Master thesis
University of Twente
Supervisors: Dr.ir.A.A.M Spil
Dr.G.W.J.Bruinsma

Serious Games in Medical Education

as Learning Tools

Written by:
Juliane Teschner (s0105848)
May , 2016

Acknowledgments

Firstly, i wanted to thank very much both supervisors, Dr.Spil and Dr. Bruinsma for their patience, endurance, flexibility, confidence, advice, feedback and guidance during this project- without which this Master thesis would be impossible to manage.

Secondly, I want to thank very much the 10 respondents for their time and effort in taking part in the interview sessions and their valuable inputs.

Lastly, I express my heartfelt thanks to the persons, who supported me in all those difficult, enduring years, never stopping to believe in me, being my mother, Mr.Evertzen, my friends and my deceased grandpa.

Summary

Serious Games are used in the fields of military, government, corporation aviation, education, city planning and healthcare (De Wit (2011).Two major shifts contributed to this emergence. Firstly, the fast developments in the field of ICT in general and the evolution of Virtual Reality in particular enabled the uprising of Serious Gaming. Secondly, education as practice is in the 21st century a learner-centred, competency –focused and interactive endeavour.

Games have the potential of making academic learning content more learner-centred, easier, more enjoyable, more interesting and thus more effective.

To signify the high potential of SG, Lynch exemplifies that surgeons who played video games in the past for at least 3h per week had 37% fewer errors, were 27% faster and scored 42% better overall than surgeons with no video game exposure. (Lynch, 2010). Moreover, until 75% of the learned content can be memorized by practising the acquired skills.One special attention point was how SG can lower the high cost of surgical training and improve the and patient safety (high-stake training in safe environment).

The main research question is how Serious Games can enhance medical learning?
Based on the literature research and interview sessions several themes were addressed such as the definition of games, game attribute and learning outcomes. The practical side concerns the best training methods for surgeons, (dis)advantages of SG, barriers for using SG and the future of SG in medical learning.

Definition of a SG based on Ben Sawyer:
"Any computerized game whose chief mission is not entertainment and all entertainment games which can be reapplied to a different mission other than entertainment." (Ben Sawyer, BIT Conference , 2010)

The **research methodology** consisted of a qualitative approach, by interviewing
8 experts and 2 medical students with background in virtual learning environment and (medical)learning.

The key intriguing **results** show that:

- Experts acknowledge that there is no universal definition of SG, because every game has its own in-vivo setting(context, gamers, aim)
- A convergence is taken place between simulation and serious games, but the most distinguishing factor between both is the lack of fun in case of simulations
- The most important perceived advantages of SG are: fun and flow-state as motivation driver, safe environment for high-stake actions, immediate, unbiased system feedback, especially well-suited to learn knowledge (factual and procedural) and partial task (but not complex operations).
- The main disadvantages of SG concerns the validation issues, high development costs, lack of stakeholder approach and the institutional, professional and individual barriers/fears to use a new, innovative learning method, high total cost of preparation and training.
- Dissection/Vivisection will remain the best method to train surgeons, although SG will be most used in an effective mix of learning methods
- Regarding the game attributes for training surgeons, representation/accuracy (low quality for beginners, high quality for experts), challenge, fun/flow and feedback were named as highly relevant
- Particularly the opportunity for feedback and debriefing are considerably valuable for reflection and should be included whenever it is feasible
- SG are not yet ready for the 3D-VR trend (OculusRift, Hololens)
- All experts agreed that SG have the potential to be an effective, assisting learning tool.

List of tables

List of figures

Abbreviations

SG:	Serious Game(s)
EM:	Ethnomethodology
E 1-8:	Expert 1-8
MS or S 1-2:	Medical students 1-2

Table of contents

Chapter 1: Introduction

In this chapter I introduce my research, being SG, narrow down the target group (medical educators and surgical trainees), present the societal and practical importance of the subject and lastly address the research gap and questions, which will be answered by using a qualitative design.

Serious Games Serious are used in the fields of military, government, corporations, aviation, education, city planning and healthcare (De Wit (2011).Two major shifts contributed to this emergence. Firstly, the fast developments in the field of ICT in general and the evolution of Virtual Reality in particular enabled the uprising of Serious Gaming. Secondly, education as practice is in the 21st century a learner-centred, competency –focused and interactive endeavour.

Games in general are optimal to educate, train and to provoke a change in the behaviour. SGs appeal to the aspects that render learning effective, being repetition, reinforcement, association and the use of multiple senses (De Wit, 2011). In Serious Games the application of learned skills and knowledge are a key element. Up until 75% of the learned content can be memorized by practising the acquired skills.
Games have the potential of making academic learning content more learner-centred, easier, more enjoyable, more interesting and thus more effective.

Target group of this thesis
McCallum (2014) developed taxonomy for Games in healthcare based on his 10 year experience of working in the SG-field. This taxonomy illuminates the different users and activities of healthcare-focused games.

Table 1. The Games for Health Taxonomy developed by the Games for Health Project.				
Area of health activity	**Personal**	**Professional practice**	**Research and academia**	**Public health**
Preventive	"Exergaming" Stress	Patient Communication	Data Collection	Public health Messages
Therapeutic	"Rehabitainment" Disease management	Pain distraction Cyberpsychology Disease management	Virtual humans	First responders
Assessment	Self-ranking	Measurement	Inducement	Interface and visualization
Educational	First Aid Medical information	Skills and training	Recruitment	Management simulations
Informatics	Personal health records	Electronic medical records	Visualization	Epidemiology

Table 1: Taxonomy for health-related SG

Source: McCallum (2014)

Based on this taxonomy, the relevant users can be found in the personal, professional, research and public health field. The activity area is aimed at preventive, therapeutic and educational, assessment and/or informatics purposes. The focus point of this thesis is marked by the orange oval, namely the field consisting of the professional healthcare education that is aimed at the transfer of factual, procedural knowledge and (low functional, non-complex, minimal invasive) skills and the appliance and training of both. This thesis is specifically aimed at surgical trainees and their educators as target group. Robotic operation or full-scale, complex training of operations are not addressed, because my interest area is the opposite (see above).

Societal importance of the research subject

Wattanasoontorn (2013) points out that good health is a vocal point of interest of every human being to survive and prosper. It is the second most basic need in Maslow's pyramid of needs regarding the safety dimension. At the macro level, a peaceful society/ nation state can only exists on the long-term of it offers it constituents a good healthcare system. At the micro level-the bottom line- serious games have a societal importance because they help medical professionals to avoid mistakes and saving lives, and thereby increasing the patient safety. Moreover, they reduce the total training, education and certifications cost of the medical student and professionals.

As two authors (Graafland and Bridges) put it: Serious Games probably can be an effective tool to reduce the education costs of medical professionals and “*are directed at reducing medical errors and subsequent healthcare cost*“(Graafland, 2012, p.1322).

“(The) *high monetary cost suggests the need for digital skills, selection criteria, the development of training curriculum and resource facilities, the pre-operating room need for suturing and stapling techniques, and perhaps the acquisition of virtual surgery training modules*” (Bridges, 1999, p.28).

Cost of surgical training

A medical academic study costs the tax payer between 200 000€ in Germany and 450000 francs in Switzerland. (DEA, 2014) In Germany alone, in 2014 there were 19000 casualties in the hospital attributable to medical malpractice. (Zeit, 2014).

New forms of surgical training are emerging because of the inherent limitations of the clinical setting. In recent years the operating training for trainees has been curtailed because night operations are not allowed anymore (in the USA) and there is a continuous drive to reduced mistakes by inexperienced surgeons. Hence, computer simulations of surgical techniques are a good solution to get a realistic training experience that avoids any fatal errors. (Lynch, 2010). To signify the high potential of SG, Lynch exemplifies that surgeons who played video games in the past for at least 3h per week had 37% fewer errors, were 27% faster and scored 42% better overall than surgeons with no video game exposure. (Lynch, 2010)

Shortly said: There are many mistakes that occur in the medical setting and the training costs are very high. Hence I assume that SG can contribute to lower the training costs and reduce the mistakes level (by offering a trial-and-error, risk-free and safe training environment).

Theoretical and practical contribution

The theoretical contribution of thesis is to elucidate and understand Serious Games as a new phenomenon, helping to manifest SG as a field of study as a sub-area of games and simulations. The practical contribution will be based on the insights of the expert interviewees. Those insights will support educators to find the right balance of using games in the curriculum, surgeons residencies to achieve a better performance in shorter time due to the learning elements and the appeal of games and lastly, game designers to develop effective and relevant games for educators and practitioners.

Research gap and research goal

Several authors underline the need for more research on SG concerning health(care). For instance, Connolly (2012) calls for more qualitative studies to comprehend the disposition of engagement in games. According to De Wit (2012), there is a shortage of studies clarifying the clinical effectiveness of serious games. Wattanasoontorn(2013) reasons that no single definition of SG as concepts exists. Lastly, *"little is known about the processes and mechanisms through which behaviour change occurs in a serious games*" (Thompson,2008, p.9).

Graafland (2012) reveals that the effectiveness of learning through games is a new form of education and hence there is a paucity of research. Concluding, Ma (2014) also demands more systematic research efforts attained to the evaluation of how satisfied the players are and in how far the knowledge is retained. He points to the need for SG to complement existing and specific training methods and that they are no full replacement alternative.
The research goal is to close those gaps by studying which SG are used to train surgeons and how they be can be used as effective learning tools for surgeons.

Research design

A qualitative research design was chosen, because this set-up enables sufficient flexibility to investigate the SG-phenomenology. Informants were purposefully selected due to ease of access and expert knowledge of the SG-domain. Data were collected by using a semi-structured interview design. The steps in analysis (data reduction, data display and conclusions/verification) are described in Chapter 3: Methodology.

Research questions

Main RQ: How can Serious Games enhance medical learning?

Sub-questions:

<u>Questions answered by theoretical approach:</u>

What are Serious Games?

What are the games attributes of Serious Games?

What are the possible learning outcomes of Serious Games?

Which are the most relevant game attributes that enhance the learning outcomes?

Questions answered by practical approach:

What are the perceived advantages and disadvantages of Serious Games?

What are the perceived problems of using Serious Games

How important is feedback and debriefing?

How will 3DVR-tools influence Serious Gaming?

Other area of focus are: the differences between SG and simulations, games that are used at medical universities, R&D costs, best methods to train surgeons, recommended relevant game attributes, recommended learning models and a finally a assessment of the effectiveness of SG in medical education.

In the following chapter the background literature is addressed concerning serious games for training surgeons and the learning aspects. The empiricism feature is realized by interviewing experts with a healthcare education background. Based on those interviews the research questions will be answered and conclusion and advises will be formulated for the several stakeholders.

Summary; In this chapter I introduced my research top, being SG, narrowed down the target group (medical educators and surgical trainees), present the societal and practical importance of the subject (use of SG to reduce high training costs of medical education, whilst reducing the human failure rate). The main research question is: **How can SG enhance medical learning?** Further topics of interest were outlined such as the (dis)advantages and problems with SG, the importance of gaming attributes and the future of SG. Those and other background questions provide a step-wise framework for this research.

Chapter 2: Literature review

In this chapter, the literature research strategy is described, SG are displayed as concept, the conceptual model is presented at a preliminary stage to highlight the set-up of this chapter, following an input-process-output logic. Lastly, learning outcome dimensions are presented.

Literature research strategy

The guidelines for a systematic literature review (Wolfswinkel, 2013) served as best practice to draw up this review.

Inclusion and exclusion criteria (Define stage and search stage)

Articles were included in this review if they addressed the subjects of serious games in general and healthcare in particular and were written in English, German and Dutch. The fields of research covered areas such as ICT, psychology, pedagogy, learning theories, game-based education and medical education. The online database consulted were: GoogleScholar, FindUT, Heidi of the University of Heidelberg, PubMed and the search engine of the University in Bielefeld (base-search.net). The chosen search terms were: serious game or healthcare or surgical training or medical education.

Select stage

46 articles were read and screened for relevant insights and concepts and were classified into very non relevant articles offering background knowledge (28 articles) and key relevant articles (18 articles) on the literature review is based.

Analyze stage

Only the open coding practice was applied to identify the key insights, factors and empirical data found in the read articles.

Present stage

The results of the literature session are presented beneath. Furthermore, a table of articles summarizing the key findings of concepts of all the key relevant articles can be found in the appendix D.

Conceptual model

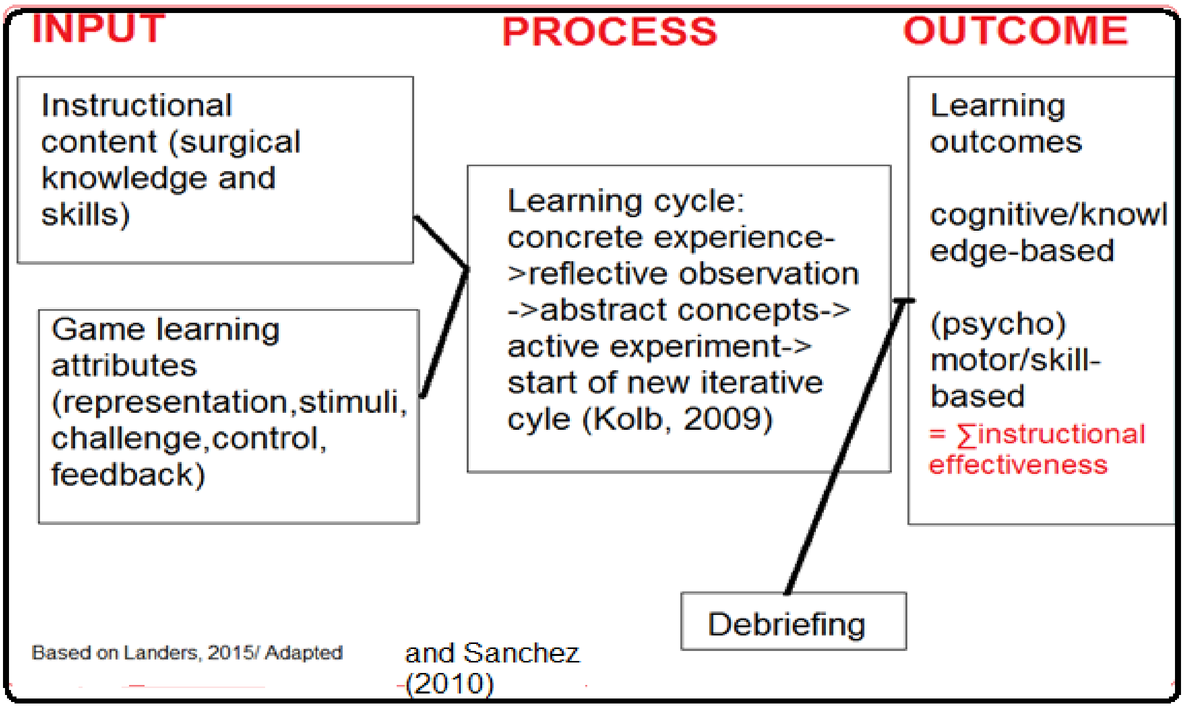


Figure 1: Conceptual Framework

Source: own model based on Landers (2015) and Sanchez (2010)

This part of the chapter follows the logic of an input-process-output model. The INPUT consists of the game characteristics (a.k.a. learning elements) and the instructional content. The instructional content is based on the medical curriculum of the individual and step-by-step procedures (based on the EPA competencies and activities). The PROCESS dimension is addressed by describing how learning takes places via experience in an iterative cycle (ELT-model). At last, the OUTCOME dimension is about learning outcomes concerning cognitive/knowledge-based and (psycho) motor skill-based aspects. The conceptual model is presented at this stage to give a logical pre,-and overview of the pivotal concepts.

