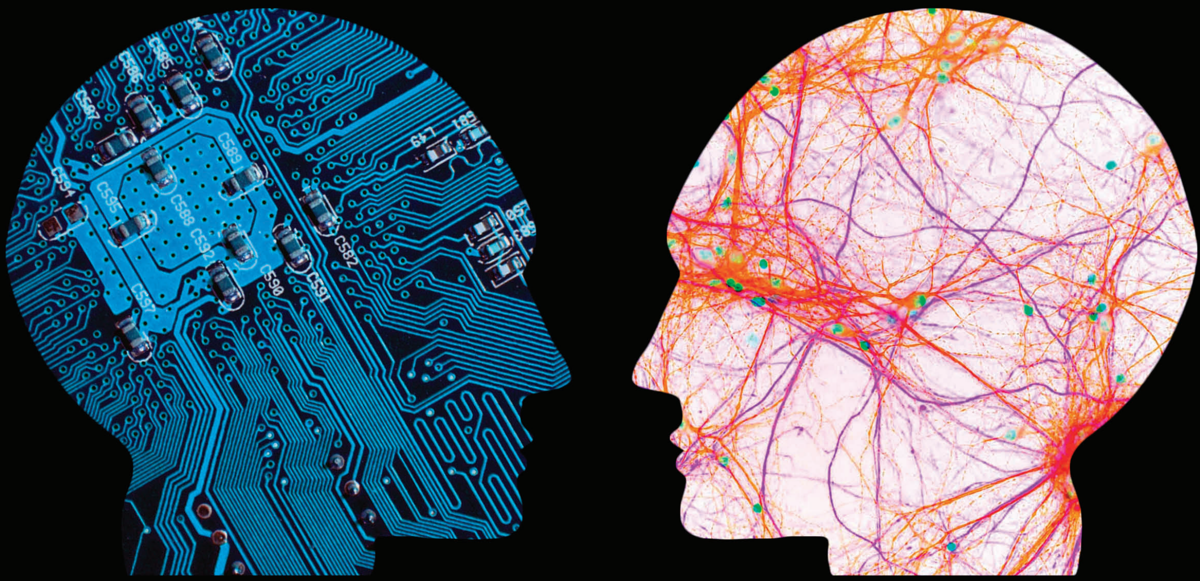
SERIOUS GAMES AND BLENDED LEARNING

Effects on Performance and Motivation in Medical Education



Mary Dankbaar

SERIOUS GAMES AND BLENDED LEARNING

Effects on Performance and Motivation in Medical Education

Mary Dankbaar

Dankbaar, Mary E.W.
Serious games and blended learning;
effects on performance and motivation in medical education
Thesis, Erasmus University Rotterdam, the Netherlands

ISBN: 978-94-6169-741-7
Cover design: Mathijs Doets & Mary Dankbaar
Lay-out and printing: Optima Grafische Communicatie, Rotterdam, the Netherlands

SERIOUS GAMES AND BLENDED LEARNING
Effects on Performance and Motivation in Medical Education

SERIOUS GAMES EN BLENDED LEREN
Effecten op Prestaties en Motivatie in het Medisch Onderwijs

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr.H.A.P Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 18 november 2015 om 11.30 uur

