Serious games in surgical education
Graafland, M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Systematic review of validity of serious games for medical education and surgical skills training

Maurits Graafland, Jan Maarten C. Schraagen, Marlies P. Schijven

The British Journal of Surgery
2012;99(10):1322–30
ABSTRACT

Background: The application of digital games as training modality for medical professionals is on the rise. These so-called serious games form training tools that provide a challenging simulated environment, ideal for future surgical training. Ultimately, serious games are directed at reducing medical error and subsequent healthcare costs. This systematic review aims to give an overview of current serious games for training medical professionals and scrutinize validity testing.

Methods: PubMed, Embase, The Cochrane Database of Systematic Reviews, PsychInfo and CINAHL were searched for available studies up until April 2012. Studies were selected on predefined inclusion criteria. The primary endpoint was achievement of steps in a formal validation process, according to current standards of validation.

Results: A total of 25 articles were found relevant, describing a total of 30 serious games. The games were divided into two categories, either developed for educational purposes (n=17) or games associated with but not specifically developed for improvement of skills relevant to medical personnel (n=13). Six serious games were identified that underwent a process of validation. Of these six games, three games were developed for team training in critical care and triage; and three games were commercially available games applied to train laparoscopic psychomotor skills. None of the serious games completed a full validation process for the purpose of use.

Conclusions: Blended and interactive learning by means of serious games may be applied to train both technical and non-technical skills relevant to the surgical field. Games developed or used to train medical professionals need validation before integration into surgical teaching curricula.
INTRODUCTION

Patient safety concerns call upon the need to train medical personnel in simulated settings to reduce cost and patient morbidity. Technological innovations, such as virtual reality simulation and e-learning applications show consistent improvement on learning outcomes, and already play a role in surgical residency training programs\textsuperscript{1–3}. A potent concept for medical education is interactive learning through ‘serious games’. A serious game is formally defined as “interactive computer application, with or without significant hardware component, 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\textsuperscript{4}.” Serious games are designed with primary objectives other than entertainment and therefore clearly differ from conventional videogames. They can be played on platforms such as personal computers, smartphones or videogame consoles and can apply multimodal interactive content in any virtual environment. They present an ideal playground to engage players in simulated complex decision making processes like those required in medical training\textsuperscript{5}.

Serious games form a balanced combination between challenge and learning. Playing the game must excite the user, while assuring that the primary goal (i.e. acquiring knowledge or skills) is reached seemingly effortlessly, thus creating a ‘stealth mode’ of learning\textsuperscript{5,6}. Players are challenged to keep on playing to reach the game-objective. This corresponds well to Ericsson’s theory of deliberate practice: as players are not naturally ‘good’ at a game, intentional repetitive training makes a player become an expert\textsuperscript{7}. Games hold clear advantages over conventional learning methods due to their competitive elements, entertainment aspects and feedback mechanisms\textsuperscript{8,9}.

To date, many medical professionals may still have a rather outdated view on the average ‘gamer’, as being someone who is too young to vote, afraid of daylight and killing mystical dwarves in games like \textit{World of Warcraft} (Blizzard Entertainment, Versailles, France) in their parents’ basement. Contrary to this view, adults are avid users of digital devices, and playing videogames is in fact an important part of their lifestyle. The average videogame-player is 37 years old and has been playing games for over twelve years\textsuperscript{10}. Forty-two percent of all game-players are women and women over 18 years old represent a significantly greater share of the game-playing population (37 percent) than boys ages seventeen or younger (13 percent)\textsuperscript{10,11}.

Although game-based learning is becoming a new form of education throughout healthcare, scientific research on its effectiveness is rather limited. Ideally, training instruments measure certain parameters, ‘game-metrics’, to assess the trainees’ performance. If training and testing of health care professionals such as surgical trainees is to be carried out in digital game-based environments, strict requirements should be met. Use of these games and interpretation of underlying game-metrics must be reliable,
Chapter 1

valid and cause-specific. Thorough, scientific research on validity testing is mandatory, before serious games may be applied to surgical training curricula in a valid manner. The aim of this review is to identify the value of serious games for training professionals in the medical and, in particular, the surgical field. The first objective is to assess the background of serious games for the purpose of training professionals in medicine and their usability in surgical postgraduate training. The second objective is to assess the validity of serious games as a teaching method according to criteria regarded as best evidence.