Input dimension: Serious games as learning tool

General introduction to Serious Games

Belonging mainly to the information system field, Serious Games can be approached at best in an interdisciplinary way (Webster, 2002). Being an interdisciplinary venture, psychology, biology, cognitive science and sociology theories can be applied to learn more about games (Boyle 2011).

Origin of Serious Games, definitions and description of their aspects/characteristics on a general level

The term SG was brought to life by Clark Abt in a book published in 1970. (De Wit, 2011). The launch of the game America's Army in 2002 by the US Army is seen as the starting point of the serious game evolution. SG need to manifest an *“explicit and carefully thought-out educational purpose”* (De Wit, 2011, p.17).The Serious game concept is often confused with the terms game-based learning, digital game-based learning, e-learning and edutainment (Susi, 2007). According to Boyle (2011), SGs are a subset of game-based learning.

In short, serious games are games that educate, and train players, changing their behaviours (Thompson, 2008). They represent a mix of *“game-based methods and concepts and game technology (which) are combined with other ICT technologies and research areas; and are applied to a broad spectrum of applications domains ranging from training, simulation and education (...)”)Göbel,2010 p. 1663).*

SGs are fundamentally different from entertainment games, because in SG the knowledge and experience acquisition and application of learned skills besides the virtual reality imitation are stressed. (Susi, 2007)

Graafland (2012) defines serious games as “‘i*nteractive computer application... that has a challenging goal, is fun to play and engaging, incorporates some scoring mechanism, and supplies the user with skills, knowledge or attitudes useful in reality “* (Graafland, 2012, p. 2322). The main goal is the deliberate learning of new skills and knowledge: through repetition the players will be turned into an experts.

Hereafter the definition of Ben Sawyer will be used, because he is the ‘guru’ and founder of the Serious Games Initiative. This definition is highly recommended by expert3:*“Any computerized game whose chief mission is not entertainment and all entertainment games which can be reapplied to a different mission other than entertainment”* (BIT conference paper, 2010).

This definition suits the purpose of this thesis perfectly, because the transfer of medical knowledge and skills is a core theme and not entertainment, but edutainment.

Forward-looking, Serious Games can be conceptualised as consisting of several game attributes such as fantasy, rules and goals, sensory stimulation, challenge, mystery and control. All are described in a section underneath.

Historical development

The work of Faria (2009) can be used to describe the development of games, although his main interest is business games. In the first phase, hand-scored games were used to transfer knowledge and skills. In the second and third phase, mainframe games were used with more variables. In the fourth phase, games were offered via installations on PC in a 2nd reality. Nowadays, in the fifth phase, decentralised web-based games and cloud-games with 3rd realty options are common ways to offer learning applications.
Eigenfeldt-Nielsen (2007) describes the development of computer games from natural science (behaviourism) based 1st generation games to post-modern science with 3rd generation games such as SG that have 3d realities. In the first generation, the teacher has authority and the student just listening. As learning is democratised, the teacher becomes a facilitator of learning. The students in the post-modern education context construct and de-construct knowledge by themselves.

Preview of general benefits/ effectiveness of serious games

In dependence on the modern theory of learning, Connolly (2012) states that games in general are well-suited for learning purposes if games corresponds with the "*active, experiential, situated, problem-learning and (...)immediate feedback" aspects of learning.* "(Connolly, 2012, p. 661)

Connolly (2012) performed a meta-analysis systematic literature review. He classifies the mostly positive outcomes alongside the following dimensions:

- learning and behaviour and behaviour change
- knowledge acquisition and content understanding
- perceptive and cognitive skills
- motor skills
- social and psychological (affective and motivational)skills
- physiological results
- (un)intended results.

One surprising finding his study was that even games that were not purposely designed to enhance learning, did exactly that. In 2016 an update was published on that former article, which confirmed the findings of the 2012 paper. The update report found 512 papers (2012:129 papers)that corroborated the positve effects of SG. New was the amount of higher quality papers (34 vs.8) 'Hard' sciences like health and science researched the effectiveness of SG in a strict-designed, quantitative way, whereas the soft domains (such as business and social science) use a qualitative, loose-designed way. North America and Europe are still the leading innovation regions for SG.SG are foremost used for knowledge acquisition, not so much to induce skills or behaviour changes. More unintentional learning was determined more often. Surprisingly, simulations were the most popular game genre. Competition, uncertainty of information and varying training modes were the game features that engaged players, fostered learning and enhanced motivation. Flow, engagement and appeal of games make them pleasant.

Getting more specific, De Wit (2011) refers to the competitive, entertainment and feedback aspects of serious games that renders learning effective. Compared to the traditional, offline simulation method, SG are more enjoyable and flexible better portable, cheap to distribute over the internet, storable, individual scalable regarding the skill development, and creating cost-effective testing environments.

Lynch (2010) makes some cautionary comment in stating that "*manual surgical technique is just one of the competencies it takes to become a good surgeon, but clinical outcomes are strongly influenced by interpersonal communication skills and good judgement* "(Lynch, 2010, p. 188)

Remark: Before Serious games are presented as tool for medical education, other available learning methods for medical education are portrayed.

Serious games in the medical education

Before the issue of serious games in surgical education can be addressed, it is necessary to outline the classical and modern tools of learning medical procedures.

Overview of the surgical tools portfolio

Based on Luursema (2010) and the French medical training centre (ILUMENS.fr), dissection, vivisection, standardized patients, box-trainers and technology-enhanced dummies, computer-assisted simulations, hybrid simulation systems, autonomous robotic systems and lastly SG are used as surgical training tools- all are described in Appendix A. Dissection is still seen as gold standard, especially to gain respect and empathy for the individual patient and haptic feel of the tissue. Vivisection's pro is the functioning blood circle. **Serious games** offer an opportunity to transfer knowledge and skills for surgical procedures by using a 2D or 3D gamifying approach. The highest form of technology-evolved, full operation equipment, are autonomous robotic systems for surgery. An example of such a system is STAR: Smart Tissue Autonomous Robot, which can treat gall bladders and blind gut completely alone. It can operate just as well as surgeon. Improved rates for efficacy, safety, consistency and access to optimized surgical techniques" are stated as general benefits.(Shademan, 2016 and NOS, nl,2016)

Based on Perrow's framework, it is tried to make a classifications of the different surgical training tools. His model is applicable to a wide variety of tasks and activities. He uses task variability and task analyzability as two dimensions to classify organizational technology (here: medical institution).Task variability is the degree to which the labour process involves routinized standard procedures (non-variable) or whether a wide range of exceptions (variability) must be handled. Task analyzability is about the degree to which formal procedures are needed for handling the non-standardized exceptional cases (analyzability) or whether exceptions require workers to improvise (non- analyzability).

Based on the two dimensions, four kinds of organizational technology can be distinguished: Routine technology has standard operating procedures for dealing with exceptions (low variability, high analyzability).Craft technology entails standard procedures for most tasks, but when exceptions and problems happen, workers must innovate and improvise due to the inability to formalize all possible contingent cases. Engineering technology is relevant if variability is high (work cannot be reduced to a standardized protocol), but analyzability is also high (as range of problems have been predetermined and solutions are formalized). The variety of cases is handled via a set of specific rules, procedures and forms (Jaffee, 2001, p.188).

low task variability high task analyzability	Routine technology task environment: stable and certain (Jaffee,2001,p.211)	Robotic surgery,box-trainer,standized patients and dummies,CA hybrid systems	low fidelity system opportunity
high task variability high task analyzability	Engineering technology task environment:unstable and uncertain(Jaffee, 2001,p.211)	Simulations,SG, vivisection,dissection	high fidelity system opportunity

Table 2: Technology classification and task environment for surgical education tools

Source: Adapted, based on Jaffee, 2001, p.188/ 211

Games in surgical curriculum

Training in medical education is mostly aimed at achieving competency in one specific surgical procedure. Trainees must become proficient in skills of this activity such as technical skills, cognitive skills, procedural knowledge, judgment, decision making, leadership, communication and teamwork "(Graafland, 2014, p.13)

The surgical postgraduate curriculum in the Netherlands aim to turn novice medical students into proficient, competent experts, providing 1200 hours of operating time. Serious Games have the potential of being a powerful training and performance assessment tool for individual, step-by-step procedures, limiting the time required in a costly training in a 'live'-setting operating room.

Examples of surgical SG

Graafland (2012) gives some examples of SGs for surgical trainees such as Pulse, 3Di Teams, CliniSpace and the predecessor Virtual ED as platforms to train critical care. The Off-Pump Coronary Artery Bypass and Total Knee Arthoplasty are examples of games to train decision steps in a virtual operating room. Also commercial, 'unintended' aims games can improve surgical skills. For example games belonging to the Microsoft, Nintendo and Mii platform were shown to improve laparoscopic handling speed.

Serious Games as accelerator for evidence-based medicine knowledge diffusion

Diner (2007) demands a effective knowledge translation of evidence-based medicine in graduate medical education, ensuring that best evidence is applied to ensure optimal patient care. Evidence based medicine skills are about learning critical

appraisal and information mastery skills concerning a (novice- undergraduate) medical student. During their professional endeavour, doctors must be aware of the continuous change of knowledge body due to new insights in the EBM-literature. The final stage is the efficient information mastery and real-time utilization in the clinical setting.
To bridge the gap between moment of EBM-finding and implementation of these method in clinical settings to benefit patients it takes up to 13 years and only 50% of best evidence is introduced:
18% of evidence is lost from time of research discovery to manuscript submission, 46% and 6 months from submission of findings to a journal to their acceptance, 50% from journal publication to erroneous or incomplete indexing and 35% more and 6 up to 13 years for incorporation into reviews, guidelines and textbooks.

Although back in 2007 Diners favored the introduction of a journal club to accelerate the diffusion of EBM-interventions, serious games could have the potential of serving as accelerator for knowledge transmission

Gaming attributes that foster learning

The work of Wilson (2009) offers a sound description of the main game attributes that foster learning, mainly developed by Garris (2002). He adds up representation and assessment as dimensions. Pavlas (2009) focused himself on the dimension of the game realty, conflict/challenge and assessment categories of a game that manifestly contribute directly to learning. The focus point in thesis is the feedback attribute of a serious game as a facilitator for learning, but the three other relevant categories of a serious game are shortly outlined before feedback is addressed.

Game attribute Wilson (2009) **=7 in total**	**Game attribute Garris (2002)** **=6 in total**
Fantasy: make-believe environment, scenario or characters	Fantasy: imaginary or fantasy context
Mystery:gap between known and unknown information (meta-information)	Mystery: optimal level of informational complexitiy
Assessment: immediate feedback of user's progress	Rules /Goals: clear rules and goals and feedback on progress towards goals
Sensory Stimuli: visual, auditory, tactile	Sensory Stimuli: dramatic or novel visual

stimulations to create alternative reality Representation: precision of reproduction: physical and psychical similarity between game and environment it represents	and auditory stimuli
Challenge: ideal amount of difficulty and improproabilty of obtaining goals	Challenge: optimal level of difficulty and uncertain goal attainment
Control:user's ability to influence elements of their learning environment	Control: active learner control

Table 3: Game attributes

Source: Wilson

Because my main area of interest is realistic 3D serious games, i will focus on 5 of the game attributes (in red) regarding the game reality dimension (representation, sensor stimuli), challenge/Conflict dimension, control dimension and feedback dimension (Pavlas, 2009). Because of the paramount importance of feedback for learning, this attribute is elaborately described. Fantasy and mystery are assessed to be non-critical, because firstly, fantasy is an issue of Adventure/Fantasy games (which is not our focus). Secondly mystery is a relevant attribute concerning the perceived discrepancies or inconsistencies in the player's knowledge. Because mystery is a very individual, intrinsic, subjective characteristic- it is not treated here, hence our focus lies on the objective, extrinsic attributes of games. All are outlined underneath.

Game reality: Representation and Sensory Stimuli

Representation is about the precision of the reproduction. Representation concerns the physical and psychological similarity between a game and the environment. Firstly, the physical fidelity is how accurately the game replicates the real world context. Secondly, psychological fidelity is highly relevant for the use of SG, because it should safeguard that players use the same cognitive processes of task completion in the real-world that they learnt by gaming.

The sensory stimuli quality should offer vivid visual, auditory and tactile effects that can be used as feedback for performance.

Challenge

A game should not be boring, rendering the player frustrated. An optimal game challenge level is neither too easy nor to difficult. The optimal level should match the player's skill abilities to reach a pre-defined goal. Then the right amount motivational tension is produced. Challenge is linked with intrinsic motivation and 'effectance'-motivation*. Furthermore, "motivation is maintained by creating uncertainty about goal attainment. Uncertain outcomes are challenging because of the variability depending on the user's actions, multiple goals, hidden information, and randomness*"(Wilson, 2009, p.232).

Control

Users should be able to influence the elements of the learning process (e.g. the type of feedback received, content navigation, and the pace of the game). If users exercise control, they invest more time and apply more complex strategies.

Feedback

Describing effective training methods,Bluestone (2013) manifests feedback as important driver for outcomes and skills development.

According to Wilson (2010), feedback takes place as "something in the game changes as the result of the players action".(Wilson, 2010,p.233) Feedback belongs to the assessment part as game attribute. Players get a immediate negative or positive feedback based on their progress and their chosen decision path. Based on the feedback, "players learn from their actions and adjust their performance accordingly "(Wilson, 2010, p.233).

One other researcher defines feedback as performance information, which gives users indications about their status and ability level. Feedback is kind of a road of progress, in which players spot and remedy their failures (Wattanasoontorn, 2013, p.235). Feedback can be given in audio, visual or haptic form.

Sanchez describes two kinds of feedback. The outcome feedback indicates the performance level solely based on data. The process feedback specifies the performance regarding the chosen strategy and need of guidance by an expert or teacher.

The best down-to earth definition of feedback is: ***"***(that of a)...***central mechanism by***

which learners can regulate their own performance and understand how to improve”(Wilson, 2010).

Types of feedback

Wilson (2010) presents 3 forms of assessment, namely completion assessment (Was the task completed? With the right answer?), in-process assessment (What steps players go through to complete the task?) and teacher evaluation. In the in-process assessment the chosen decision path is more important than choice of the right decision result. Regarding the teacher evaluation, the teacher should be a facilitator who provides guidance and support.

Garris (2008) portrays two forms of feedback. The first one is system feedback, which gives knowledge about the results. The second one is debriefing stage that represents the key link between the game cycle (experiences) and the realisation of the learning outcomes.

