(4): 1385–1391.

[79] Leong, JJH, Leff, DR, Das, A, Aggarwal, R, Reilly, P, Atkinson, HDE, Emery, RJ and Darzi, AW [2008]: Validation of orthopaedic bench models for trauma surgery, Journal of Bone & Joint Surgery, British Volume **90**(7): 958–965.

[80] Lingard, Lorelei, Espin, Sherry, Whyte, Sarah, Regehr, Glenn, Baker, GR, Reznick, Richard, Bohnen, J, Orser, B, Doran, D and Grober, E [2004]: Communication failures in the operating room: an observational classification of recurrent types and effects, Quality and Safety in Health Care **13**(5): 330–334.

[81] Lohman, David F [1979]: Spatial ability: A review and reanalysis of the correlational literature., Technical report, DTIC Document.

[82] Looker, Anne C, for Health Statistics (US), National Center et al. [2012]: Osteoporosis Or Low Bone Mass at the Femur Neck Or Lumbar Spine in Older Adults, United States, 2005-2008, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics Hyattsville, MD.

[83] Lundin, Karljohan, Cooper, Matthew, Persson, Anders, Evestedt, Daniel and Ynnerman, Anders [2007]: Enabling design and interactive selection of haptic modes, Virtual Reality **11**(1): 1–13.

[84] Lundin, Karljohan, Gudmundsson, Björn and Ynnerman, Anders [2005]: General proxy-based haptics for volume visualization, Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint, IEEE, S. 557–560.

[85] Luursema, Jan-Maarten, Buzink, Sonja N, Verwey, Willem B and Jakimowicz, JJ [2010]: Visuo-spatial ability in colonoscopy simulator training, Advances in health sciences education **15**(5): 685–694.

[86] Martin, JA, Regehr, G, Reznick, R, MacRae, H, Murnaghan, J, Hutchison, C and Brown, M [1997]: Objective structured assessment of technical skill (osats) for surgical residents, British Journal of Surgery **84**(2): 273–278.

[87] Martin, Steven and Hillier, Nick [2009]: Characterisation of the novint falcon haptic device for application as a robot manipulator, Australasian Conference on Robotics and Automation (ACRA), Citeseer, S. 291–292.

[88] Mathis, JohnM. [2006]: Percutaneous vertebroplasty: Procedure technique, in J. Mathis, H. Deramond and S. Belkoff (eds), Percutaneous Vertebroplasty and Kyphoplasty, Springer New York, S. 112–133.

[89] Mcgaghie, William C., Issenberg, Saul B., Barsuk, Jeffrey H. and Wayne, Diane B. [2014]: A critical review of simulation-based mastery learning with translational outcomes, Medical Education **48**(4): 375–385.

[90] McGreevy, James M [2005]: The aviation paradigm and surgical education, Journal of the American College of Surgeons **201**(1): 110–117.

[91] McReynolds, Tom, Blythe, David, Grantham, Brad and Nelson, Scott [1998]: Advanced graphics programming techniques using opengl, Computer Graphics S. 95–145.

[92] Meng, Ma, Fallavollita, Pascal, Blum, Troels, Eck, Ulrich, Sandor, Christian, Weidert, Simon, Waschke, Jens and Navab, Nassir [2013]: Kinect for interactive ar anatomy learning, Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on, IEEE, S. 277–278.

[93] Miller, Gregory E [1990]: The assessment of clinical skills/competence/performance., Academic medicine **65**(9): S63–7.

[94] Mishra, A, Catchpole, K and McCulloch, P [2009]: The oxford notechs system: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre, Quality and Safety in Health Care **18**(2): 104–108.

[95] Moorthy, Krishna, Munz, Yaron, Adams, Sally, Pandey, Vikas and Darzi, Ara [2005]: A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre, Annals of surgery **242**(5): 631.

[96] Moorthy, Krishna, Munz, Yaron, Forrest, Damien, Pandey, Vikas, Undre, Shabnam, Vincent, Charles and Darzi, Ara [2006]: Surgical crisis management skills training and assessment: a stimulation-based approach to enhancing operating room performance, Annals of surgery **244**(1): 139.

[97] Mulconrey, Daniel S [2013]: Fluoroscopic radiation exposure in spinal surgery: In vivo evaluation for operating room personnel., Journal of spinal disorders & techniques .

[98] Okuda, Y, Bryson, E, Demaria, S, Jacobson, L, Quinones, J, Shen, B and Levine, A [2009]: The Utility of Simulation in Medical Education: What Is the Evidence?, Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine **76**(4): 330–343.

[99] O'Regan, N [2010]: Simulation of malignant hyperthermia improves patient safety, Can J Anaesth **57**: S177.