door

Maria Elisabeth Wilhelmina Dankbaar
geboren te Lichtenvoorde

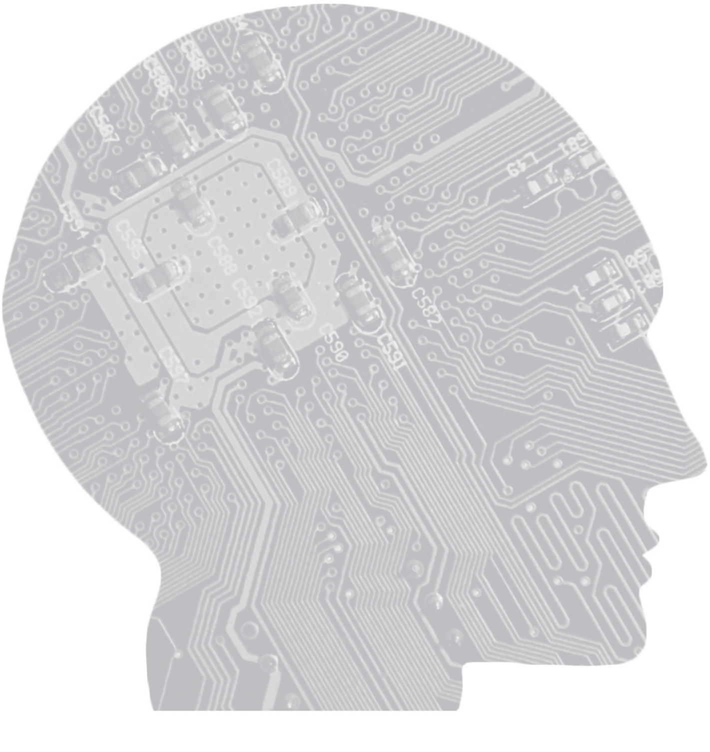
Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotoren:	Prof.dr.J.L.C.M. van Saase
	Prof.dr.J.J.G. van Merrienboer
Overige leden:	Prof.dr.M. de Hoog
	Prof.dr.M.P. Schijven
	Prof.dr.ir.A.P.N. Themmen
Copromotor:	Dr.S.C.E. Schuit

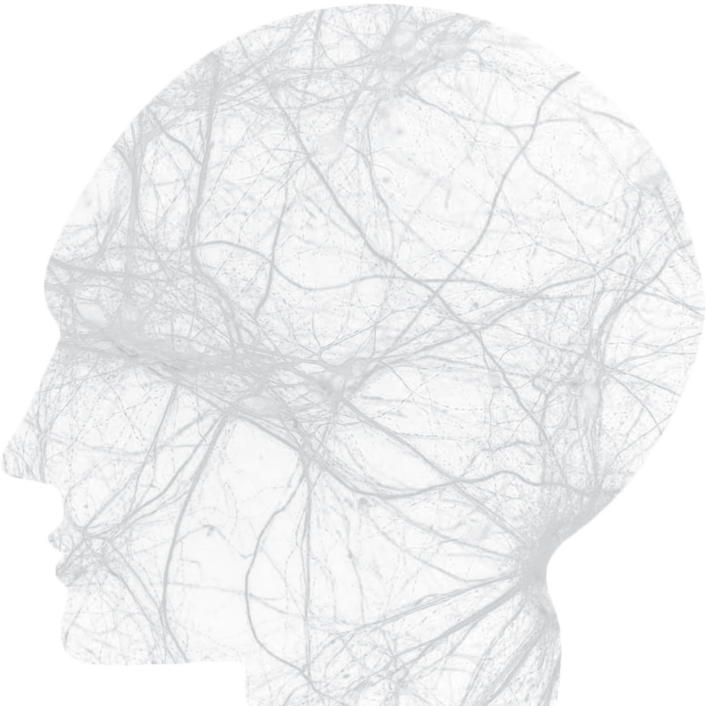
Voor mijn moeder

Contents



Chapter 1

Introduction



In health care and medical education, with its exponential growth in knowledge and increasing demands on the needed competencies of doctors, there is a need for new and cost-effective training models[1]. Online learning, web-based learning or e-learning is the use of Internet technologies to deliver a broad array of solutions that enhance learning and performance[2]. Online learning is a powerful concept with a variety of expression forms. It can be used for self-study or in groups, completely online or 'blended' with face-to-face training, and in different formats to acquire knowledge, skills or attitudes. Blended learning may combine the advantages of online and classroom learning. Its use is increasing quickly and a precipitous growth is predicted[3]. Just adding technology to classroom learning does not automatically result in better learning outcomes. Blended learning requires a course redesign, making the most of the educational affordances technology offers, and using classroom time for interaction and feedback[4]. As a result classroom time may be reduced, making it more cost-effective, or improved learning outcomes may be realized[5,6]. Some studies show blended learning is more effective than face-to-face learning[7,8], while others show no difference[9]. The **first aim of this thesis** is to investigate whether a blended training design is more effective or equally effective as classroom training for the acquisition of knowledge.