Link with learning outcomes and motivation

Modern theories of learning state that learning is most effective when it is “active, experiential, situated, problem-based and provides immediate feedback**” (** Connolly, 2012, p.661). Learning is only effective and really achieved, when learners review their learning outcomes and reflect upon them (Guillen-Nieto ,2011). The bottom line of effective learning is based on the Dewey equation: **: experience +reflection = learning, stating that experiential learning must always be matched with appropriate learning support.**

Several other researcher reason that feedback is a critical learning support tool (Garris, 2008 and Wooley, 2007).

Players need to make the link between their action and the outcomes to improve their performance or learning. (Wilson, 2010). The learning preferences and capabilities of the individual learner must be matched with the demands of the learning task in order to enhance the learning effectiveness (Kolb, 2009, p.318).Garris (2008) underscores the pivotal and strong importance of feedback in regulating motivation, because it is a critical input in the judgement-behaviour-feedback cycle. Motivation is regulated by constant comparison between the actual performance levels with set standards (goals). If the feedback indicates that the performance level is constantly achieved, then the game is perceived as too easy and motivation declines. But if the feedback states that the current performance level is below envisioned standard, then the motivation (efforts) increases, because of the general human tendency to avoid any discrepancy, in this case the this feedback- standard discrepancy. For this to be happen, the goal of the game (=learning outcome) must be clear, so that the learner is committed to this goal and can be focused his attention of the goal-relevant task.

Especially, debriefing is judged to “*to be most critical part of the simulation/gaming experience”* (Garris, 2008, p.454). It is the link between the game cycle and the attainment of the learning goals a.k.a. outcomes. Debriefing can be defined as the constant review and analysis of game event. By debriefing, the players draw analogies between his gaming experiences and the real-world instances:

“Learning by doing must be coupled with the opportunity to reflect and abstract relevant information for effective learning to occur and for learners to link knowledge gained to the real world” (Garris, 2008, p.455).

Concluding, Sanchez points to the lack of research regarding the question how feedback can be best integrated as a learning tool in a game.

PROCESS Dimension: learning background and process

Difficulty of grasping of learning

"Human learning is distinguished by the range and complexity of skills that can be learned and the degree of abstraction that can be achieved" (Meltzoff, 2009, p.1). He argues for a new science of learning that is based on principles drawn from psychology, neuroscience, machine learning and education. He presents three principles of learning, namely that learning is computational, social and supported by the neural-cognitive system.

Another author defines learning are all non-directly observable processes in an organism, especially in the central nervous system (brain), that are conditioned by experiences and which result in a relatively permanent change and extension of the behavioural repertoires (Gröhlich, 2007, p.12). Gröhlich distinguish three learning dimensions, being knowledge, skill learning and life learning. The first is a cognitive, reflexive experience process that leads to in-depth knowledge about a thingness. The second is about practical, technical knowledge that is learned via experience. The last is about how to survive and cope as human being. To sum up: "*to learn means to gain or change knowledge, understanding, skill, habit, or attitude by instruction, study, observation and experience*" (Wenzler, 1999, p9).

There exist several learning theories how people learn in general. Due to their clear line of thought, I chose the input –output model of Garris and the ELT-model of Kolb to shed more light on the learning process and gaming. The goals are to combine them as a theoretical effort to arrive at comprehensive model of learning by games. Connolly (2012) highlights the interdisciplinarity of the SG-area and that many theories are involved and overlap. For example, knowledge transfer is a subject of learning theories, but perceptual and cognitive outcomes are rooted in cognitive theory.

Regarding this thesis, learning is conceptualize as process, in which experience is transformed into knowledge. (Kolb, 2009). In the next section, the learning theories of Garris (2002) and Kolb (2009) will be merged into one conception model.

The generational impact

Learners are fed up with learning via traditional, authoritarian, in-front way, where learning content accentuates the learning of facts (Pannese, 2007). New forms of learning which are interactive and democratic conquer the education landscape. Interactive learning (case reports, technology assisted, problem-based and open discussion) is more effective than traditional learning. It yields a higher student satisfaction, a deeper learning and better knowledge retention (Sandhu, 2012). SGs are the perfect training tools to transfer knowledge and skills in the competency-based training systems of medical postgraduate education in the Western hemisphere. (Graafland, 2012)

Furthermore, there are generational differences between students of the millennial generation and their educators that are of importance. This generation prefers instant communication, feedback, knowledge and multi-tasking and is adept in virtual reality. (Sandhu, 2012) Games must take account of those differences

Global Game Learning Model: Input-process-outcome model by Garris (2002)

The input-process-outcome model by Garris (2002) can be seen as extension of the activity theory model, linking games and learning. Instructional content and game characteristics are the inputs; the game cycle is the process as a repetitive judgement-behaviour-feedback loop and lastly the output are the learning results. Concerning the inputs, dimensions for the game characteristics are fantasy and mystery, rules, sensory stimuli (video and audio), challenge and active learner control. A debriefing stage is essential to serve as self-reflection effort and catharsis. Debriefing is the link between the simulation experience and the real-world. Learning outcomes can be of a skill-based, cognitive (declarative, procedural, strategic) and affective (behaviour) nature.

Experience learning theory

Because learning via games is learning via experience, the model developed by Kolb (2009) suits our study very well. Experience has since the work of Kurt Lewin a central role in the learning process, accentuating the radical empiricism as philosophical foundation. The experience learning theory (ELT) defines learning as process whereby knowledge is created through the transformation of experience. Grasping experience occurs via the concrete experience and abstract conceptualisation mode. Transforming the experience happens via reflective observation and active experimentation mode. Hence, there are four learning modes. The ELT represents a recursive cycle where concrete experience is an input for the observation and reflections. These are associated with abstract concepts, which in turn have implications for actions. When the activities are executed, new concrete experience is created. This process is responsive to the learning context and content. The ELT- models offers meta-cognitive strategies to enhance the individual learning effectiveness.

In the following section, only the learning game attributes and learning outcomes that are assumed to be critical for the surgical training will be addressed. This is an arbitrary decision to avoid a too extensive discussion of non-relevant attributes (like fantasy and mystery) and learning outcomes (affective and communicative).

Outcome dimension: the learning outcomes of Serious Games

As Kolb, Wilson defines three kinds of learning outcomes: cognitive, skill-based and affective ones.

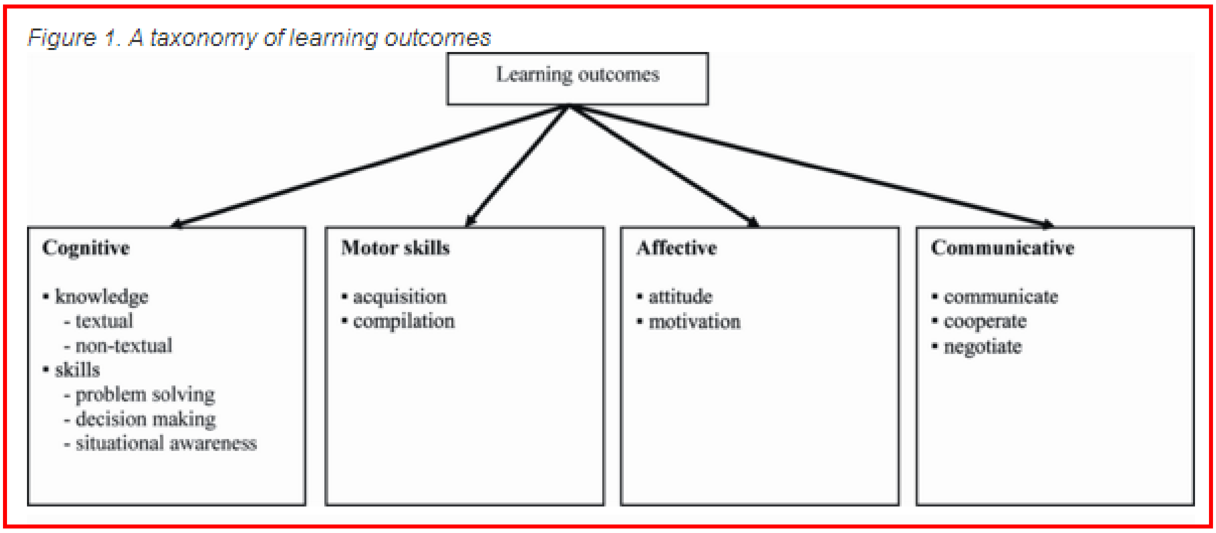


Figure 2: Taxonomy of learning outcomes

Source: http://www.igi-global.com/chapter/games-based-learning-advancements-multi/18798

Cognitive learning outcomes

There are three types of knowledge users can adopt. *Declarative knowledge* (i.e., knowledge about what), *procedural knowledge* (i.e., knowledge about how), and *strategic or tacit knowledge* (i.e., knowledge about which, when, and why) .These three types of learning outcomes describe the cognitive process of learning. Once the knowledge has be acquired, the learner has to organise the content. *Knowledge organization involves grouping meaningful pieces of information into mental models, which are stored in long-term memory for later recall.* When the knowledge is actually applied, cognitive strategies are developed. This stage involves meta-cognition and self-insight.

Skill-based learning outcomes

This is about learning and mastery technical and motor skills. Skill-based outcomes are attuned to goal achievement and *“the systematic organization of behaviours in a sequential and hierarchical manner.”*

(Psycho)Motor skills

There are several types of those skills, starting with a simple nature and progressing towards more complex skills.

1. perception (i.e., using sensory cues to guide motor activity)
2. set (readiness to act)
3. guided response (i.e., imitation)
4. mechanism (i.e., exhibiting habitual movement patterns)
5. complex overt response (i.e., exhibiting proficient, habitual movement patterns)
6. adaptation (i.e., modification of habitual movement patterns to meet a special need)
7. origination (i.e., creation of new movement patterns to meet specific situations.

Along this spectrum, the skills become more elaborated and more automated, applicable to increasingly non-standardized, variable tasks.

Linking Serious Games and Learning: the ambiguous relationships between Game(s) (Attributes) and Learning (Outcomes)

In general, games are a powerful learning environment because they provide:1) a support multi-sensory, active, experiential, problem-based learning, 2) favour activation of prior knowledge, 3) immediate feedback and lastly, they represent a kind self.-assessment and 5) social environment(Papastergiou, 2009).

Wilson complains that although the fact that researchers know a lot how learning takes places via games, there is paucity of research that links games attributes and learning outcomes. There are two streams of research, one that is stating that games offer a 'accessible, low cost and effective' way of learning, whereas others doubt the in-depth effectiveness, stating that games has only superficial learning effects. Besides, learning via games does not fulfil the learning requirements of students. (Wilson, 2009). In general, research has demonstrated in the last decades (1960s onwards) that games improve learning, motivations and performance.
Wilson warns that games are not effective universal tools that are suitable for all learners and learning situations (Wilson, 2009), but games need to be adapted to the user characteristics and the learning objectives (= content)(Sanchez, 2010).

Wilson also mentions that researchers *" do not know if the relationship between games and learning is direct or indirect, and if so, what the mediating variables may be. We must also understand whether a single game attribute leads to learning or if a combination of multiple attributes within a game has a stronger effect. Many areas of research remain unexplored*" (Wilson, 2009).

Conclusion regarding the learning Outcomes

- Matching the desired outcomes with the game attributes, or rather selecting the game attributes to produce a desired outcome, is a difficult task.
- **Learning outcomes can be considered indicators for evaluation methods.**
- Instruction should be designed in a way to achieve desired learning outcomes. One way to assess this is through measures that detect the desired learning outcomes.
- Any game that is designed for instructional purposes should be heavily linked to instructional objectives (Hays, 2005).

Concluding, though many authors (e.g. Girard, 2012) are convinced that games have a positive effect on learning, it is very challenging to enlighten how learning works by gaming due to the complexities and ambiguities of the learning dimension. But nevertheless, two authors (Graafland, 2014 and Van Dijk, 2014) corroborate the sound learning effectiveness potential of Serious Games. In general, Graafland (2014) states that SG "*should be regarded as an acceptable modality in (surgical) training*" (Graafland, 2014, p.202). In particular, Van Dijk (2014) specifically links game attributes with learning elements/outcomes. His findings indicate that SG are more effective as a learning tool (than a presentation) to teach people a certain behaviour. The opportunity to self-experience failure, the feedback loop, the ability for users to be in self-control of their learning process and being able to be consumed by their own storytelling, competition/challenge, social endeavour of team play- all are very important game attributes that contribute to an effective learning outcome, meaning transfer of knowledge. Especially the built-in debriefing stage as self-reflection endeavour was pivotal for users to ponder on their own experience and behaviour during and after the gaming project.

Remark: The instructional effectiveness is ultimately assessed by testing the clinical competence of medical trainees. The competences and activities of the medical curriculum are presented and a presentation of Maslow´s pyramid as heuristic tool to enable the assessment of clinical competence are described in Appendix B. Furthermore, Appendix B deals with validity issues of SG, which is a prerequisite that SG are accepted in the medical curriculum as proofed, fully validated tools. Only content validity (i.e. game content adequately covers the dimensions of the medical education construct) and face validity (i.e.degree of similarity between medical construct created in gameplay and in reality) are relevant concepts for the range of this thesis.

Summary; In this chapter I defined SG as *""any computerized game whose chief mission is not entertainment and all entertainment games which can be reapplied to a different mission other than entertainment"* using the definition of the 'SG-guru' Ben Sawyer, presented the intended and unintended benefits of SG, displayed the relevant game attributes for medical education (representation/sensorystimuli,challenge/control and feedback(process/outcome/debriefing). Moreover it was concluded that it is difficult to grasp the learning process. The experience learning model of Kolb will be used as abstract model to illuminate the learning process via games.Lastly, learning outcomes were classified in cognitive, skill-based and motor-skills ones.

Chapter 3: Research design and methodology

In this chapter we describe the followed qualitative approach. The purposeful sampling strategy is described as is the use of the semi-structured interview as data collection method and the open coding procedure. Lastly, the interview framework is displayed.

The motivation to use a qualitative research design (exploration approach), the post-modern philosophy of knowledge, the unit of data collection/analysis, the purposive, non-probability data sampling and information about the data analysis and nature of qualitative data is given in Appendix C. Approximately 40 potential interview partners were contacted via e-mail and phone and asked for participation. Due to natural attrition, having a too heavy workload (3), no time and interest, interviews were done with 8 remaining experts (instructor/educator side) and 2 medical students (user side), Anonymity was promised to enable a safe, respectful and trustworthy talk atmosphere. This is not a limitation, because access and contact with the respondents can be made via the researcher.

Data collection method: interview

The interview method is chosen. The researcher is the main instrument for data collection in the interview method (Boeije, 2010). The interview is a conversation, where the researcher pose questions concerning the behaviour, ideas, attitudes and experience with regard to phenomena and interviewees provide answer to those questions (Boeije, 2010)The purpose if an interview is to get ‘thoroughly tested knowledge’(Kvale, 2009). An interview framework is used to enhance reliability. A semi- structured type of interview will be used and the interview forms are of factual (to gain valid factual information about the research topic) and conceptual (conceptual clarification as point of interest) nature (Kvale, 2009).

Data collection problems

Data collection is a selective process, because it is not possible to capture a get-it -all perspective. There are 3 problems inherent to qualitative data collection. Firstly, the selectivity (bias of researcher, informants are selective-they hide certain behaviour or perception, and nonverbal information is omitted). Data overload (massive amount of notes) and data retrievals issues (figuring out the most meaningful material).