METHODS

Search criteria
A systematic search was performed on peer-reviewed literature on serious games used to educate professionals in medicine. Serious games were defined as digital games for computers, game consoles (such as PlayStation®, Sony, Tokyo, Japan; Nintendo®, Nintendo, Kyoto, Japan), smartphones or other electronic devices; directed at or associated with improvement of competence of professionals in medicine. Professionals in medicine were defined as individuals responsible for patient care (doctors, nurses, physiotherapists, paramedics etc.) in institutionalized settings. Serious gaming, e-learning and virtual reality simulation tend to overlap and strict subdivision frequently proves to be difficult. This search focused on game-based learning programs, excluding papers on virtual reality (VR) simulation and E-learning to the best of our ability.

Study selection and assessment of serious games
PubMed, Embase, The Cochrane Database of Systematic Reviews, PsychInfo and the Cumulative Index to Nursing and Allied Health Literature (CINAHL; all from January 1995 to April 2012) were searched for key terms (serious gam* OR videogame* OR video gam* OR gaming) AND (medical education OR educat* OR training). The last search date of the databases was on April 12th, 2012. Additionally, reference lists of relevant articles were searched. In the CINAHL database, the limitation of ‘peer reviewed articles’ was added. Only completed trials were regarded as relevant. No reports were excluded based on language.

The titles and abstracts of all reports were screened for the previously mentioned search criteria. All articles deemed ‘relevant’, ‘dubious’ or ‘unknown’ were examined in full text. Data on serious games was extracted from all papers, including name, type, platform, purpose, target population and the presence of validation studies. If necessary, additional information was sought on the publisher’s website or through correspondence with the authors.
Review of validation studies

Studies designed to validate serious games were assessed for achievement of steps in the validation process, according to criteria regarded as best evidence (Table 1)\textsuperscript{12,13}. The predominant question to be addressed by validity testing is whether the instrument measures what it is intended to measure\textsuperscript{12}. Content must be reviewed by experts on the subject. The instrument’s face validity must be valued by both novice trainees and experts to gain acceptance. The parameters by which performance is objectified should be scrutinized for their representation of the skills they are intended to measure (construct validity). After these measurements are performed, the instrument should be compared to current methods of teaching in order to assess alignment in outcome or unexpected deviations (concurrent validity). The transfer of skills acquired on the instrument to performance in reality (predictive validity) is the final step in the validation process.

Data on validation studies was extracted in accordance to the Cochrane Handbook for Systematic Reviews of Interventions and concerned methodological aspects (study design, intention to treat, randomisation, concealment of allocation, blinding, follow-up, other possible bias), details on the serious game (mentioned in previous section), study population, details on intervention, primary and secondary endpoints, instruments, timing, results of measurements performed and funding\textsuperscript{14}. Quality of randomized controlled trials was systematically assessed using The Cochrane Collaboration’s Tool for

<table>
<thead>
<tr>
<th>Validity Type</th>
<th>Description</th>
<th>Criteria for achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content Validity</strong></td>
<td>The degree to which a game content adequately covers the dimensions of the medical construct it aims to educate (or is associated with).</td>
<td>Uniform and positive evaluation of game content and associated testing parameters by expert medical specialist panel.</td>
</tr>
<tr>
<td><strong>Face validity</strong></td>
<td>Degree of resemblance between medical constructs featured in gameplay and in reality, as assessed by novices (trainees) and experts (referents).</td>
<td>Uniform and positive evaluation of the game as a valuable learning environment amongst novice and expert medical specialists.</td>
</tr>
<tr>
<td><strong>Construct validity</strong></td>
<td>Inherent difference in outcome of experts and novices on gameplay outcome parameters</td>
<td>Outcome differences considered to be of significance between players being of different medical specialist level-of-skill.</td>
</tr>
<tr>
<td><strong>Concurrent validity</strong></td>
<td>Concordance of study results using a concept instrument (e.g. game) and study results on an established instrument or method, believed to measure the same medical theoretical construct.</td>
<td>Outcome parameters show correlation considered to be significant between game and an alternative, established training method.</td>
</tr>
<tr>
<td><strong>Predictive validity</strong></td>
<td>The degree of concordance of a concept instrument (e.g. game) outcome and task performance in reality, based on a validated scoring system.</td>
<td>Metrics show correlation considered to be significant between outcome parameters of a game and performance results on the medical construct featured in the game in real life after performers are being trained using the game.</td>
</tr>
</tbody>
</table>

Table 1. Matrix of validity types for games relevant in educating medical professionals.
Assessing Risk of Bias, resulting in either low risk (LR) or high risk (HR) for bias\(^{14}\). Observational studies were assessed with the Methodological index for non-randomized studies (MINORS). This validated instrument grades studies on a weighted 12 item-scale, with a maximum score of 16 for non-comparative studies and 24 for comparative studies\(^{15}\). The achievement of steps in the validation process based on data extracted from the articles was judged by two reviewers that were not involved in the study articles (MG and MS). In case of disagreement between both reviewers, a third reviewer was consulted.