[100] Paige, John, Kozmenko, Valeriy, Morgan, Barbara, Howell, D Shannon, Chauvin, Sheila, Hilton, Charles, Cohn, Isidore and O'Leary, J Patrick [2007]: From the flight deck to the operating room: an initial pilot study of the feasibility and potential impact of true interdisciplinary team training using high-fidelity simulation, Journal of surgical education **64**(6): 369–377.

[101] Paige, John T, Kozmenko, Valeriy, Yang, Tong, Gururaja, Ramnarayan Paragi, Hilton, Charles W, Cohn, Isidore and Chauvin, Sheila W [2009]: High-fidelity, simulation-based, interdisciplinary operating room team training at the point of care, Surgery **145**(2): 138–146.

[102] Papaioannou, Alexandra, Morin, Suzanne, Cheung, Angela M, Atkinson, Stephanie, Brown, Jacques P, Feldman, Sidney, Hanley, David A, Hodsman, Anthony, Jamal, Sophie A, Kaiser, Stephanie M et al. [2010]: 2010 clinical practice guidelines for the diagnosis and management of osteoporosis in canada: summary, Canadian Medical Association Journal **182**(17): 1864–1873.

[103] Pati, Sarthak, Erat, Okan, Wang, Lejing, Weidert, Simon, Euler, Ekkehard, Navab, Nassir and Fallavollita, Pascal [2013]: Accurate pose estimation using single marker single camera calibration system, SPIE medical imaging, International Society for Optics and Photonics, S. 867126–867126.

[104] Pauzie, A [2008]: A method to assess the driver mental workload: The driving activity load index (dali), IET Intelligent Transport Systems **2**(4): 315–322.

[105] Pettersson, Johanna, Palmerius, Karljohan Lundin, Knutsson, Hans, Wahlström, Ola, Tillander, Bo and Borga, Magnus [2008]: Simulation of patient specific cervical hip fracture surgery with a volume haptic interface, Biomedical Engineering, IEEE Transactions on **55**(4): 1255–1265.

[106] Ploeg, Wendy T., Veldhuizen, Albert G., The, Bertram and Sietsma, Maurits S. [2006]: Percutaneous vertebroplasty as a treatment for osteoporotic vertebral compression fractures: a systematic review, European Spine Journal **15**(12): 1749.

[107] Powers, Kinga A, Rehrig, Scott T, Irias, Noel, Albano, Hedwig A, Malinow, Andrew, Jones, Stephanie B, Moorman, Donald W, Pawlowski, John B and Jones, Daniel B [2008]: Simulated laparoscopic operating room crisis: an approach to enhance the surgical team performance, Surgical endoscopy **22**(4): 885–900.

[108] Proceedings of the 4th International Conference on Healthcare Ergonomics and Patient Safety (HEPS): Bridging Research and Good Practices towards Patients Welfare [2014]: Taipei, Taiwan.

[109] Ra, JB, Kwon, SM, Kim, JK, Yi, J, Kim, KH, Park, HyunWook, Kyung, K-U, Kwon, D-S, Kang, HS, Kwon, ST et al. [2002]: Spine needle biopsy simulator using visual and force feedback, Computer Aided Surgery **7**(6): 353–363.

[110] Resnick, Daniel K, Barr, John D and Garfin, Steven R [2011]: Vertebroplasty and kyphoplasty, Thieme.

[111] Rethans, J-J, Norcini, JJ, Baron-Maldonado, M, Blackmore, D, Jolly, BC, LaDuca, T, Lew, S, Page, GG and Southgate, LH [2002]: The relationship between competence and performance: implications for assessing practice performance, Medical education **36**(10): 901–909.

[112] Risucci, Donald A [2002]: Visual spatial perception and surgical competence, The American Journal of Surgery **184**(3): 291–295.

[113] Ritter, E Matt, McClusky, David A, Gallagher, Anthony G, Enochsson, Lars and Smith, C Daniel [2006]: Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator, The American Journal of Surgery **192**(3): 379–384.

[114] Rodgers, David L [2007]: High-fidelity patient simulation: a descriptive white paper report, Healthcare Simulation Strategies **10**(4): 68–77.

[115] Rodriguez-Paz, JM, Kennedy, M, Salas, E, Wu, AW, Sexton, JB, Hunt, EA and Pronovost, PJ [2009]: Beyond see one, do one, teach one: toward a different training paradigm, Quality and Safety in Health Care **18**(1): 63–68.

[116] Rogers, Selwyn O, Gawande, Atul A, Kwaan, Mary, Puopolo, Ann Louise, Yoon, Catherine, Brennan, Troyen A and Studdert, David M [2006]: Analysis of surgical errors in closed malpractice claims at 4 liability insurers, Surgery **140**(1): 25–33.

[117] Rosen, Kathleen R [2008]: The history of medical simulation, Journal of critical care **23**(2): 157–166.