Coding

Coding is a core analysis activity as raw data is transcribed into a synthesis of meaningful parts (Miles & Huberman, 1994).

Codes are tags that assign meaning to the descriptive or inferential information that is collecting during the research. Those labels are used to retrieve and organize fragments of words, sentences and paragraphs. One can use a straightforward category label or a complex one (metaphor), although one must be aware of the embedded linguistic logic or conceptual lens that works at a subliminal part.
According to Boeije (2010), the researcher distinguishes a theme or categories in the data and names them by attributing a code a label that depicts the core topic of the segment (p.95).
“The organizing part will entail some system for categorizing the various chunks, so the researcher can quickly find, pull out and cluster the segments relating to a particular research question, hypothesis, construct or theme. “
Clustering and display of condensed chunks-sets the stage for drawing conclusions

Types of codes

Firstly, there are descriptive codes that involve some interpretation as the researcher is “attributing a class of phenomena to a segment or text”(Miles & Huberman, 2014, p.56). Interpretive codes are linked to hidden agenda issues like motives, power-and team dynamics, private vs. public posture). Pattern codes are more ‘inferential and explanatory’ regarding some leitmotiv, pattern, theme or casual link.
Code may represent deeply theoretical or analytical concept or it could be practical or descriptive (Boeije, 2010).

Code Characteristics

Firstly, codes involve different level of analysis, ranging from descriptive to inferential. Different codes are created at different analysis phases and lastly codes are astringent as they unify disparate material, enabling analysis.

Creating codes

Miles and Huberman recommends the use of a preliminary, well-structured list of (generic) codes based for example on the the conceptual framework, research questions, problem areas or conceptual variables (Miles & Huberman,1994,p.66).

Boeije (2010) differentiates between open, axial and selective coding. Open coding is the process of breaking down, examining, comparing and conceptualising and categorising the data. Open coding encourages a thematic approach since it forces the analyst to break up the text into pieces, contributes to a clear organisation of data, as it results in an indexing system (code list of tree), a coding scheme.

Open coding is applied in our research as we follow a deductive approach in which "coding that strives for specific and immediate (objective) comments or specifically identified outcomes with the aim of practical understanding"(Boeije, 2010,p.100). Based on the theoretical framework, the researcher knows what he is looking for, because the themes and categories are explicit and represent a code. The literature review, existing theories and theoretical concepts illuminate the labelling process, resulting in constructed codes. Those concepts correspond with the codes, which built upon the a-priori theoretical conceptualisation.

Moreover, sensitizing or guiding concepts can be used as a code. These are global notions and ideas that are based on a skeletal framework (outcome of literature review). These are concepts which have not been yet fully formalized (Boeije, 2010, p.23)

The **interview framework** addressed 12 themes of interest, which covered:

- The definitions of SG;
- The differences between SG and Simulations;
- The best method to train surgeons;
- SG used at the medical faculty of interview partner;
- The perceived advantages and disadvantages of SG;
- The Perceived problems with SG in general and regarding the training of operation in particular;
- The R&D costs and process for a soundly designed SG (3D, VR, Edutainment);
- The most relevant game attributes: importance of fun/flow, feedback, debriefing;
- The recommended learning model for SG;
- The suitability of SG for new generation of learners;
- 3D-VR tools potential for SG;
- Arriving at a final conclusion: SG as effective tool for medical education.

Chapter 4: Findings

In this chapter, the results of the 10 interviews are comprised, after applying a open coding and integrative procedures during the analysis stage.

Results based on the interviews

8 experts and 2 medical students were interviewed and 2 students with a (medical) education background: Expert 1 wrote his phd-thesis (2015) at the UvA on SG and surgical education. Expert 2 wrote his phd-thesis (2010) about anatomical virtual learning and simulating training and is manager of such courses at the Radboud University. Expert 3 works at the LMU as SG and medical education professional. Expert 4 is a lecturer for virtual learning on medical education (University Heilbronn and Heidelberg). Expert5 writes his phd-thesis at the University of Mannheim about universal applicable SGs where the learning content is inserted by the teacher. Expert 6 is responsible for the 3D virtual tables and medical education at the anatomical institute of the University Heidelberg. Expert 7 has a background in psychology (emotion focus) and instructional design of serious games and works at the University of Ulm.Expert 8 is chief operating officer at INSIMO, a French simulations software provider for organs, tissues and surgical procedures, the design is based on biomechanical modelling.
Furthermore, two medical students in their final year (10 semester) from the University of Heidelberg were questioned.

The key themes (green colour) were addressed during the interview sessions, which took approx.60 minutes with the experts and 10-20 min with the students. The most important findings are reported per key theme. The findings are based on Appendix E after applied an open coding procedure.

Serious Games Definition

4 of the 8 experts agree that there is no universal definition about serious games. Also 4 out of 8 experts' stresses that a SG should have learning aim as primary use, not entertainment.The core element of Serious Games is the video game character to enable professionals to learn serious knowledge, skills and behaviour, which are necessary for their job (Expert 1). A convergence is taken place between simulations and serious games: games use simulation elements, simulations use game elements

(Expert 2). For Expert 3 the gamification and edutainment approach is key input for SG. Every SG should have a learning aim, but contrarily the entertainment factor should prevail. For Expert 5, SG are digital games (not necessarily 3D) and their primary function is not entertainment, but the learning aim (training, awareness) should play the major role.SG can help to convey contents, change behaviour or assess the learned content.(Expert 7). For her, the classical learning goals of SG are cognitive and skill-based learning. Expert 8 as simulation expert and both students had no knowledge about the SG-concept.

Difference between SG and Simulation

For 4 experts (about that respective theme) mentioned that SG and simulations are an overlapping area, a gray area (Expert 1), as already mentioned a convergence is happening between both. Simulations are designed to learn one thing (action) thoroughly, but the fun part is missing, hence the trainee only uses the simulator because his boss orders him that or it is assessed learning content (Expert1). In surgery training, simulators are used in case of high-stake, expensive technologies. Simulators as system give no feedback towards the student, hence he does not know how good he is (Expert2). Gaming is suitable to learn factual knowledge or procedures (Expert 4). The fun part of games is also stressed by Expert5.Either application should make sense, as SG are not ubiquitously meaningful (for all learning content,target groups, settings), according to Expert 5. Expert 7 classifies simulations under the umbrella term SG, because most simulations have a gamification component such as ranking, score, play goal. Everything depends on the learning content: if realistic procedures are addressed, use a simulation- if factual knowledge is demanded, use a SG (e.g. Adventuregame). Expert 8 is the simulation expert. Simulation can help to experiment how to execute a task or activities in different ways, while collecting a big set of relevant data. It is not aimed to gain knowledge, but to maintain the necessary skills. Moreover, the possibility to train skills for rare procedures with a broad range of anatomical variability and complexities, creating a huge set of relevant data (recording of whole session) are pros of the simulator.

SG used at faculty or are known of

All six experts have knowledge of 3D virtual learning applications. Exert6's interest area is the 3D Anatomy Table, which is a simulation but game elements are incorporated (experimentation, fun, challenge, quizzes). Both Dutch doctors (Expert 1 and 2) mentioned the Underground and Geriatrics games, that represent adding learning possibilities at the medicine track of the University of Amsterdam and Radboud. In Germany, the UroIsland Game used at the University of Freiburg and the PatDocTalk game at the University of Cologne are renowned (Expert 3 and 4). Expert 5 develops knowledge game which can be used universally (learning content, target group, settings), but the adapted, specialised learning content is add by the lecturer/teacher. In Germany, serious games are not universally used at medical departments, because many medical education planners have no knowledge about SG (and their potential) (Expert 3 and 6), hence knowledge diffusion of SG happens mainly via international medical professional conferences (Expert 6: First acquaintance with Anatomage Incorporation tool at American Assiciation of Clinical Anatomists Conference).

Perceived advantages of SG

All experts agree that a SG provides much fun and motivation as driver for learning represents a core advantage. SG should offer a added value if they are smartly designed (Expert1). They offer a safe environment to train high-stake actions. The Feedback is rendered in an neutral approach without any assessment bias from colleagues (charming-factor) (Expert2). Expert 3 underscores the high motivation, quick feedback and the high immergence factor of games-important to attain a flow-status as the key advantages of SG. SG are especially suitable to change behaviour/attitudes as SG arouse emotions (Expert3). According to Expert5, SG are especially practical to learn factual knowledge. (Expert 4). SG are better suited to tap new target groups (difficult learners, younger generation) due to their interactive design. Another convenience is that the game adjusts itself to the knowledge level and learning speed of the learner (Expert 5). The potential of a smooth transition/generalisation to the practical reality (hospital) is stressed by Expert 6, because skills and knowledge that are learned are also an essential part in the professional daily routine. For expert 7, SG has a motivational component, but only if the game designs correspond with the target group needs and their living world. For both students, having the opportunity to train critical skills (e.g. intubation) in a game situation is a big pro. They conceive it as preparation for an emergency case, as learning without failures (death patient), while holding a relaxed attitude (emergency case: high stress level).The time flexibility, broad applicability and lower price (compared to Dummies, Simulations) are further benefits. Expert 4 agrees with the lower price criteria because SG require lesser human resources (compared to dummies or hybrid systems).

Perceived disadvantages

3 of the 8 experts (1,4,5) mentions validity problems. Games inputs and outputs are very hard to measure and to validate (Expert 1) and there is a lack of longitudinal studies to measure the effectiveness of games.Expert 4 and 7, the focus on fun can be detrimental. Fun is not enough to guarantee learning, but the learning content must always be learnt. (Expert4). For Expert 7, too many players focus on the gaming, hence fun part, trying to achieve the goal of the game by applying a trial-and error approach leading to success. But the aim of the game is to transport the learning content, this is not achieved because the player does not think about or understand them. Moreover, the high development costs of appealing, effective games can be a barrier (Expert 3 and 5).Expert 3 and 7 state the high infrastructure (e.g. high speed internet) costs, huge data amount of data volume and high quality, demanding programming skills as potential disadvantages. Expert 2,6 and MS1, mentioned the high time investment needs as non-beneficial regarding the writing of learning content material, the training of other trainers as multiplicators and the general preparation time (E1). For expert 1 and 6, there is a acceptance and knowledge barrier. There is no exchange of knowledge domains taking place between medical educators on one side (domain: learning contents) and game designers on the other side (domain: sound game design (E1). Moreover, there is a institutional, professional and personal fear regarding the introduction of a new training system (E6).Expert 6 gives also the lack of learning content standardization as entry barrier. Lastly, the potential normal risks of games such as distraction and addiction needs to be considered (E5).

Problems with SG regarding the training of operations (overlapping with disadvantages)

Expert1 and 2 reason that it challenging and difficult to train a whole operation, because this is too complex: “It is very difficult to simulate a whole operation by combining it with a videogame.”- it is better to use partial task simulators. (Expert1).

Experts2 advocates a sound task and stakeholder analysis in the development/implementation stage of a game. It is problematic that game producers aim to design highly visual representation/accuracy, but this aspect is not what user always needs. Hence, the producers accentuate a wrong dimension (high visual accuracy instead of user inclination).Expert 3 and 6 mention the acceptance barrier regarding new learning methods- on both sides: older lecturers and students who hold a pessimistic view towards VR, digital tools. Lastly there is also a knowledge diffusion barrier because the practical point ('Lehre) has until know no high valuation in Germany (as opposed to NL/USA). Here the lecturing part is more important as lecturers are evaluated how they perform in the classical setting (frontal, crowd lessons).Only in the last years, the practical side is more accentuated as the Master of Medical Education can be done at 5 main universities based on the Harvard/Bern model (Expert4). Furthermore, games can be a have a high technical expectancy and requirement. Contrarily, if mobile apps are used, there is a limited area to display the relevant content and to give limited feedback in a mostly textual form. Games should always be used in combination with other learning tools. Moreover, there are validation problems as many components contribute to learning, e.g. personification effects (individual differences: learner type and learner requirements) have great influences on the individual learning effectiveness. For example learners with a high beginning amount of interest, it is no so much important how the learning environment is designed. Lastly, student2 states the incapability of games to foster complex, deeper understanding instead of just learning facts by rote. Also there could be individual problems with 3D thinking and positioning.

R&D costs for sound SG (VR, 3D, entertainmentEducation)

There was no agreement on how much a good quality SG cost to develop. If the game should be 3D, VR and of high quality, taken commercial of the shelf games (like EA-games) as gold standard, then the budget will be in the millions € range up from 1,5 mio € to 16 mio € (Expert1, 3,7 and 8). Expert 1 and 2 agree that for a budget of few hundred thousand € a soundly designed, effective game can be developed. Expert 6's figures corresponds with that (60 000€ for 3D-VR AnatomageTable) For Expert 4 and 5 highlight the huge saving potential by developing platform games (universal knowledge games and special knowledge and skills games used at medical departments).

But all expert stress that an 'one game fits all'-approach is difficult, as every educational setting has its own embedded, in-vivo reality with special learning aims and target groups and power and team dynamics.

Best method to train surgeons

3 of the 8 experts (3,4,6 and student 2) agree that the dissection will remain the standard to train anatomical knowledge in the future. For expert 2 animal vivisection is best training method because the haptic texture and blood flows are very important. Expert 1 favours a partial task simulator in which only one task is simulated because to simulate a whole operation would be too complex and not being cost efficient or representing effective learning. Situational awareness training can be providing by videogames. Expert 3 and 4 advocates a mix of education methods, because the learning and training of practical skills is highly constrained by using technology and software alone, hence some kind of vivisection, dissection training will always be necessary (E4 and E6). Expert 3 recommends dissection for anatomical learning, SG for procedural knowledge and simulations for motor skills. Student 2 also stresses highly the advantages of live-dissection, being the high complexity in general (human tissue structure, particularities, and subtleties of every human posture). The games or simulations do not achieve such a high complexity/ fidelity level, giving like books only a constrained schematic representation of the items. For expert 8, only simulations can provide the huge range of anatomical variability, complexities and representation of rare cases in general.

Most relevant game attributes for training of surgeons

3 of the 8 experts (E1,2,7) mourn that game attributes are too complex, abstract and theoretical to be assessed and to be proofed. There are simply to many constellation possibilities concerning the matching of game attributes and learning outcomes (E7).Challenge is stated most often by 4 of the 8 experts (E3,5,7,8). Via active control the learner can influence the challenge level (E3). Surgeons in general have an inclination for competition and challenge (E8).

Representation is named by expert 1 and both students as highly pivotal. A high fidelity is more relevant for 6th year students, than for 1st year medical students, because 6th year students demand a more precise level of detail that fits with their advanced knowledge base.(E1). Having a reward system to propel incentive is stated by E1.Feedback is stated by 2 of the 8 experts (E3 and E7) as essential, flow by expert3 as factor for immersion and informal learning (E3). For E3, fun is the most important driver for learning and motivation.

Principles presented in classical Game Design books provide a good guidance how to design a appealing game (E5).Lastly, expert7 points out that the relevancy of game attributes highly depends on the learner profile (e.g. appeal and influence of colours) and learning content.