RESULTS

The systematic search rendered 1151 articles. Figure 1 demonstrates the search and screening strategy used for the articles. A total of 25 articles were found relevant, describing a total of 30 serious games. These 30 games were divided into two categories: Category 1, which consists of serious games specifically developed for educational purposes (\(n = 17\)) and Category 2, which consisted of commercially available games associated with (but not specifically developed for) improvement of skills relevant to
medical personnel (n = 13). Due to heterogeneity of study designs and serious games, pooling of data was not performed.

**Serious games designed for an educational purpose.**

Nineteen articles discuss seventeen serious games specifically designed to train professionals in medicine (Category 1). The majority is highly relevant to surgical trainees. Table 2 presents an overview of these games. *Pulse!* (Breakaway Ltd., Hunt Valley, MD, USA), *3DiTeams* (Duke University, Durham, NC, USA), *Clinispace* (CliniSpace, Los Altos Hills, CA, USA) and its predecessor *Virtual ED* (Stanford University, Stanford, CA, USA) were developed as platforms for training critical care, e.g. advanced trauma life support\(^{16-18}\). These platforms provide possibilities for team training. *Burn Center* (360 ED, Inc. Orlando, FL, USA) is used to train treatment of burn injuries in an accredited multidisciplinary course\(^ {19}\). The *Off-Pump Coronary Artery Bypass (OPCAB)* game and the *Total Knee Arthroplasty* game (University of Ontario Institute of Technology, Toronto, Canada), have been developed to train decision steps in a virtual operating room\(^ {20-21}\). Other topics include triage and basic life support. Two augmented reality environments were developed, in which virtual reality is projected over a real environment by either wall-projection (*Cave Automated Virtual Environment (CAVE)*, Electronic Visualization Laboratory, University of Illinois, Chicago, IL, USA\(^ {22}\)) or via head-mounted display (*Project Touch*\(^ {23}\), Center for Telehealth, Univ. of New Mexico Albuquerque, NM, USA).

Ten of these serious games have multiplayer functions, useful for team training and seven of these ten games were specifically designed for this purpose\(^ {16-18,24-25}\). Methods for measuring performance vary: articles on three games, i.e. *Pulse*, *OPCAB* and *Total Knee Arthroplasty*, describe an intrinsic scoring system based on the accurate or inaccurate choices made by the trainee\(^ {16,20-21}\), whereas other games rely on external assessors to judge performance\(^ {17,18,22-24}\).

Eight serious games underwent steps in validity testing\(^ {18,21,22,24,26-30}\), results of which are added in Table 3. *Virtual ED* was evaluated in a randomized controlled trial, in which 30 novices were randomly assigned to manage six patient cases in either *Virtual ED* or human patient simulators\(^ {18}\). Team leadership performance was assessed by three assessors in one pre-test and one post-test case on a standardized scale. Results showed similar skills improvement for both groups, thus proving concurrent validity. Medical content was designed by an independent institution\(^ {24}\).