[118] Ruspini, Diego C, Kolarov, Krsimir and Khatib, Oussama [1997]: Haptic interaction in virtual environments, Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on, Bd. 1, IEEE, S. 128–133.

[119] Safar, Peter, Brown, Torrey C., Holtey, Warren J. and Wilder, Robert J. [1961]: Ventilation and Circulation with Closed-Chest Cardiac Massage in Man, JAMA **176**(7): 574–576.

[120] Satava, Richard M. [1993]: Virtual reality surgical simulator, Surgical Endoscopy **7**(3): 203–205.

[121] Schout, BMA, Hendrikx, AJM, Scheele, F, Bemelmans, BLH and Scherpbier, AJJA [2010]: Validation and implementation of surgical simulators: a critical review of present, past, and future, Surgical endoscopy **24**(3): 536–546.

[122] Sexton, B, Marsch, S, Helmreich, R, Betzendoerfer, D, Kocher, T, Scheidegger, D et al. [1998]: Participant evaluation of team oriented medical simulation, Simulators in anesthesiology education, Springer, S. 109–110.

[123] Seymour, Neal E [2008]: Vr to or: a review of the evidence that virtual reality simulation improves operating room performance, World journal of surgery **32**(2): 182–188.

[124] Seymour, Neal E, Gallagher, Anthony G, Roman, Sanziana A, O'Brien, Michael K, Bansal, Vipin K, Andersen, Dana K and Satava, Richard M [2002]: Virtual reality training improves operating room performance: results of a randomized, double-blinded study, Annals of surgery **236**(4): 458.

[125] Society of American Gastrointestinal and Endoscopic Surgeons: Fundamentals of laparoscopic surgery [2010].
URL: *http://www.flsprogram.org/*

[126] Spencer, F [1978]: Teaching and measuring surgical techniques: the technical evaluation of competence, Bull Am Coll Surg **63**(3): 9–12.

[127] Srinivasan, Dushyanth, Than, Khoi D, Wang, Anthony C, La Marca, Frank, Wang, Page I, Schermerhorn, Thomas C and Park, Paul [2014]: Radiation safety and spine surgery: systematic review of exposure limits and methods to minimize radiation exposure, World neurosurgery **82**(6): 1337–1343.

[128] Steele, RJC, Walder, C and Herbert, M [1992]: Psychomotor testing and the ability to perform an anastomosis in junior surgical trainees, British Journal of surgery **79**(10): 1065–1067.

[129] Stevens, Louis-Mathieu, Cooper, Jeffrey B, Raemer, Daniel B, Schneider, Robert C, Frankel, Allan S, Berry, William R and Agnihotri, Arvind K [2012]: Educational program in crisis management for cardiac surgery teams including high realism simulation, The Journal of thoracic and cardiovascular surgery **144**(1): 17–24.

[130] Sutherland, Leanne M., Middleton, Philippa F., Anthony, Adrian, Hamdorf, Jeffrey, Cregan, Patrick, Scott, David and Maddern, Guy J. [2006]: Surgical Simulation: A Systematic Review, Annals of Surgery **243**(3): 291–300.

[131] Sweller, John [2011]: Cognitive load theory, The psychology of learning and motivation: Cognition in education **55**: 37–76.

[132] Taffinder, N, Smith, S, Mair, J, Russell, R and Darzi, A [1999]: Can a computer measure surgical precision? reliability, validity and feasibility of the icsad, Surg Endosc **13**(suppl 1): 81.

[133] Tan, Shaw Boon, Pena, Guilherme, Altree, Meryl and Maddern, Guy J [2014]: Multidisciplinary team simulation for the operating theatre: a review of the literature, ANZ journal of surgery **84**(7-8): 515–522.

[134] Touchie, Claire, Humphrey-Murto, Susan and Varpio, Lara [2013]: Teaching and assessing procedural skills: a qualitative study., BMC medical education **13**: 69.

[135] Undre, Shabnam, Koutantji, Maria, Sevdalis, Nick, Gautama, Sanjay, Selvapatt, Nowlan, Williams, Samantha, Sains, Parvinderpal, McCulloch, Peter, Darzi, Ara and Vincent, Charles [2007]: Multidisciplinary crisis simulations: the way forward for training surgical teams, World journal of surgery **31**(9): 1843–1853.

[136] Vaccaro, Alexander R and Bono, Christopher M [2007]: Minimally invasive spine surgery, CRC Press.

[137] Volk, Mark S, Ward, Jessica, Irias, Noel, Navedo, Andres, Pollart, Jennifer and Weinstock, Peter H [2011]: Using medical simulation to teach crisis resource management and decision-making skills to otolaryngology housestaff, Otolaryngology–Head and Neck Surgery **145**(1): 35–42.