Importance of fun and flow factor

4 of the 8 (E1,3,5,6) experts agree that fun is a very pivotal element to promote learning. Expert6 acknowledge that learning theories proofs that fun increases the learning effect and makes learning easier. Expert4 plays the role of a daredevil by warning that fun is not self-sufficient for learning, but a respective game has to be incorporated in a formalized curriculum. In case of content that is not exam-relevant 95% of students do not take part in the game if it is voluntary and a free ride. Expert3 highlights the importance of the immersion factor to attain a flow state, which is created by a perfect balance of requirements ('Anforderung') and creativity level ('das Schaffen). Flow is among other things dependent on the level of the trainee and the context (training content). The problem is that individual inclination to attain a flow status: some achieve it, others not-hence flow is always influenced by a specified goal, situation and the individual posture(Expert 1).

Importance of feedback: Visual,audio, haptic

All experts and both students agree that feedback is a very important input for learning, with visual feedback as the most commonly used form: *"learning without feedback is difficult"*(Expert4). In every game the player should know what he does and when, hence it is critical that he receives any feedback. There is a small feedback feature (press key, light on key) and a big feedback feature (total(system) feedback that highly influences the learning effect (Expert5 and 1). The maturing level of the trainee is important: for junior students visual feedback is important, but a senior student needs more haptic feedback to train the touch dimension (Expert1)

All games have visual feedback, audio is of additive value, but haptic is only meaningful for special cases. (Expert5).

3 of the 6 (Expert 1,3, 5)experts say that the importance of feedback is highly dependent on the learning content and aim.

.According to Expert2, feedback in general is highly essential; the modality of feedback (how feedback is delivered) is of secondary interest. What matters is that feedback is well timed and specialized.

To learn motor skills, haptic feedback can be given via WII-like motion sensors, which immediate react upon the actions of the player. Virtual feedback is the best form to learn cognitive and procedures: if a patient suffers a virtual death, the player immediately knows that he has done something wrong. (Expert3)

A learner should receive feedback on what and when he does something and that he is informed of doing the right thing. (E5). E7 plays the daredevil by stating that feedback concerning SG are not well-researched, often being approached by theoretical approaches. She pinpoints the role of feedback in case of failures, leading to extra frustration. Also giving feedback to often is frustrating.E7 also highlights the importance of self-reflection concerning the gaming experience as a key form of feedback: by doing self-reflection players adapt themselves to a given challenge level. Moreover, the timing of feedback is essential (at which point (of failure). Negative feedback can also be a driver to accomplish something. The common proclamation:"in case of positive emotions and feedback, one learns better" is not universally correct. Feedback is a complex issue, especially in connection with emotions and learning, being very differentiated in the learning process and the results (E7). For expert 8, feedback are just a parameter in the data set computed by the system. Positive and negative feedback are created by creating respectively positive or negative consequences.

Outcome or processfeedback

5 of the 8 experts (E1,3,5,7 and 8) and 1 students underscore the importance of process feedback.

.A player needs to get a score to know where he stands (outcome), but to know that he becomes better is a flow factor (Expert1). For expert 2, the learning stage is a factor to give feedback: a beginning learner without knowledge references needs more often feedback what they do (outcome) and how(process), whereas advanced learners need a compact feedback at the end of the task (E2). Processfeedback is simply more meaningful and concrete The importance of a high score is overrated, because it is the social comparison that matters (how good is the player compared to his peers) (E3). The player should also be given the opportunity to experiment without receiving any feedback (E5).As technical people, expert 6 and 8 stresses the system feedback component. Simulators record every parameter of the session, hence both kinds of feedback can be used as reflection and debriefing input (E8).

Importance of debriefing stage

7 of the 8 experts and 1 student state that a kind of human debriefing is very valuable:" Debriefing (and feedback) are very important for reflection and knowledge gain"(Expert4). Especially for training games,Expert5 advocates the use of a debriefing stage in any case (if that is feasible).

The human factor is very important. The lecturer should inform the student about his study progress and skill development (Expert2). Expert3 thinks that some kind of social feedback (either tutorial or at least online discussion group) is terrific to enable a valuable peer-to-peer exchange. Above all, debriefing can be best given at the end of a session, so the gamer has a compact overview of his performance and improvement gaps. (Expert 5)

2 experts (2 and 6) illuminate the possibility of technical feedback data, that the system creates automatically. The system gives a very detailed performance feedback at the end of the session. Here, the administrator/lecturer is not involved feedback statement.

2 experts (E2 and E8) and 1 student highlight the concept of a total feedback, one produced by the system (insided) and one by the instructor (outsided).

Expert 1 is very sceptical of the debriefing concept as a myth because learners can become better on one skill by playing a completely unrealistic game (e.g. surgeons exercise by using WII). The learning effect occurs, because certain brain areas are triggered. In such a case, debriefing would be 'utterly useless'. Hence the need for debriefing relies on the context, means and target group (E1). Expert 7 recommends the use of a pre-debriefing stage to highlight the important issues of a game. Debriefing is especially pivotal for the tutor-based approach in medical training.

Recommended learning model for Serious Gaming

6 out of 8 experts agree that Kolb's model is suitable to depict the learning models of games, because it highly values the pivotal role of experience. Expert2 is very sceptical using this model as it is not making any sense and was never empirically proven. He recommends using instead some model of skill development (deliberate practice model of Erikson or skill automation level). This makes more sense as one starts with partial skills (conscious execution) and over the time one repertoire,

enabling an automatic, unconscious execution. Expert3 advocates the social-cognitive learning theory of Bandura (eventually linked with Kolb's model).Both students would favour a apprenticeship-or tutored model, because one always needs a surgeons/medic who guide the trainee.

Better suitability of SG for new generation of learners

5 of the 8 experts (E1,2,3,4 and 8) agrees that serious games are matching the expectancies and demands of new generations of learners ('smartphone'-generation). Expert 3 mourns about the fallacy and inefficiency of ex-cathedra teaching: 500 students sit there, shutting up and have no opportunity for questions or answers.

A pessimistic viewpoint is taken by experts 6 and 7 and student1: frontal, mass class will in some instances be necessary (anatomic course with 500 students) and learning facts by rote remains a matter of fact. What counts is the transfer of knowledge and that is a question of the mediating learning tool (E7 and S1): "*Not everything is medial in the sense of digital*", as E7 frames it. If games offer added learning value, they should be used. (E7 and S2). Lastly, E8 mentions that also older (eye) surgeons are fond of using the high fidelity simulator (E8).

Future of SG based on 3D-VR tools (OculusRift, Samsung VR, Hololens,etc)

7 of the 8 experts hold a very critical and pessimistic viewpoint regarding the potential of those tools. To make an appealing application, one needs a huge budget (E1). The augmented reality potential for the use in clinical settings is stressed by E2. But he also that technology is not a redeemer in itself, but one needs to know what are the challenges of a technology, before applying it. Moreover, human interaction is difficult to imitate-this remains a challenge for VR scenarios-because much non-verbal subliminal communication is lost (E2). For expert 3, the VR-glasses are too expensive (700-800€ per piece), hence a waiting period of two years is necessary to allow the mass acceptance of this technology. But it will remain a trend for the next 3-5 years. At this moment, the hardware and software is very premature, but the potential is huge (E4).E5 is more optimistic, because GameEngines like Unity already support OculusRift. The game designer only has more work effort to invest. E6 mentions the NIMBY-syndrome: educators want to include their own learning content, not using prefabricated content. For expert 7,the programming complexity is too demanding, the freedom of movement (bound to a cable) and the possibility of natural movements are too constrained, hence the environment feels too artificial. E8 doubts the immersion

potential of such tools and student2 is a proponent of live settings due to the logic: Seeing a natural artifact live, one can imagine it better.

Conclusion: Are SG an effective tool for medical education

4 of the 8 experts agree that SG are an effective tool for medical education, not as replacement of the traditional learning method (frontal class, ppt, books), but they favour SG are efficient supplement. Expert 1 highlights that a SG can be effective but only to a certain range of improvement. Expert 2, although he is very knowledgeable about the topic, he holds a pessimistic view, stresses that a final conclusion about the effectiveness of SG is not yet possible. Expert 3 accentuates a stakeholder approach to promote the awareness and acceptance of SG in medical education. By offering flow and immersion, SG can tap the ‘Pokémon-potential’ of every learner, even when interest and relevancy (content) is low.

SG can only be effective if they are used in connection with other learning methods Expert 4). Expert6 highlights the need of voluntary coercion, so that game participation is voluntary, but that the learning content of games is assessed in formal exams. Student1 also perceives SG as a good supplement tool to learn factual knowledge, whereas student2 thinks they are better suited to learn practical skill or partial-tasks of operations. He doubts that SG can help to holistically assess and understand complex relationships of given concepts. For the medicine domain, learning by rote will remain a matter of fact, which is not possible to circumvent. The key question is how the learning content can be best provided by a certain learning tool; this depends always on the learner profile, focal context and content itself (E2 and S1).

Comparison theoretical and practical insights

Serious Games Definition

For many experts there SG is an umbrella term and many mourn that there is no universal definition. The edutainment and gamification aspect of SG is stressed, but the weighted balance is a point of controversy: Expert 3 says that the entertainment factor should prevail, whereas the other say that the learning aim should play the major role. SG as umbrella term should even incorporate simulation that use gaming attributes. The definition of Sawyer fulfils this all-in-thinking, hence it should be used as universal definition to describe SG.

“Any computerized game whose chief mission is not entertainment and all entertainment games which can be reapplied to a different mission other than entertainment.” (BIT conference paper, 2010)

Difference between SG and Simulation

A convergence between both is taken place: there are serious games with simulation features (focus to learn one activity soundly, no process feedback, partial task, high-stake, no fun) and simulations that have gaming features (challenge, ranking/score card, fun, feedback, reward system). This convergence is not addressed in any article, but the happening can be seen in the names of a journal such as ‘Serious Games and Simulations’.

SG used at faculty or are known of

Very few SG are used in at medical universities in Germany, only UroIsland and PatTocTalk are known of on a voluntary basis. Instead simulations such as the AnatomageTable (3D anatomical simualtion on a VR-table) are used at 3 out of 32 universities. In the Netherlands, more SG are used at the academic medical centres, such as Underground, Geriatrix. ABDEesion or Medialis surgical mastering-these games are part of the medical curriculum. Moreover, France is also a leading innovator with the ‘ilumens’-system and incumbent simulator providers (like Insimo).
This is a new insight, as France and the Netherlands are established as leading innovators concerning the SG technology.

Perceived advantages of SG

The possibility of receiving feedback, attaining a flow-state and high motivation and immersion effect were given by respondents and also stated in the theory. SG were thought to be especially suited for factual knowledge learning (2D, Adventuregame), but not so much to learn high-stake, complex activities (better use of simulations). Games could be suited to learn partial-task, simple and not complex activities- this is a new insight. SG could target new groups of learners and offer a safe-risk free training environment, corroborating the theoretical insights.

Perceived disadvantages

The cost benefit of SG were not perceived as realizable, because it cost millions to develop a appeasing 3D, VR-game with a high fidelity, variability and complexity that matches the high expectancy of commercial off-the-shelf games (such as from EA, or ‘Need for speed’. This cost benefit is realizable if platform games are developed with a case database, but there are too many institutional acceptance barriers. This is in stark contrast to literature statements, as they stress the cost-effectiveness of SG, that are often hard to deliver. Overall, the potential disadvantages are overlooked in literature.

R&D costs for sound SG (VR, 3D, entertainmentEducation)

In theory, there was no cost estimation given. My guess estimation is 30 000€ for a 2D Adventuregame for knowledge learning, whereas a 3D-VR SG that imitates the high fidelity, high quality of commercial games (Microsoft flight simulator, Need for Speed-EA games) cost in the range 1,5-4,5 millions. The US army paid 16 mio $ for the development of a emergency management SG .Good games are very difficult to develop because a sound stakeholder and task analysis must be made, using a network approach, hence many agents, much time and monetary investment. An innovative approach, that saves resources, is possibility of a knowledge game platform, wherein a universal game (2D, Adventure) is provided), and the special learning content is added by the instructor.

Best method to train surgeons

Common to theory, dissection will remain the gold standard to train new surgeons, due to the haptic and visual (tissue) fidelity. SG and simulations are best provided as additive, supportive learning tools in an voluntary way. Vivisection is beneficial due to a working blood cyle and the direct action-reaction (cut=blood) logic.

Problems with SG regarding the training of operations

The main problem that was not mentioned in the theory was the acceptance problems of SG as learning tool, because there is a communication gap between medical educators and game developers, each not knowing about the others knowledge domain. The best driver for knowledge diffusion (about the SG-concept and their potential) is international conferences and incorporation in the curriculum in the master of educational medicine (only offered at 5 universities in Germany since 5 years).

Importance of fun and flow factor, feedback and debriefing.

All were assessed to be a critical component of a effective SG, especially the fun and immersion factor fostering indirect, subliminal learning. Process feedback was thought to be more pivotal than outcome feedback. Immediate feedback and a mix of audio, visual and haptic feedback were perceived as mostly benificial. Debriefing (with instructor)was found essential for self-reflection. All issues correspond with the theoretical insights.

Most relevant game attributes for training of surgeons

Representation and challenge, besides feedback were thought to be the most relevant attributes for medical education. This is a new insight, although the foremost importance of feedback for learning is stressed in literature.

New generation of learners

There is a new generation of learners that is technology-savvy and has a higher expectancy of the learning environment. These should be interactive, no frontal lesson, media-enhanced).This is corroborated by literature.

3D-VR tools (Samsung VR,Hololens, OculusRift) potential for SG

This was a new insight, not addressed before by literature. In general these trends are perceived as promising, but not yet ready for broad application. Firstly, the instructional content is missing. Secondly, the software and hardware infrastructure is too demanding. The costs are too high (VR lens costing 700-900$ per set) and the spacial contraints (cable-bounded to PC) and hardware prepositions (high performing graphic card, processing capacity) are also a limiting factor. Furthermore, many users complain about difficulties in 3Dthinking and positioning, feeling noxious (catching nausea).

Recommended learning model for Serious Gaming

The experience-based model of Kolb was thought to be a good, although abstract model of learning that fits the serious gaming approach. This model could be added by the Social Cognitive Theory (Bandura), Bloom’s taxonomy or the cognitive Apprenticeship-Model to be more practically applicable and empiricism-proofed as the Kolb model is mostly beneficial for abstract conceptualisation.

Better Suitability of SG for new generation of learners

SG, belonging to that new learning formats that are learner-centred, competency –focused and interactive, SG were found to be more suitable for the ‘Smartphone’-generation, that have high expectations of innovative, adaptive learning tools. This is in line with common literature findings.

Future of SG based on 3D-VR tools (Oculus Rift, Samsung VR, Hololens,etc.)

This issue is often too brightly assessed in the literature, in an optimistic and ‘technology is a universal redeemer’-approach. Contrarily, the opinion of the experts is that the SS as technology is not yet ready for the 3D-VR tools as appealing, high quality content is missing, the infrastructure is not ready (too many constraints of software & hardware) and acceptance barriers of practitioners. Regarding the user side, they prefer the live setting (dissection/vivisection) due to the haptic and high fidelity qualities of both learning tools.