*Virtual ED II* (Stanford University, Stanford, CA, USA) was tested on face validity by means of a questionnaire on its usability among 22 physicians and nurses with an average of 4 years of experience\(^ {29}\). A majority felt immersed in the virtual world, felt the game improved their confidence and believed the cases were useful in learning clinical skill management, proving face validity.
<table>
<thead>
<tr>
<th>Serious Game (N=17)</th>
<th>Game type</th>
<th>Platform</th>
<th>Purpose</th>
<th>Multi-player</th>
<th>Target groups</th>
<th>Implemented in clinical practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute and critical care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3DiTeams 16</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Team training in acute and critical care.</td>
<td>Yes</td>
<td>Physicians, nurses</td>
<td>No</td>
</tr>
<tr>
<td>Clinispace 17</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Team training in acute and critical care.</td>
<td>Yes</td>
<td>Physicians, nurses</td>
<td>No</td>
</tr>
<tr>
<td>HumanSim 25</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Platform for scenario-based education, e.g. team training in acute care, critical care.</td>
<td>Yes</td>
<td>Physicians; nurses; emergency medical personnel, students</td>
<td>No</td>
</tr>
<tr>
<td>Pulse 16</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Acute care and critical care.</td>
<td>Yes</td>
<td>Physicians.</td>
<td>No</td>
</tr>
<tr>
<td>Virtual ED 18</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Team training: Acute and critical care</td>
<td>Yes</td>
<td>Physicians, nurses</td>
<td>No</td>
</tr>
<tr>
<td>Virtual ED II 24</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Team Training: Acute care, triage in mass casualty events involving hazardous materials.</td>
<td>Yes</td>
<td>ER physicians, ER nurses</td>
<td>No</td>
</tr>
<tr>
<td>Virtual Operating Room</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off-pump Coronary Artery Bypass Surgery game 30</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Training operation steps for OPCAB procedure</td>
<td>Yes</td>
<td>Surgical trainees</td>
<td>No</td>
</tr>
<tr>
<td>Total knee arthroplasty game 21</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Training operation steps for total knee arthroplasty procedure</td>
<td>No</td>
<td>Surgical trainees</td>
<td>No</td>
</tr>
<tr>
<td>Triage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAVETriage training 22</td>
<td>Immersive Learning Environment</td>
<td>Projected</td>
<td>Triage training</td>
<td>No</td>
<td>Physicians</td>
<td>No</td>
</tr>
<tr>
<td>Code Orange 34</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Triage and organisation in mass casualty incidents</td>
<td>Yes</td>
<td>Physicians, nurses</td>
<td>No</td>
</tr>
<tr>
<td>Serious Game</td>
<td>Game type</td>
<td>Platform</td>
<td>Purpose</td>
<td>Multi-player</td>
<td>Target groups</td>
<td>Implemented in clinical practice</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------------------------------------------------------</td>
<td>--------------</td>
<td>-------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Nuclear Event Triage Challenge</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Triage in nuclear events.</td>
<td>No</td>
<td>First responders</td>
<td>No</td>
</tr>
<tr>
<td>Peninsula City</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Team Training: triage in mass casualty events, hazardous materials</td>
<td>Yes</td>
<td>physicians, nurses</td>
<td>No</td>
</tr>
<tr>
<td>Triage Trainer</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Triage in mass casualty incidents</td>
<td>No</td>
<td>First responders</td>
<td>No</td>
</tr>
<tr>
<td>Other purpose</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>(1) Triage and (2) resuscitation of burn patients</td>
<td>No</td>
<td>Physicians, nurses</td>
<td>Multimodal training course</td>
</tr>
<tr>
<td>Burn Center</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Training basic life support</td>
<td>No</td>
<td>Medical personnel (not specified)</td>
<td>No</td>
</tr>
<tr>
<td>OLIVE Cardiopulmonary Resuscitation</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Platform for scenario-based education, e.g. team training in acute care, critical care.</td>
<td>Yes</td>
<td>Physicians, nurses, students</td>
<td>No</td>
</tr>
<tr>
<td>Project Touch</td>
<td>Immersive Learning Environment</td>
<td>Projected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radiation Hazards Assessment Challenge</td>
<td>Game-based simulation</td>
<td>Computer</td>
<td>Assess radiation hazard after nuclear event</td>
<td>No</td>
<td>Physicians, nurses, emergency medical personnel</td>
<td>No</td>
</tr>
</tbody>
</table>

**Table 2.** Overview of serious games specifically developed for educational purpose (Category 1), ranked according to purpose.