[138] Walker-Bone, Karen, Dennison, Elaine and Cooper, Cyrus [2001]: Epidemiology of osteoporosis, Rheumatic Disease Clinics of North America **27**(1): 1–18.

[139] Wang, Li-jun, Yang, Hui-lin, Shi, Yuan-xin, Jiang, Wei-min and Chen, Liang [2012]: Pulmonary cement embolism associated with percutaneous vertebroplasty or kyphoplasty: a systematic review, Orthopaedic surgery **4**(3): 182–189.

[140] Wanzel, Kyle R, Hamstra, Stanley J, Caminiti, Marco F, Anastakis, Dimitri J, Grober, Ethan D and Reznick, Richard K [2003]: Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance, Surgery **134**(5): 750–757.

[141] Weigl, Matthias, Stefan, Philipp, Abhari, Kamyar, Wucherer, Patrick, Fallavollita, Pascal, Lazarovici, Marc, Weidert, Simon, Euler, Ekkehard and Catchpole, Ken [2016]: Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure., Surg Endosc **30**(2): 559–566.

[142] Wetzel, Cordula M, Black, Stephen A, Hanna, George B, Athanasiou, Thanos, Kneebone, Roger L, Nestel, Debra, Wolfe, John HN and Woloshynowych, Maria [2010]: The effects of stress and coping on surgical performance during simulations, Annals of surgery **251**(1): 171–176.

[143] Wetzel, Cordula M, Kneebone, Roger L, Woloshynowych, Maria, Nestel, Debra, Moorthy, Krishna, Kidd, Jane and Darzi, Ara [2006]: The effects of stress on surgical performance, The American Journal of Surgery **191**(1): 5–10.

[144] Wieczorek, Matthias, Aichert, André, Fallavollita, Pascal, Kutter, Oliver, Ahmadi, Ahmad, Wang, Lejing and Navab, Nassir [2011]: Interactive 3d visualization of a single-view x-ray image, Medical Image Computing and Computer-Assisted Intervention, Springer, S. 73–80.

[145] Wilson, Mark R, Poolton, Jamie M, Malhotra, Neha, Ngo, Karen, Bright, Elizabeth and Masters, Rich SW [2011]: Development and validation of a surgical workload measure: the surgery task load index (surg-tlx), World journal of surgery **35**(9): 1961–1969.

[146] Wilson, MS, Middlebrook, A, Sutton, C, Stone, R and McCloy, RF [1997]: Mist vr: a virtual reality trainer for laparoscopic surgery assesses performance., Annals of the Royal College of Surgeons of England **79**(6): 403.

[147] Wucherer, Patrick, Stefan, Philipp, Abhari, Kamyar, Fallavollita, Pascal, Lazarovici, Marc, Weidert, Simon, Peters, Terry M, Ribaupierre, Sandrine De, Eagleson, Roy and Navab, Nassir [2015]: Vertebroplasty Performance on Simulator for 19 Surgeons Using Hierarchical Task Analysis., IEEE transactions on medical imaging **19**(c).

[148] Wucherer, Patrick, Stefan, Philipp, Weidert, Simon, Fallavollita, Pascal and Navab, Nassir [2013]: Development and Procedural Evaluation of Immersive Medical Simulation Environments, International Conference on Information Processing in Computer-Assisted Interventions S. 1–10.

[149] Wucherer, Patrick, Stefan, Philipp, Weidert, Simon, Fallavollita, Pascal and Navab, Nassir [2014]: Task and crisis analysis during surgical training., International journal of computer assisted radiology and surgery **9**(5): 785–94.

[150] Yule, S, Flin, R, Paterson-Brown, S, Maran, N and Rowley, D [2006]: Development of a rating system for surgeons' non-technical skills, Medical education **40**(11): 1098–1104.

[151] Zausig, YA, Grube, C, Boeker-Blum, T, Busch, CJ, Bayer, Y, Sinner, B, Zink, W, Schaper, N and Graf, BM [2009]: Inefficacy of simulator-based training on anaesthesiologists' non-technical skills*, Acta anaesthesiologica Scandinavica **53**(5): 611–619.

[152] Zendejas, Benjamin, Brydges, Ryan, Wang, Amy T. and Cook, David a. [2013]: Patient outcomes in simulation-based medical education: a systematic review., Journal of general internal medicine **28**(8): 1078–1089.

[153] Zevin, Boris, Aggarwal, Rajesh and Grantcharov, Teodor 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.

[154] Ziewacz, John E, Arriaga, Alexander F, Bader, Angela M, Berry, William R, Edmondson, Lizabeth, Wong, Judith M, Lipsitz, Stuart R, Hepner, David L, Peyre, Sarah, Nelson, Steven et al. [2011]: Crisis checklists for the operating room: development and pilot testing, Journal of the American College of Surgeons **213**(2): 212–217.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht.

February 14, 2017

Patrick Wucherer