Conclusion: Are SG an effective tool for medical education

Yes, they can be, especially for knowledge, factual learning games rendered in a 2D-Adventure game way. SG are less suited to learn high-stake,high variablity,high fidelity and complexity-enriched tasks, because the programming is not demanding and the literature content is not yet developed, hence SG are not ready for a 3D-VR-world. This is a new insight that has not yet been treated in the literature.

Definition of SG	✓
Differences between SG and Simulations;	✗
Best method to train surgeons	✓
SG used at the medical faculty	✗
Perceived advantages Perceived disadvantages	✓ ✗
Perceived problems	✗
R&D costs	✗
game attributes: importance of fun/flow, feedback debriefing	✓ ✓ ✓
learning model	✓
Suitability of SG for new generation of learners	✓
3D-VR tools potential for SG	✗
SG as effective tool for medical education	YES, for knowledge acquisition and simple, partial task ✓ NO for full, complex tasks ✗

Table 4: Comparison theoretical and practical insights

Source: own creation

The **results** show that:

- Experts acknowledge that there is no universal definition of SG, because every game has its own in-vivo setting(context, gamers, aim)
- A convergence is taken place between simulation and serious games,but the most distinguishing factor between both is that the lack of fun in case of simulations
- In the Netherland (Radboud, UvA), SG are already a formal, inherent part of the medical curriculum and can be found in the Skills-Lab. On the contrary, Germany, SG are only used in an informal and volontary way. Freiburg is using UroIsland and Emerge, and Cologne is using PatDocTalk
- The most important perceived advantages of SG are: fun and flow-state as motivation driver, safe environment for high-stake actions, immediate, unbiased system feedback, especially well-suited to learn knowledge (factual and procedural) and partial task (but not complex operations).
- The main disadvantages of SG concerns the validation issues, high development costs, lack of stakeholder approach and the institutional, professional and individual barriers/fears to use a new, innovative learning method, high total cost of preparation and training.
- Dissection/Vivisection will remain the best method to train surgeons, although SG will be most used in an effective mix of learning methods
- Regarding the game attributes for training surgeons, representation/accuracy (low quality for beginners, high quality for experts), challenge, fun/flow and feedback were named as highly relevant
- Visual feedback is automatically inherent in many games an relevant for all gamer types, whereas haptic feedback is more important for experienced senior students
- Process feedback is more significant as outcome feedback, because learners need to know their improvement gap and how they can improve
- A debriefing stage is considerably valuable for reflection and should be included whenever it is feasible
- Regarding the suitable learning model for SG, Kolb's model was considered to be a good fit due its descriptive power and paramount role of experience -although it should be adapted by using skill development models (ICAP, Erikson) or social cognitive model (Bandura)
- SG are especially suitable for the New Generation of learners ('Smartphone-Generation')
- The SG-potential of new 3D-VR tools like (OculusRift, Hololens) is viewed critical and pessimistic by the respondents as there are too many practical and acceptance limitations.
- Nearly all experts agreed that SG have the potential to be an effective, assisting learning tool, that is especially well-suited for cognitive learning outcomes (factual and procedural knowledge).

Chapter 5: Conclusions and Recommendations

In this chapter, the key conclusions are presented and an assessment of the research quality is made. Moreover, recommendations for future research are formulated.

Based on those results, the practical and theoretical insights, the following intriguing conclusions (inferences about trend, not as summary) can be made:

1. A general universal, ubiquitous definition of Serious Games is possible, but very problematic as every game has its own setting (learning aims, target group)
2. Due to the convergence of simulations and serious games, a new terminology, distinctive research stream should be crystallized out based on the target group and the learning aim (here: healthcare, learning, surgeons)- at best under the gamification umbrella
   To increase knowledge diffusion about the SG and their potential, more interaction should take place between medical educators and game developers. Now both knowledge domains are separated. International academic conferences like the one organised by the institutional bodies (such as DSSH.nl, ilumens.fr, SEGAN, GALA, Serious Games Society or Association) can help to bridge the gap and exchange, helping to promote and foster SG as a effective learning tool.
3. A balanced viewpoint should be striven for regarding games development, applying a stakeholder and balance-scorecard approach
   A sound user analysis should be made before the game development as often a low quality, low-medium complex, low cost solution is better suited as learning tool, than a high quality, high complex, high cost solution. Game design should be based on the players profile and preferences.
4. Although the learning environment for surgeons, will become more complex and technology-enriched, dissection and vivisection will remain the main training methods for surgical training as both offer feedback that is direct and realistic: haptic (tissue structure and movement reaction) and visual (in case of vivisection), working blood flow in case of vivisection and direct, observable consequence (blood,death).
5. Total feedback dimension should be a solid unit of analysis, meaning that the machine and system feedback together should be assessed in a holistic way.

Methodological considerations and limitations

Quality of the research: Reliability and validity

In the past the quality of research was judged by using implicit criteria. But since the year 2000, more explicit criteria and have been formalized. Boeije (2010) emphasizes that the generated knowledge only represents the partial truth and not the actual (social) reality. Quality of research comes down to two indicators, being reliability and validity. According to Steinar (2009), "reliability *refers to the consistency and trustworthiness of research findings, linked to questions whether a finding sis reproducible at other times, by other researchers."* Validity "*refers to the extent to* *which the observations indeed reflect the phenomena or variables of interest .(...) Qualitative research can lead to valid scientific knowledge"(*Steinar,2009, p.122).

Validity

Succinctly, validity asks if the measure that represents specific concepts really reflects this concept. Steinar (2009) differentiates between communicative and pragmatic validity. There are three forms of communicative validity, namely member validation, audience
validation and peer validation. Pragmatic validity concerns actual behaviour changes of the participants due to the researcher's interventions. The motto is: action involves a reaction. In this study, thus only the first form is of interest, because only member validation is tried, although failed (see beneath).

Reliability

Due to the vagueness of qualitative research, replication (reliability) is difficult. Reliability is a necessary, but not a sufficient condition to achieve validity, because when the measurements are unreliable, the researcher fails to measure what he intends to measure. Furthermore reliability is difficult to attain in qualitative research, because there are no standard measurement instruments available in qualitative research. For example, in interview research the instrument is always a human, being the researcher.

Nevertheless, validity and reliability should be striven for as aims regarding the quality of the research. Transcription as the interpretive construction of reality enhances both reliability and validity. Reliability is promoted because others can access the material and will hopefully yield the same findings. Validating is improved because by re-listening the researcher analyse if the interviewee have understood he concepts in the intended way. (Steinar, 2009). Transcriptions were produced for all interviews.

Safeguarding quality

Boeije (2010) has created a list of measures to ensure the quality of research, namely the use of methodological accountability, reflection on the researcher`s role, triangulation, member validation and multiple researchers.

Firstly, the methodological accountability was enforced, because the researcher described all the activities of the methodological steps, hence others can retrace what he has done. At least, virtual replication is feasible, since the researcher has described all those steps.

Secondly, the researcher's role was the one of a careful listener. Some kind of involvement is necessary, because the researcher is interested in the experience, motives and interests of the interviewees. Although none researcher is value-free and everything is value-laden, it is judged that bias is not a potential risk, since the questions are based on a sound literature review and subsequent interview framework.
Thirdly, triangulation is about the use of more sources of data to examine the research subjects. Theoretical triangulation was applied, since in the literature review was sourced from various renowned databases. Methods triangulation was not pursued.
Fourthly, only 1 interviewee gave feedback on his transcription, being a limitation for validity. But all transcriptions were rechecked if the utterance were faithful (valid). Hence, this issue is satisfactorily solved. Because the respondants (mostly holding phd-degrees) had time contraints, many were not able to give feedback or recheck the transciptions altough asked politely to do that.To ask interviewees to verify the interpretations is a direct test of reliability- but this was not possible.
Lastly, researcher triangulation, the use of multiple researchers was not intended, because this is a master dissertation, hence a research project for one researcher. I think that the risk of bias was not present, because literature review guided the interview questions. The literature review is based on the peer-reviewed articles of several other researchers. In my view, this counts as triangulation of researchers as well.

Generalizability

Furthermore, there is the external quality to be judged, concerning the generalizability of the conclusions. Thus the question is whether the results of a study can be generalised to other contexts. This is one of the most difficult endeavours in qualitative research. If the research is not externally valid, results only apply to the research case. If the research is externally valid, results can be extrapolated to other cases and contexts. The key question is then if the cases hold for other not-studied cases. Because every researcher is engaged in theorising, he is automatically concerned about generalizability. There are several types of inductive generalisation such as statistical generalisation, theoretical generalisation and variation-based generalisation. (Boeije, 2010) In inductive reasoning, generalisation runs from research results to a population or to a theory scope. Theoretical generalisation is the application of the principle of replication. The researcher theorises on the basis of a case and then tests the findings. Afterwards the theory is adapted. Theoretical generalisation was one of the side aims of this dissertation. Finally, new theory formulation is the driver for generalisation to cases that have not been researched but still is in the scope of the theory. Furthermore, Boeije (2010) mentions analogical and communicative generalisation. The first one is not appropriate for this study, because it is only relevant if the research design consists of multiple case studies. Communicative generalisation is done by the readers as targeted audience of the research. They make inferences by themselves while linking findings and contexts that matter to them. The researcher needs to accurately describe the case in order to enable the reader constant comparison of the studied and the self-referent cases.

Contribution to literature

This thesis pulls together and promotes coherence in literature by combining insights of the model of Garris (2002) and Wilson (2009) in connection with the experience-based learning model of Kolb. Moreover, a need for future research is pinpointed to address the the game cycle with other learning theories (like ICAP (learning activities) or Bloom's taxonomy to classify educational goals based on their complexity and specificity).

Managerial implications

Medical educators are the main driving force to induce the development of medical SG As knowledge domain experts and holder of tacit knowledge, they need to work together with designers and game developers, trying increase the instructional effectiveness (corresponding learning content =objectives and game attributes.
A game should be developed based on an network approach, doing a sound stakeholder (user profile) and task and learning content analysis.

Recommendation to use of 'new angle' -method in future research

Use technomethdology to assess use of new SG and interaction interface-user

Rooke (2005) proposes that designer should be more aware of how work-flows are organised by technologies and how this relationship conflict with other 'organisational imperatives'. The main purpose is to explicate the problems faced and resolved by the workers.Designers need to take in a 'holding up a mirror '– approach in the design process, where ethnomethodology-derived concepts have ownership of the design concepts.. By doing this, a new discipline is realized – the technomethodology. As an example of such a technomethodology approach Rooke refer to the 'CSCW'(Computer Supported Cooperative Work)-design practice. EM and the design process create a symbiotic relationship- although system design deals with abstractions and EM with generalities. EM promotes the use of ”sensitising concepts that stimulate particular sensibilities in the designer”(Rooke, 2005). Moreover, Rooke recommends the use of breaching experiments to study the 'schemes of interpretation' which users adopt during the use of technologies. The outcomes of these experiments provide then the guiding concepts for design.

In a future research project, i would make more use of axial coding, because the purpose is to describe and delineate categories, determine their relevance and increase the level of abstraction. Whilst open coding provides a certain kind of flexibility and openness for data interpretation, axial coding offers more profound possibilities to reduce the amount of data to a manageable, scalable level.

Concluding, more longitudinal experimental/ quasi- experimental studies regarding SG should be made to assess the validity, effectiveness and evaluation of the edutainment approach. The ultimate goal should be to link certain game attributes to certain learning outcomes, hence that games can be designed more purposefully.

In this final chapter, I conclude that a universal answer to the main research question (How can SG enhance medical learning?) is not ineffectual at a local level, because there are too many individual constellations (game context, game attributes, learning content learning outcomes, individual inclinations and profile regarding learning and gaming) and the working relationships between certain game attributes and their learning outcomes is only researched at a premature level. Furthermore there is a huge knowledge/research gap regarding the questions how learning takes place at a meso/micro level. Hence, only insights at abstract, global level are futile, i.e. about trends related to SG in medical education. The three most significant trends are:

1. A convergences is taking place between SG and simulations;
2. A fertile exchange between medical educators and game designers should be promoted via international conferences, because the other party has the knowledge domain the other needs, thus both need to collaborate and cooperate to develop effective SG.
3. Dissection will remain the gold standard in surgical education as 3D-VR tools do not yet provide a realistic haptic and cognitive simulation of the complex operation environment.

Due to the properties of a qualitative research design, the validity, reliability and generalizability were not so high as in a quantitative research design-all achieved a moderate level. They were striven for as principles and several precautions were made to safeguard them.

References

Articles

Aziz, M. A., Mckenzie, J. C., Wilson, J. S., Cowie, R. J., Ayeni, S. A., & Dunn, B. K. (2002). The human cadaver in the age of biomedical informatics. *The anatomical record*, *269*(1), 20-32.

Bluestone, J., Johnson, P., Fullerton, J., Carr, C., Alderman, J., & BonTempo, J. (2013). Effective in-service training design and delivery: evidence from an integrative literature review. *Hum Resour Health*, *11*, 51

Boyle, E., Connolly, T. M., & Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Computing*, *2*(2), 69-74.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., ... & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games.*Computers & Education*, *94*, 178-192.

Bridges, M., & Diamond, D. L. (1999). The financial impact of teaching surgical residents in the operating room. *The American Journal of Surgery*, *177*(1), 28-32.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661-686.

De Wit-Zuurendonk, L. D., & Oei, S. G. (2011). Serious gaming in women's health care. *BJOG: An International Journal of Obstetrics & Gynaecology*,*118*(s3), 17-21.

Egenfeldt-Nielsen, S. (2007). Third generation educational use of computer games. *Journal of Educational Multimedia and Hypermedia*, *16*(3), 263-281.

Faria, A. J., Hutchinson, D., Wellington, W. J., & Gold, S. (2009). Developments in Business Gaming A Review of the Past 40 Years. *Simulation & Gaming*, *40*(4), 464-487.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & gaming*, *33*(4), 441-467.

Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *British Journal of Surgery*, *99*(10), 1322-1330.

Graafland, M.(2014). Serious Games in Surgical Education. *E-Pub*

Göhlich, M., Wulf, C., & Zirfas, J. (2007). Pädagogische Zugänge zum Lernen.*GÖHLICH, M., WULF, C. & ZIRFAS, J.(Hrsg.): Pädagogische Theorien des Lernens. Weinheim, Basel (Beltz). S. 7*, *19*.

Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of It'sa Deal!. *Computers & Education*, *58*(1), 435-448.

Göbel, S., Hardy, S., Wendel, V., Mehm, F., & Steinmetz, R. (2010, October). Serious games for health: personalized exergames. In *Proceedings of the international conference on Multimedia* (pp. 1663-1666). ACM.

Hogle, N. J., Widmann, W. D., Ude, A. O., Hardy, M. A., & Fowler, D. L. (2008). Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games?. *Journal of surgical education*, *65*(6), 431-435

Kolb, A. Y., & Kolb, D. A. (2009). The learning way meta-cognitive aspects of experiential learning. *Simulation & Gaming*, *40*(3), 297-327.

Landers, R. N. (2014). Developing a Theory of Gamified Learning Linking Serious Games and Gamification of Learning. *Simulation & Gaming*, *45*(6), 752-768.