*Abbreviations: CAVE = Cave Automated Virtual Environment; OPCAB = Off-Pump Coronary Artery Bypass;*
<table>
<thead>
<tr>
<th>Serious Game</th>
<th>Study type</th>
<th>Validity steps tested</th>
<th>Methodological Quality</th>
<th>Groups (N)</th>
<th>Results</th>
<th>Validity step achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual ED II</td>
<td>Cohort Study</td>
<td>Face</td>
<td>8/16 (MINORS)</td>
<td>Novice and experts (N=22)</td>
<td>Questionnaire: Favourable results on immersion, confidence, usefulness, content.</td>
<td>Face</td>
</tr>
<tr>
<td>OLIVE Cardiopulmonary resuscitation</td>
<td>Cohort study</td>
<td>Face</td>
<td>6/16 (MINORS)</td>
<td>Novices (N=12)</td>
<td>Questionnaire: Increase self-efficacy, concentration, no increase mental strain</td>
<td>-</td>
</tr>
<tr>
<td>Total Knee Arthroplasty</td>
<td>Cross-sectional Study</td>
<td>Face</td>
<td>9/16 (MINORS)</td>
<td>Novices (N=14)</td>
<td>Questionnaire: Easy to use, intuitive, stimulating</td>
<td>-</td>
</tr>
<tr>
<td>Virtual ED</td>
<td>RCT</td>
<td>Concurrent</td>
<td>LR (Cochrane tool)</td>
<td>Serious Game (N=16) vs. control (N=14)</td>
<td>Score improvement for both groups (posttest; p &lt; 0.05), no difference between groups</td>
<td>Concurrent</td>
</tr>
<tr>
<td>Triage Trainer</td>
<td>RCT</td>
<td>Concurrent</td>
<td>HR (Cochrane tool)</td>
<td>Serious Game (N=44) vs. control (N=47)</td>
<td>Increased triage tagging- and step accuracy by intervention group (p &lt; 0.05)</td>
<td>Concurrent</td>
</tr>
<tr>
<td>CAVE Triage</td>
<td>RCT</td>
<td>Concurrent</td>
<td>HR (Cochrane tool)</td>
<td>CAVE (N=7); control (N=8)</td>
<td>Effect size control group larger than CAVE group on knowledge (pre-/posttest design)</td>
<td>-</td>
</tr>
<tr>
<td>Nuclear Event Triage Challenge</td>
<td>Cohort study</td>
<td>Concurrent</td>
<td>10/24 (MINORS)</td>
<td>Serious game (N=40) vs. control (N=43)</td>
<td>Results favourable for intervention group (p &lt; 0.01). *</td>
<td>-</td>
</tr>
<tr>
<td>Radiation Hazards Assessment Challenge</td>
<td>Cohort study</td>
<td>Concurrent</td>
<td>10/24 (MINORS)</td>
<td>Serious game (N=40) vs. control (N=43)</td>
<td>Results favourable for intervention group (p &lt; 0.01). *</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3. Studies on validity testing of serious games specifically developed for educational purpose (Category 1), ranked according to steps in validation process tested and achieved.

Abbreviations: HR = High risk for bias; LR = Low risk for bias; MINORS = Methodological index for non-randomized studies; RCT = Randomized Controlled Trial;
Systematic review of validity of serious games for medical education and surgical skills training

_Triage Trainer_ (Blitz Games Studios, Ltd. Leamington Spa, UK) was compared to a conventional card-sort exercise in a non-randomized controlled trial among 91 doctors, nurses and paramedics30. Raters found a significant increase in triage accuracy for the _Triage Trainer_ group in post-test cases, compared to the control group, proving concurrent validity.

Triage training on the CAVE system was evaluated among 15 residents, who were randomly assigned to triage 14 cases in either CAVE or on human patient simulators. Concurrent validity was not proven as the control group performed significantly better on the posttest22.

_Nuclear Event Triage Challenge_ and _Radiation Hazards Assessment Challenge_ (HST division, Harvard Medical School, Boston, MA, USA) were evaluated in one study26. The author failed to describe the results in detail, so criteria for concurrent validity were not met.

The _Total Knee Arthroplasty_ game was tested for usability among 14 orthopaedic residents but criteria for face validity were not met21. The effect of cardiopulmonary resuscitation training on the On-Line Interactive Virtual Environment (Science Applications International Corporation, McLean, VA, USA) system was evaluated among 12 medical students for increase in self-efficacy but the study design did not correspond to formal validity testing27.

Steps in the validation process and face validity have been made for _Triage Trainer_30, _Virtual ED_18 (concurrent validity) and _Virtual ED II_29 (face validity) but these steps have yet to result into a completed formal validation process.

**Commercially available games associated with training skills.**

Six studies assessed thirteen commercially available games that were associated with (but not specifically designed for) training laparoscopic psychomotor skills (Category 2). Table 4 provides an overview of these games. They included sports games, action games, adventure games and shooting games on different platforms. Every game had an intrinsic scoring system. Performance in these games was compared to performance on different instruments for training laparoscopic psychomotor skills to test their concurrent validity. An overview of the results is provided in Table 531–36.

The study by Rosser _et al_34 showed a clear association between performance in three videogames (i.e. _Silent scope_ (Konami, Tokyo, Japan), _Star Wars Racer Revenge_ (LucasArts, San Francisco, CA, USA) and _Super Monkey Ball 2_ (Sega Corporation, Tokyo, Japan)) and laparoscopic handling speed and errors made in laparoscopic box trainer exercises, thus proving concurrent validity. Studies by Badurdeen _et al_31 and Rosenberg _et al_33 show clear correlations between laparoscopic handling speed and videogame performance in five games (_Amped 2, Top Spin_ (Microsoft, Redmond, Washington, USA); _Charge, Pose Mii, Shooting Range_ (Nintendo Co. Ltd. Kyoto, Japan)), compared to an animal model and a laparoscopic box trainer. This correlation only partially resembles concurrent validity, as movement proficiency was not significantly correlated.
Schlickum et al. randomized two groups of students to systematic videogame training with *Halflife* (Valve Corporation, Bellevue WA, USA) and *Chessmaster* (Ubisoft Entertainment, Inc., Montreuil, France). The *Halflife* group had significantly improved...
scores on a validated laparoscopy simulator and an endoscopy simulator and the Chess-
master group improved only on the endoscopy simulator. No clear correlation between
measured parameters was shown, so no concurrent validity was proven.

Studies on Marble Mania (Nintendo Co. Ltd., Kyoto, Japan) and Super Monkey Ball (Sega
Corporation, Tokyo, Japan) had insufficient design to draw conclusions on validity for
learning laparoscopic psychomotor skills32,35.

DISCUSSION

Serious games form an innovative approach towards the education of medical profes-
sionals and surgical specialties are eager to apply them to various training purposes.
They have been adopted for various different goals, e.g. as adjunct to existing simulator
training or as a stand-alone method. Two forces play a driving role in the development
and introduction of serious games. First, game-developers do not want to ‘miss out’ on
the medical market and may be afraid that thorough validation studies will postpone
their novelty. Second, the market is eager to adopt serious gaming as it appears to be
more attractive than learning ‘the old fashioned way’. Marketing and commercial forces
may lead to haphazard introduction of educational instruments that are not scrutinized
for their content in itself, nor for their proper transfer-of-content. It is important for
game designers and educators to cooperate in designing and validating a serious game
for a specific educational problem or hiatus. Only then do they have solid grounds to
integrate the serious game as a teaching tool in surgical curricula12,13. Serious games
first undergo testing of the system’s reliability, to address if the same measurement tool
yields stable and consistent results when repeated over time. Subsequently, the applica-
tion must undergo a validation process, preferably in the order described in Table 1.
Errors and deficiencies should be corrected when encountered. When the outcomes
of validity studies are unfavourable, the instrument cannot be seen as a valid teaching
instrument for a specific skill.

The search provided 17 serious games specifically designed for educational purpose
in medicine, of which several were of specific interest to surgical practice. Other games
were not directly linked to the surgical practice, but could be viewed as generally
interesting because of methods of education. Further research should define valid per-
formance parameters and complete formal validation programs, before serious games
can be seen as full-fledged teaching instruments for professionals in the medical and
surgical field.