Lynch, J., Aughwane, P., & Hammond, T. M. (2010). Video games and surgical ability: a literature review. *Journal of Surgical Education*, *67*(3), 184-189.

Luursema, J. M. (2010). *See me, touch me, heal me: The role of visuo-spatial ability in virtual anatomical learning and surgical simulator training*. University of Twente.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*(5938), 284-288.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance.*Academic medicine*, *65*(9), S63-7.

McCallum, S.(2012). Gamification and Serious Games for Personalized Health. *pHealth, B.Bobel et al (Eds), IOS Press*

Orvis, K. A., Horn, D. B., & Belanich, J. (2008). The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human behavior*, *24*(5), 2415-2433.

Pannese, L., & Carlesi, M. (2007). Games and learning come together to maximise effectiveness: The challenge of bridging the gap. *British Journal of Educational Technology*, *38*(3), 438-454.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation.*Computers & Education*, *52*(1), 1-12.

Pavlas, D., Bedwell, W., Wooten, S. R., Heyne, K., & Salas, E. (2009, October). Investigating the attributes in serious games that contribute to learning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, No. 27, pp. 1999-2003). SAGE Publications.

Rooke, J. (2005). Studies of Work: Achieving Hybrid Disciplines in IT Design and Management Studies. *Human Studies Volume 28*

Sanchez, A., Cannon-Bowers, J. A., & Bowers, C. (2010). Establishing a science of game based learning. *Serious game design and development: Technologies for training and learning*, 290-304.

Sandhu, S., Afifi, T. O., & Amara, F. M. (2012). Theories and practical steps for delivering effective lectures. *Journal of Community Medicine & Health Education*, *2*(158), 2161-0711.

Squire, K. (2002). Cultural framing of computer/video games. *Game studies*,*2*(1), 90.

Shademan,A.et al (2016). Supervised autonomous robotic soft tissue surgery.*Science Translational Medicine 4th May 2016, p.337*

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games: An overview.

Todescan, R. (2011). Increasing New Faculty Teaching Effectiveness Through Gaming Strategies: Tips, Tools, and Resources [Presentation slides] . Retrieved from: www.adea.org/publications/.../gameshofinal.pdf

Thompson, D., Baranowski, T., Buday, R., Baranowski, J., Thompson, V., Jago, R., & Griffith, M. J. (2008). Serious video games for health: how behavioral science guided the design of a game on diabetes and obesity. *Simulation & gaming*.

Yusoff, A., Crowder, R., Gilbert, L., & Wills, G. (2009, July). A conceptual framework for serious games. In *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on* (pp. 21-23). IEEE.

Van Dijk, T. (2014). Present or Play-the effect of serious gaming on demonstrated behaviour. *Publication source unknown*

Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet*, *357*(9260), 945-949.

Wattanasoontorn, V., Boada, I., García, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, *4*(4), 231-247.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*,*26*(2), 3.

Wenzler, I., & Chartier, D. (1999). Why do we bother with games and simulations: An organizational learning perspective. *Simulation & Gaming*,*30*(3), 375-384.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., ... & Conkey, C. (2009). Relationships between game attributes and learning outcomes review and research proposals. *Simulation & Gaming*, *40*(2), 217-266.

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45-55.

Books

Boeije, H. R. (2009). Analysis in qualitative research. Sage.

De Vet, M. (1999). *Leren communiceren*. Groningen: Wolters-Noordhoff.

Jaffee, D.(2001).*Organisation Theory-Tension and Change.*McGraw-Hill

Kvale. S.(2007).Doing interviews.Sage

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Ma, M. et al (2014). *Virtual, Augmented Reality and Serious Games for Healthcare 1. Springer Verlag*

Punch, K. F. (2006). Developing effective research proposals. Sage.

Websites:

BIT conference: Serious Games für die Informations,-und Wissensvermittlung.Retrieved from http://www.b-i-t-online.de/daten/BIT_Innovativ_28_Auszug.pdf

Die Zeit (2014,January 21).*19.000 Klinikpatienten sterben jährlich durch Behandlungsfehler.* Retrieved from:http://www.zeit.de/wissen/gesundheit/2014-01/krankenhausreport-aok-behandlungsfeher

DEA (2014). *Was kostet das Medizinstudium den deutschen Staat?.* Retrieved from:http://www.deutsche-eliteakademie.de/load.php?name=News&file=article&sid=234

New York Times (2015, May 22). *What to Learn in College to Stay One Step Ahead of Computers.* Retrieved from: http://www.nytimes.com/2015/05/24/upshot/what-to-learn-in-college-to-stay-one-step-ahead-of-computers.html?abt=0002&abg=0

NOS (2016, May 5). *Onder het mes bij doktor robot*.Retrieved from: http://nos.nl/artikel/2103207-onder-het-mes-bij-dokter-robot.html

Appendix A: Surgical training methods

Dissection

Dissection is the training of the actual body of an deceased person. Dissection is the gold standard for medical training that is promoted by Aziz (2002). According to him, dissection triggers the ultimate empathy and compassion in the trainee, whereby the student applies a 'patient comes first'-strategy throughout his career. Moreover, dissection offers the understanding of the bodily structure in a multidimensional way and with a thorough touch-mediated perception. Dissection should create a knowledge base whereby the basic terminology of medicine is learned in a real-live setting. Dissection is especially renders learning effective when it is linked with computer-assisted learning, competence-learning of diagnostic imaging (CRT, MRT, etc), peer-group learning and training for other medical specialties (such as immunology, biology, pathology, chemistry, etc.).

The great advantage of dissection is the training on the actual body, which is the best dummy available and the lower need of intensive supervision (as compared with the live operating scenario). The disadvantage consists of the lack of any sign of life, hence the tissue structure is changed. There is an age bias (more older people die) and the morgue is expensive to maintain.

Standardized / simulative patients is the use of a living agent, who is trained to act and behave as real patients would do.

Vivisection is the training on alive animals, mostly pigs or rats. The advantage is the opportunity of having a live bodily function, whereby less intensive supervision is necessary. The disadvantages are the cost, the fallacy of overtraining and ethical concerns.

Procedural simulations consist of box trainers and simulation dummies. Box trainers are real instruments attached to a box for endoscopic surgery. A video system offers visual feedback (similar to the one used in laparoscopic surgery). Box trainers need less maintenance and no special facility. The lower degree of fidelity and the lack of anatomical variation are the main drawbacks of this tool.

Source: https://www.youtube.com/watch?v=GUheFddh86A

Simulation dummies with high and low fidelity are also an effective tool to learn surgical procedures. Some dummies even offer smart functions with haptic feedback. **Virtual reality trainers and computer-assisted simulations** simulate the haptic feedback**,** consisting of force and tactile feedback via the sense of touch. Compared with the open surgery, during the simulation the importance of haptic feedback is largely reduced, so it is very easy to simulate by using VR-trainers. The huge advantage of this tool is the little maintenance need. Moreover, there is a advanced portfolio of anatomical variability that can be offered and the possibility of an objective assessment of the surgical performance of the trainees. **Hybrid simulation systems** are based on virtual and haptic controls such as the 'Da Vinci'- skills simulator. This training system for gynaecologic and urological minimal invasive surgery consists of a EndoWrist® manipulation instrument, a camera and clutching device, a 'Fourth Arm' integration, a console (system settings), a needle control and driving tool and energy and dissection outcomes.

Source : http://www.intuitivesurgical.com/products/skills_simulator/

STAR: Smart Tissue Autonomous Robot

The highest form of technology-evolved, full operation equipment, are autonomous robotic systems for surgery. An example of such a system is STAR: Smart Tissue Autonomous Robot, which can treat gall bladders and blind gut completely alone. He can operate just as well as surgeon, although being never tired or distracted, reducing the human failure rate. Improved rates for efficacy, safety, consistency and access to optimized surgical techniques"are given as general benefits.

Innovative is that STAR can deal with soft tissue, the surgeon only supervises and intervenes in case of failure. On this moment, the surgeon needs to make small corrections at 50% of the time Tested in vivisection, it was shown that STAR works as fast as a surgeon, but is more accurate (more regular setting of stitches): *"We demonstrate that the outcome of spervised autonomous procedures is superior to surgery performed by expert surgeons and RAS techniques."(Shademan, 2016, p.337)*

Because the human tissue is in constant movement, the robot needs to anticipate them by following marker points on the tissue via a 3D-camera. The surgical robot should be fully operable in 2018, doing millions of routine operations per year. (This is a too optimistic assessment in my point of view, because there are many accreditation and institutional acceptance barriers.) (NOS.nl, 2016)

Appendix B: clinical competences, assessment and validity of SG

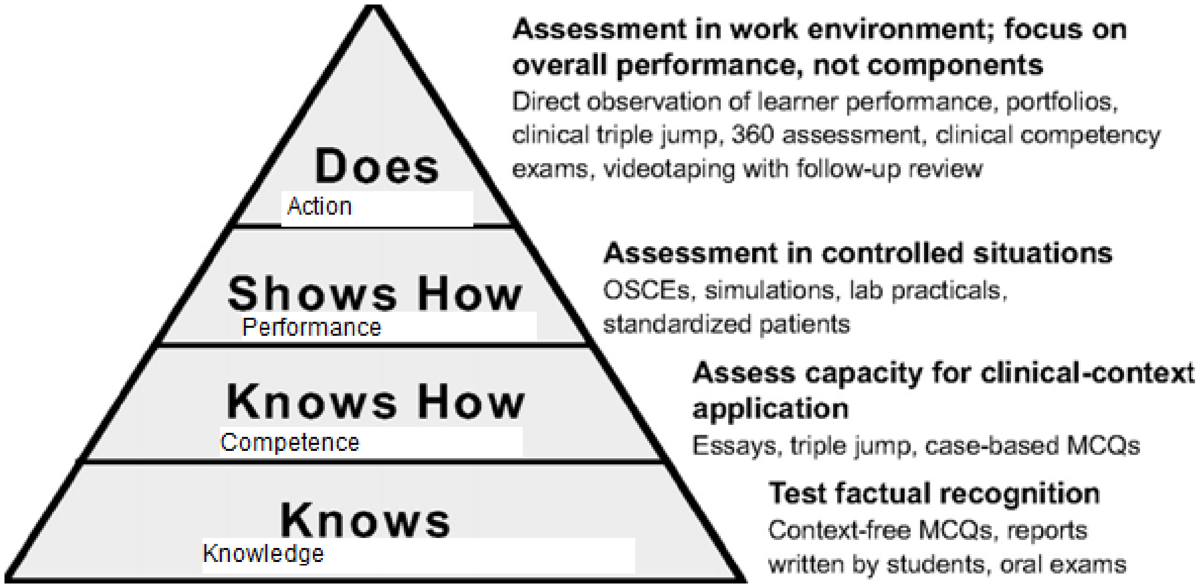
Competency and activities of medical curriculum and assessment of clinical competences by Miller's Pyramid / Validity dimensions of SG

Competency and activities of medical curriculum

To ascertain the potential need of serious games for special competencies based on activities profiles, Graafland did a Delphi study with 149 medical specialists from seven areas and identified 62 EPA.EPA are 'entrustable professional activities' in a clinical setting that are solely allowed to be performed by a certified and entrusted sufficiently proficient, competent surgeon. The five specialty areas consisted of anesthesiology, general surgery, psychiatry, gastenterology and emergency medicine.
Especially 11 EPA were assesses as mostly valuable for patient care concerning 5 specialty areas These are: management of trauma patient;chest tube placing;lacroscopic cholecystectomy;assessment of vital signs;airway management;Induction of general anaesthesia;assessment of suicidal patient;psychiatric assessment;gastroscopy;colonoscopy;resuscitation of emergency patient.
Graafland propose that game developers and educators work together on those EPA's to further technology enhanced trainings in which serious games have the potential of being a effective training tool.

Assessment of clinical competence

Luursema highlights the transition from the traditional master-apprentice training model to the evidence-based training model –which is the gold standard in medical education (Luursema, 2010,p. 101).
Miller's pyramid can be used to characterize the four levels of clinical competence. The assessment of those components (knowledge, competence, performance and action) allows the medical examiner to judge the medical qualification of the trainee and whether the trainee is fit to practise.
Assessment of 'knows' and 'knows how'



Source: http://www.jdentaled.org/content/72/12/1405/F4.large.jpg

Nevertheless, Miller warns that there exists no single assessment method that can gauge and judge the complexity of a professional surgical service.

1) Assessment of 'knows'

At the base of the pyramid is the agent (student, resident, physician) who knows what he needs to do in order to effectively execute the professional functions. This component is about the factual recall of accumulated knowledge. The knowledge base is assessed by objective test methods. Although those knowledge recalling/memorizing tests are always incomplete tools, because the focal interest of medicine is practising rather than knowing.

2) Assessment of the 'knows how'

Graduates must then apply the knowledge they have learned and develop skills regarding the clinical problem-solving and decision making. Information management skills are especially important in the clinical setting and encompassing the collection of information from different sources (human or laboratory), analysis and interpretation of data and their translation into a diagnostic treatment plan. A distinction can be made concerning intellectual skills (knowledge application) and technical skills (execution of diagnostic and therapeutic procedures).

This is about competence, because competence is the "quality of being functionally adequate or of having sufficient knowledge, judgment, skill or strength for a particular duty" (Miller, 1990), Both components (know and knows how) are conveyed in the early, undergraduate stages of the medical curriculum.

Because this second level/component is skill development, it can linked with the Dreyfus model of skills acquisition. This pragmatic model outlines the stages of skills acquisition on a continuum from a novice until an expert level. Novices require conscious processing to execute a procedure, leading to slow execution, unnecessary actions and susceptibility to error, they follow strictly the rule-based decisions paths VS experts perform procedures seemingly effortless, intuitive and fast, are less susceptible to errors and mental strain, they make superior decisions and respond to emerging task demands with little preparation Achieving expert level requires prolonged deliberate practice, where trainees need a well-defined goal, motivation to improve, feedback and ample opportunities to repeat and refine performance.

3) Assessment of 'show how'

Students must also show how they do something. The assessment of their performance is a real challenge, because judgments often focus on limited direct observation and restricted sampling of typical clinical problems. This component is found at the graduate level, because skills-teaching becomes vertically integrated and complex.

4) Assessment of 'does'

On this level, the postgraduate student shows that he can transfer the knowledge learned in an artificial setting to a real clinical practice setting. This action component is the most difficult to gauge and measure in an objective, accurate and reliable way.

Validity issues of Serious Games

In his systematic review of serious games for medical education, Graafland reasons, that none of the 30 games (17 games specifically developed for medical education) completed a full validation process. But this is an absolute imperative before serious games can be used in the medical curriculum.

He outlines six validity types:

- Content validity: Does game content adequately covers the dimensions of the medical education construct ?