Although a serious game does not necessarily have to be developed for an educational
purpose to be an educational tool, these games cannot be seen as fully completed train-
ing tools. All games found in the search in this category were used to train laparoscopic
<table>
<thead>
<tr>
<th>Serious Game</th>
<th>Study type</th>
<th>Validity steps tested</th>
<th>Methodological Quality</th>
<th>Groups (N)</th>
<th>Results</th>
<th>Validity steps achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Super Monkey Ball 2 34</td>
<td>Cohort study</td>
<td>Concurrent</td>
<td>11/16 (MINORS)</td>
<td>Population (mixed experience, N=32)</td>
<td>Significant correlation baseline VG score, score lap box trainer (3 exercises: time, error; ( p &lt; 0.001 ))</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>Star Wars Racer Revenge 34</td>
<td>Cohort study</td>
<td>Concurrent</td>
<td>11/16 (MINORS)</td>
<td>Attendings (N=12), Residents (N=22)</td>
<td>Significant correlation baseline VG score, score lap box trainer (3 exercises: time, error; ( p = 0.004 ))</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>Silent Scope 34</td>
<td>Cohort study</td>
<td>Concurrent</td>
<td>11/16 (MINORS)</td>
<td>Attendings (N=12), Residents (N=22)</td>
<td>Significant correlation baseline VG score, score lap box trainer (3 exercises: time, error; ( p = 0.003 ))</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>Amped 2 33</td>
<td>RCT</td>
<td>Concurrent</td>
<td>HR (Cochrane tool)</td>
<td>Systematic Gameplay (N=5) No gameplay (N=6)</td>
<td>No score improvement in laparoscopic tasks after systematic VG play. Significant correlation baseline VG score, baseline VG-score vs. time in laparoscopic tasks (animal model, ( p &lt; 0.05 ))</td>
<td>Partial concurrent validity (time)</td>
</tr>
<tr>
<td>Charge 31</td>
<td>Cross-sectional</td>
<td>Concurrent</td>
<td>10/16 (MINORS)</td>
<td>Serious Game (N = 20)</td>
<td>Correlation composite score on VG, score on lap box trainer ( r = 0.63 ) (( p &lt; 0.01 ))</td>
<td>Partial Concurrent validity (time)</td>
</tr>
<tr>
<td>Pose Mii 31</td>
<td>Cross-sectional</td>
<td>Concurrent</td>
<td>10/16 (MINORS)</td>
<td>Serious Game (N = 20)</td>
<td>Correlation composite score on VG, score on lap box trainer ( r = 0.78 ) (( p &lt; 0.01 ))</td>
<td>Partial Concurrent validity (time)</td>
</tr>
<tr>
<td>Shooting Range 31</td>
<td>Cross-sectional</td>
<td>Concurrent</td>
<td>10/16 (MINORS)</td>
<td>Serious Game (N = 20)</td>
<td>Correlation composite score on VG, score on lap box trainer ( r = 0.77 ) (( p &lt; 0.01 ))</td>
<td>Partial Concurrent validity (time)</td>
</tr>
<tr>
<td>Top Spin 33</td>
<td>RCT</td>
<td>Concurrent</td>
<td>HR (Cochrane tool)</td>
<td>Systematic Gameplay (N=5) No gameplay (N=6)</td>
<td>No improvement score laparoscopic tasks after systematic VG play. Significant correlation baseline VG score, baseline VG-score vs. time in laparoscopic tasks (animal model, ( p &lt; 0.05 ))</td>
<td>Partial concurrent validity (time)</td>
</tr>
<tr>
<td>Chessmaster 36</td>
<td>RCT</td>
<td>Concurrent</td>
<td>LR (Cochrane tool)</td>
<td>Novices (N=15)</td>
<td>Significant improvement score on 1 VR simulator (Gl-Mentor; ( p &lt; 0.05 )).</td>
<td>-</td>
</tr>
<tr>
<td>Halflife 36</td>
<td>RCT</td>
<td>Concurrent</td>
<td>LR (Cochrane tool)</td>
<td>Novices (N=15)</td>
<td>Significant improvement score on 2 VR simulators (MIST-VR; ( p &lt; 0.05 )) and Gl-Mentor; ( p &lt; 0.05 ).</td>
<td>-</td>
</tr>
<tr>
<td>Marble Mania 32</td>
<td>Prospective, Controlled trial</td>
<td>Concurrent</td>
<td>16/24 (MINORS)</td>
<td>Game (N=14), vs. No game (N=7)</td>
<td>Significant decrease time, errors, increase hand movement efficiency (( p &lt; 0.05 )) on VR simulator compared to control group</td>
<td>-</td>
</tr>
<tr>
<td>Serious Game</td>
<td>Study type</td>
<td>Validity steps tested</td>
<td>Methodological Quality</td>
<td>Groups (N)</td>
<td>Results</td>
<td>Validity steps achieved</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------</td>
<td>-----------------------</td>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Project Gotham Racing 2</td>
<td>RCT</td>
<td>Concurrent</td>
<td>HR (Cochrane tool)</td>
<td>Systematic Gameplay (N=5) No gameplay (N=6)</td>
<td>No score improvement laparoscopic tasks after systematic videogame play. No significant skills improvement in laparoscopic tasks (animal model)</td>
<td></td>
</tr>
<tr>
<td>Super Monkey Bal</td>
<td>Cohort Study</td>
<td>Concurrent</td>
<td>7/16 (MINORS)</td>
<td>Population (mixed experience, N=30)</td>
<td>Significant improvement of performance in pre-/posttest (4/8 measures in 3 tasks) on Lap box trainer after 10 min of videogame play ($p &lt; 0.05$)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.** Studies on validity testing of commercially available games associated with -but not specifically developed for- training medical skills (Category 2). Abbreviations: HR = High risk for bias; LR = Low risk for bias; MINORS = Methodological index for non-randomized studies; NFS = not further specified; RCT = Randomized Controlled Trial; VG = videogame; VR = Virtual reality;
psychomotor skills. Eight serious games showed a statistically significant correlation with speed of handling in specific tasks on a box trainer or a live animal model.\textsuperscript{31,33,34} However, only three of these videogames improved movement proficiency of laparoscopic handling in a box trainer setup, which implied concurrent validity for teaching of laparoscopic psychomotor skills\textsuperscript{34}. Until researchers have completed a full validation process for these games, they cannot be considered to be of true value in curricula for surgical resident training that currently employ validated VR simulators or laparoscopic box training and Objective Structured Assessment of Technical Skills (OSATS)\textsuperscript{37}. Schlickum et al\textsuperscript{36} showed that the improvement of performance on laparoscopic psychomotor abilities was not only accomplished by games that actually used the trainees' visuospatial ability, but also by simple two-dimensional games that required only cognitive and attention skills. Videogames have been shown to increase visuospatial and attention skills.\textsuperscript{38,39} Furthermore, visuospatial abilities and human visual working memory were associated with laparoscopic handling performance.\textsuperscript{40} The relationship between visuospatial and cognitive skills, videogames, and laparoscopic psychomotor skills is complex and therefore an interesting subject for future research. Optimizing the game metrics of these games to suit a validation process may lead to novel methods for teaching laparoscopic psychomotor skills. Challenging serious games for training laparoscopic psychomotor skills could lead to solutions for the popularity problem of VR simulators among surgical residents.\textsuperscript{41,42}

Serious games allow multiple professionals to simultaneously train on one case (teamwork) and allow one professional to train multiple cases simultaneously ('multitasking'). These non-technical skills are recognized as critical in reducing medical errors in dynamic high risk environments, like the operating room or the emergency department.\textsuperscript{43–45} The current trend in healthcare to reduce error in clinical practice has led to recognition of team training in managing crisis situations, e.g. anaesthesia crisis resource management (ACRM)\textsuperscript{46,47} and emergency medicine crisis resource management (EMCRM) and may also be of use to surgical residents.\textsuperscript{48} Crisis resource management (CRM) is derived from aviation and focuses on nurses and physicians together in crisis situations.\textsuperscript{49} Serious games allow such training in a relatively cheap, readily available environment with a large variety of cases, providing an alternative for expensive high-fidelity simulators.\textsuperscript{16–18,25} Serious games also present training environments for disaster situations and mass casualty incidents, including combat care.\textsuperscript{19,22,24,26,29,30} Realistic virtual surroundings in which sights, sounds and confusion are mimicked, provide a complete experience and improved preparation.\textsuperscript{22,30} Alongside the training of crisis management, serious games can be used for training everyday clinical activities and skills for junior doctors, e.g. decision-making abilities in surgical procedures or burn-patient care.\textsuperscript{19–21} Serious gaming as a way to prevent medical error will function optimally if games are designed to fit into residency teaching programs.\textsuperscript{50} Postgraduate education in most
Western countries is based on competency-based training, in which assessment and performance of the trainee is integrated. Competency frameworks like the CanMEDS framework\textsuperscript{51} have been developed for this purpose. The more recent introduction of Entrustable Professional Activities (EPAs), aimed at integrating these competencies in everyday clinical activities, allows a true outcome-based approach to specialist training\textsuperscript{52}. Simulation and serious gaming form ideal teaching methods to optimize knowledge and skill of residents before they are entrusted with procedures in real patients. Educators and game designers should direct serious game at training these entrustable professional activities to maximize their benefits for patient safety.

Initial development costs of serious games can be high, sometimes resulting in multimillion dollar projects\textsuperscript{53}. The expected revenue, in terms of better patient care and prevention of error, provides a decisive argument for investing in the development altogether. Insurance companies can play an important role. When a basic game-structure has been developed, it can function as platform for different institutions and departments to upload their content of choice\textsuperscript{16}. This can lead to games becoming widely usable training methods, keeping additional development costs relatively low.
REFERENCES


52 ten Cate O. Trust, competence, and the supervisor’s role in postgraduate training. *BMJ* 2006; 333(7571):748–751.