- Face validity: Degree of similarity between medical construct created in gameplay and in reality (as assessed by novice and experts)
- Construct validity: Difference in outcome of novice and experts players based on gameplay outcome parameters
- Concurrent validity: Concordance of study results using a concept instrument (game)and study results on an established instrument believed to measure the same medical theoretical construct
- Predictive validity: degree of concordance of a concept instrument (game) and task performance in reality, based on a validated scoring system

Appendix C: Quality research background

Qualitative research design, philosophy of knowledge, philosophy of knowledge, the unit of data collection/analysis, sampling, data analysis and nature of qualitative data, mode of analysis and integrative procedures

Qualitative research design

A qualitative research design will be used. Firstly, i want to understand the phenomena of serious games in healthcare and what meaning people attribute to it. Secondly, the flexible research methods allow me to enable contact with informants to get rich data. Lastly, the findings are re-interpretive descriptions based on the informant's utterances. Hence, data collection sampling and analysis are linked with each other. (Boeije, 2010). The research purpose has a fundamental research aim, because one focus is the knowledge gain about a new phenomenon (SG)- thus an applied research is done. This study is of exploratory and descriptive nature, because the SG-phenomenon is" *a newly emerging field of interest that has not yet been extensively examined"*(Boeije, 2010, p.32). It is descriptive, because literature and experience/ views of the informants are used to understand what is going on in the field of study.

Philosophy of knowledge

A post-modern approach is intended, in which the interview is seen as site of knowledge construction, it is the "*interchange between two persons conversing about a common them"* (Kvale, 2009, p.70). People construct their own reality and the researcher uses methods to uncover their perspective, experience and behaviour. (Boeije, 2010). The key feature of empirical research is that direct, observable data is used to answer the research question and in which experience is the source of knowledge. (Punch, 2006)

Unit of data/ analysis

The unit of data collection and inquiry are the individual interview partners (educators, surgeons or designers. whereas the unit of analysis are the utterance of the interviewees are distilled into empirical facts (findings).

Data collection: Sampling, recruitment and access

A purposive, non-probability sampling of informants is used, because this suits the needs of the study and because this research is informed a-priori by an existing body of theory from the outset. (Boeije, 2010, p.24). About 6 interview partners were recruited (will be 10) because the access to their expert knowledge and contact with them is feasible. The experts will be surgeons residencies, educators and game designers.

Qualitative data is resource for "well-grounded and rich descriptions and explanations (of processes)"(Miles and Huberman, 1994,p.1) Researchers use sound qualitative data to go further than the preliminary conceptions, revising the conceptual frameworks as new insights emerge (Miles and Huberman, 1994,p.1). A serious problem with qualitative data is the lack of well-explicated canons or guidelines for the analysis.

Data analysis

For Boeije (2010), "*qualitative analysis is the segmenting of data into relevant categories and the naming of these categories. ...*Analysis is the breaking up, separating and disassembling of research material into pieces...with facts broken down to manageable pieces...aim of this process is to assemble or reconstruct the data in a meaningful or comprehensible fashion (based on Jorgensen, Boeije, 2010, p.76). The researcher segments the data to figure out relevant and meaningful data parts, which are compared to find similarities and differences. By disassembling the data, the topics emerge out of the raw data.

Reassembling not an issue here, because we are not interested in uncovering the relationships between the building blocks (as it would be the case in a grounded field approach), hence we are not using axial and selective coding.

Unfortunately, very few standards exist in qualitative research how to do a sound data analysis and integration of data. One option consists of using the bricolage-approach of Kvale (2009) as free interplay, in which the analysis is done without any specific methods. Thereby a researcher picks the method that fits the needs of the study.

Regarding the integrative procedures, Boeije (2010) recommends the use of heuristics to filter out the most meaningful parts and to integrate the consistently. The best guide to decide what data is tantamount to proof and inference is logical reasoning.

The data analysis is interwoven, iterative cycle which consists of three activities, to be called data reduction, data display and conclusion drawing/verification .(Miles and Huberman, 1994, p.9)

data reduction (coding) = data condensation	process of selecting, focusing, simplifying, abstracting, transforming the data that appear in transcriptions
data display (matrix, cognitive map)	display is an organized, compressed assembly of information that permits conclusions drawing and action logic: “you know what you display”
Conclusions drawing and verification	analyst decide what things mean-noting regularities, patterns, explanations, possible configurations, causal flows and proposition conclusions are also verified: meanings emerging from the data have to be tested for their plausibility, their sturdiness, their confirmability= their validity

Nature of qualitative data

All the data is are converted into words and it is qualitative because reference is made about people (experiences), artefacts and situations.The words are sourced from qualitative methods like ‘watching, asking or examining’. The basic raw data needs to be processed before any analysis. In our case, the recording is transcribed into a simplified text that is clear to the reader.

The strength of qualitative data is their focus on “naturally occuring, ordinary events in natural settings-so we have a strong guess of what real life is” and the emphasis on a specific phenomenon (context embedded and local groundedness). Futhermore, the data is rich and holistic enough to enable a in-vivo-realistic description of complex phenomena. Lastly, the the emphasis is put on lived experiences, attuned to the meaning agents gave to events. (Miles and Huberman, 1994)

Modes of analysis

There are no standard methods to crystallize out the key meanings and implications of the verbal content of an interview. In order to achieve a sound quality of the analysis the researcher should have sufficient knowledge about the covered research topics, should be sensitive towards the interviewees and should master the language and analytical tools of qualitative research (Boeije, 2010).

Normally, an interview analysis would consist of meaning condensation, meaning categorisation and meaning interpretation (Boeije, 2010).

The researcher decided to follow the 'bricolage'-approach to interview analysis (Steinar, 2009), in which the analysis is done without following any specific methods and the researchers chooses by himself a method that fits the needs of this study at most: The ‘bricolage’ approach is a ‘free interplay’ of methods in the analysis stage that should pinpoint the relevant links and structures in a study (Steinar, 2009).Last of all, Steinar (2009) proposes to use the interview analysis effort as a simultaneous endeavour for theoretical reading and reflection.

Integrative procedures

The integration of data into a consistent format is the most challenging task of the research process, because even the literature about qualitative analysis is not addressing this aspect due to the vague, unique and specific nature of the qualitative research project. According to Boeije (2010), there is no general panacea how to integrate the data. She proposes that the researcher can make use of heuristics (thinking aids) to filter out the most meaningful parts from the data and to subsequently integrate those parts. Boeije describes ten heuristic devices, out of which visual displays, reading memos and the construction of arguments seem to be the most appropriate ones for this study. According to her, logical reasoning is the best guide to decide what data is tantamount to proof and inference, helping the reader to connect the data with the equivalent inferences. Moreover, there are two levels of reasoning involved: The first level concerns the logic of the informant's account. The researcher's claim must be grounded on plausible arguments. The second level of reasoning is about the study as a consistent whole in which the argument of a section builds up the line of argumentation of the whole report. That is why all parts of the research process are intertwined. (Boeije, 2010)

Appendix D: Table of key articles: key findings and concepts

Author (Year) and Title	Reason for importance/key findings	Concepts, Classifications or Taxonomy
Wattanasoontorn,V. et al.(2013) "Serious games for health."	Core process of SG and their functionalities Detailed overview of health-related SG	Functions of SG, core components (also stakeholders), classifications based on health purpose and players
Tolks, Daniel, (2013)"Serious Games for Health: Spielend lernen und heilen mit Computerspielen?	6 games for medical training: Pulse, Heart Sense,Uro Island, Immune Attack, Cellcraft, PatDocTalk	Criteria for SG for a valuable inclusion in the medical curriculum: -digital format -game elements (no simulations that have no game content) -medical curriculum themes and practical inclusion possible -skills need to be relevant for medicals -at least one learning goal
Aziz, M. Ashraf (2013), et al. "The human cadaver in the age of biomedical informatics"	9 reasons why dissection is the goldstandard and irreplacable and 9 reasons why there is a continous drive to reduce dissection need -	Primacy of the patient Multidimensional body structure importance of touch-mediated perception for student anatomical variablity anatomical terminology peer-group learning CA-Learning and diagnostic imaging ability
Diner, Barry M. (2007), et al. "Graduate medical education and knowledge translation: role models, information pipelines, and practice change thresholds."	Knowledge transfer, diffusion and barriers in medical education caution: educators assume that if they teach residents they learn and apply what the y have learned--> bold idea	Pathmen's pipeline Knowledge translation via knowledge to action-framwork: acceptance, application, ability and evidence-based actions importance of evidenced based medicine
Sandhu (2012), Sumit, Tracie O. Afifi, and Francis M. Amara. "Theories and	Postmodern needs to deliver effective lectures	constructive knowledge transfer, interactive lectures and principles for designing effective lectures (case reports, technology-assisted, problem-based, open discussion)

practical steps for delivering effective lectures."		
Bluestone, J., et al (2013). "Effective in-service training design and delivery: evidence from an integrative literature review."	Literature review to get evidence for effective training design CA- learning can be more effective than live instruction repetitive interventions superior for learning outcome realistic setting (similar to workplace) improves skill acquisition and performance in-service training must be evidence-based	Case-based learning,clinical simulators, practice and feedback as effective techniques
Graafland, Maurits, (2012)"Systematic review of serious games for medical education and surgical skills training	Literature review of SG for surgical training 30 SG --> only 6 were validated	Content validity face validity construct validity concurrent validity predictive validity
Wilson, Katherine A., et al (2009). "Relationships between game attributes and learning outcomes review and research proposals." .	Game attributes and link with learning outputs--> diffiult to investigate relationship (composite vs.unique RS)	Instructionall effectiveness=overlap between instructional content=goals and learning outcome Game attributes: representation,sensor stimuli,challenge,assessment, control Learning outcomes:cognitive,skill-based and affective
Garris, R.et al (2002)l. "Games, motivation, and learning: A research and practice model."	Focus instructional games input-process-outcome model of learning cycle for games	Key feature of games game cycle with repeated user judgements-behaviour- feedback-loop feedback=knowledge of results--> critical for performance and motivation Discrepancy performance and desired standards Debriefing

		types of learning outcome (skill-based,cognitive(declarative,procedural,strategic), affective)
de Wit-Zuurendonk, L.D.(2011). "Serious gaming in women's health care."	Benefits of SG (theoretical: competitive element,entertainment feedback practical: games are enjoyable,transferable, cheap to distribute digitally, restarted, score and skills recording,cheap testing facility) Literature review (30 articles) conclusions: effectiveness of SG has not been conclusively demonstrated yet	Learning theory based on Knowles (1970)-4 elements of adult learning: they are automatic, want independence/ use their past experience/goal oriented/ problem-based learners-not content-learned cognitive theory of multimedia learning (Mayer) theory of connectivism (Siemens) Evaluation of effectiveness of new training method based on 4 levels of Kirkpatrick: improved learning (1),clinical behaviour (2), clinical behaviour change (3) and outcome results (4)
dit Dariel, O. J. P. et al. (2013) "Developing the Serious Games potential in nursing education."	There can be no single model of clinical reasoning new tools are needed to learn skills in high risk-high tech setting	Clinical reasoning novice and expert level automaticity information-processing theory clinical reasoning model SG: contructivitst learning theory Bloom's taxonomy(1956): 6 levels of competency (Knowledge,Comprehension,Application,Analysis, Synthesis,Evalutation)
Kolb, Alice Y., and David A. Kolb. "The learning way meta-cognitive aspects of experiential learning." *Simulation & Gaming* 40.3 (2009): 297-327.	Experience has central role in learning process based on Kurt Lewin	Kolb's learning cycle(concrete exp.-->reflective obs.-->abstract conceptualisation-->active experimentation)
Wass, Val, et al. (2001)"Assessment of clinical competence."	criterias for test of clinical competence:Blueprinting,validity,reliability and standard setting	Miller's pyramid of competence:knows,knows how,shows how,does

Connolly, T. M., et al. (2012) "A systematic literature review of empirical evidence on computer games and serious games."	Systematic literature review (129 papers) games good for knowledge acquistion and content understanding and affective and motivational outcomes	Cognitive skills motor skills behaviour change soft skills affective/motivation outcomes unintended consequences
Lynch, J. et al(2010) "Video games and surgical ability: a literature review."	Negative consequences of video game playing:decreased organisational skill and academic performance cautionary comment: manual surgical technique, just one competence - interpersonal communication skill and good judgement also pivotal to be good surgeon	not relevant
Girard, C.,et al (2013). "Serious games as new educational tools: how effective are they? A meta-analysis of recent studies."	Lack of empircal studies to show effectiveness of SG in learning need to compare SG with other training methods-lonitudal studies SG might be powerful tools for learning	not relevant
Chang, Huan Ying, et al. (2015)"Student Preferences on Gaming Aspects for a Serious Game in Pharmacy Practice Education: A Cross-Sectional Study."	Students prefered fantasy/medieval/mythic games most prefered game style: 3D and collaborative, unlocking reward system lower grade students prefered adventurer 3D game, upper grade prefered authentic plot, 2D	Descriptions of gameplay aspects regarding reward system, game setting,game storyline, game perspective, game style and game scenario

Appendix E: Interview Framework

Main issues	Questions
Introduction to interview	• Describe goal and purpose of thesis
Background interview partner	• Personal details of interviewee • Research experience • Experience with Serious Games
Serious Games	• Describe and define Serious Game as a concept • Which applications are used at your university to train (new) surgeons? • Who are the key players in the Serious Games market? • What are the perceived advantages and disadvantages of SG(cost-benefit analysis)?
Medical Learning	• Which method is used at your institution to train new surgeons (dissection, standardized patients, box trainers, simulation dummies, virtual reality trainers and computed assisted simulations, hybrid simulation systems, serious games) • Serious Games: used or not? • How is the clinical competence of the trainees assessed (analogue and digital)?--> instructional content? IF SERIOUS GAMES ARE USED--> Input: • What is the instructional content (surgical knowledge and skills)? • Please descibe your experience and the relative importance regarding the game attributes: representation, stimuli, challenge, control and feedback) • Feedback focus (positive, negative / audio, visual, haptic/ outcome vs.system/ process feedback) • Which game attributes are the most effective/important for medical learning? • Which game attributes are especially important for knowledge transfer and for skill acquisition? • Is a universal game engine or a specifically developed game engine more suitable? • What is the advantage/ disadvantage of using SG in comparision with simulation s (dummies and CAD)? • Could SG training is able to replace dissection method? • What are the perceived problems of using SG in medical learning? • Which game posture is more appropriate (fantasy vs. Real, 2D vs 3D) PROCESS: show interviewee Kolbs learning cyle picture--> Does this cycle is a good desciption of the learning process regarding serious games?

	If not: Which learning model would you regard as more suitable? How important is the debriefing stage? OUTCOME: • Please describe the opportunity of SG to enhance learning in general and to learn cognitive knowledge and (psycho)motor skills in particular • How do you assess the overall instructional effectivess of a serious game (overlap between learning goal and game attributes) • How is the validity of SG assessed ? • Which statement is more appropriate? 1) single game attributes leads to learning 2) OR combination of attributes contribute to learning • Are learning outcomes be used as evaluation method?
Conclusion	• Are Serious Game an effective tool for medical education?
Context	• development cost of SG • saving potential of SG as training method to drive down high cost of medical training • potential of SG in medical education in the future • applications for (Hololens, Oculus Rift, etc.) • SG more suitable for New Generation of learners (learning context: interactive, democratic, teacher as facilitiator, postmodern)? • How could the design process of SG regarding medical education could be improved?
Other interesting points that came up during session	
Transcribe interview & send for review to interviewee	