Online learning, as part of a blended course or as a fully online course, can be used to improve the efficiency, flexibility and effectiveness of education and training[1,10]. One of the challenges of designing online learning is that of engaging students during self-study. A serious game is a promising new format to develop complex cognitive skills, while motivating learners through direct experiences[11]. Further study is required on the effectiveness of serious games, including design choices involved in the development of games[12]. The **second aim of this thesis** is to investigate the effectiveness and critical design features of serious games for performance, specifically training complex cognitive skills and attitudes, while enhancing motivation.

Designing online or blended learning is a multi-level process, starting with the learning goals and the training setting, followed by choice of delivery of instruction (online, blended or classroom-based) and, if relevant, the selection of online formats, didactic features and presentation types. In order to structure design decisions and enable systematic comparison of the effectiveness of educational programs resulting from these decisions, a framework for instructional design choices is described. This chapter starts with a brief introduction of the digital revolution and online and blended learning. Next, the framework for instructional design is described, a summary of available research in relation to the design choices is provided and the research questions are formulated. This chapter ends with an outline of the research projects that together form this thesis.

THE DIGITAL REVOLUTION

Since the establishment of the World Wide Web in the 1990s, Internet has had a revolutionary impact on culture and commerce, including the rise of near-instant communication by e-mail, VOIP telephone calls, interactive video calls, discussion forums, social networking and online shopping sites. The Internet's takeover of the global communication landscape was almost instant in historical terms: It only communicated 1% of the information flowing through two-way telecommunications networks in the year 1993, 51% by 2000, and more than 97% of the telecommunicated information by 2007[13].
This digital revolution, including the mass use of computer technology and mobile technology, has profoundly changed our daily life. Worldwide information resources, services and communication tools are available anytime and anywhere. As a consequence, nowadays students have grown up in a media culture; information, services and connection to their peers is available at their fingertips. Computer games, email, the Internet, cell phones and messaging are integral parts of their lives[14].
In Europe, a recent study showed on an average day, 80-90% of the population watched television, a similar proportion used the phone and e-mail; ca. 45% browsed the Internet and a similar proportion read a book; ca. 28% played online games and 22% read an e-book[15]. Research shows there is no simple generational divide in these figures (e.g. between teachers and students) regarding the use of digital media; demographic factors (e.g. ethnicity, social class) have a much more important impact on access to technology and digital literacy, both between and within countries[16].
The music industry, the financial world, publishers and retail have changed significantly as a consequence of this digital revolution. In contrast, so far in the educational system technology has replaced standardized tasks, such as lecturing, only in a limited way[17].

ONLINE AND BLENDED LEARNING

Online learning has the potential to widen access to new learners, promote learner interactivity, increase flexibility in learning and ease in updating content[1,3,10]. These types of learning solutions can go beyond the traditional paradigms of training, as they can combine informal, non-intentional learning (e.g., sharing experiences through networks) with formal learning (e.g., attending a course). Rosenberg (2001) states that with the growing use of online learning, classroom learning will not be dismissed, but gain a new role in different learning designs. Content delivery will no longer be the main function in the classroom; there will be more reliance on high quality online content, less face-to-face lecturing and more coaching in the application of knowledge, evaluation of ideas and training of skills[2]. A specific form of blended learning, currently quite popular in higher education, is 'the flipped classroom'. Web lectures are watched by students at home and class time is dedicated to solving problems and interaction[6].

In blended learning (or hybrid learning), benefits of online and classroom learning can be combined, for example increasing access while maintaining or enhancing quality of learning outcomes and social interaction[3]. Blended and online learning are not single concepts; different mixes of online and face-to-face learning exist; different online formats (e.g., e-modules, web lectures, serious games) can be used, with different didactic features (information, exercises, cases) and presentation forms (text, video, high/low fidelity simulation). Modern theories of learning suggest that learning is most effective when it is active, experiential, situated, problem-based and provides immediate feedback[18]. Online formats offer different opportunities to design this type of learning. Some formats have been used for some time and are well researched (e-modules, web lectures), some are relatively new and promising with yet little consistent research results (serious games). For educationalists, it is important to make evidence-based choices in designing learning environments.

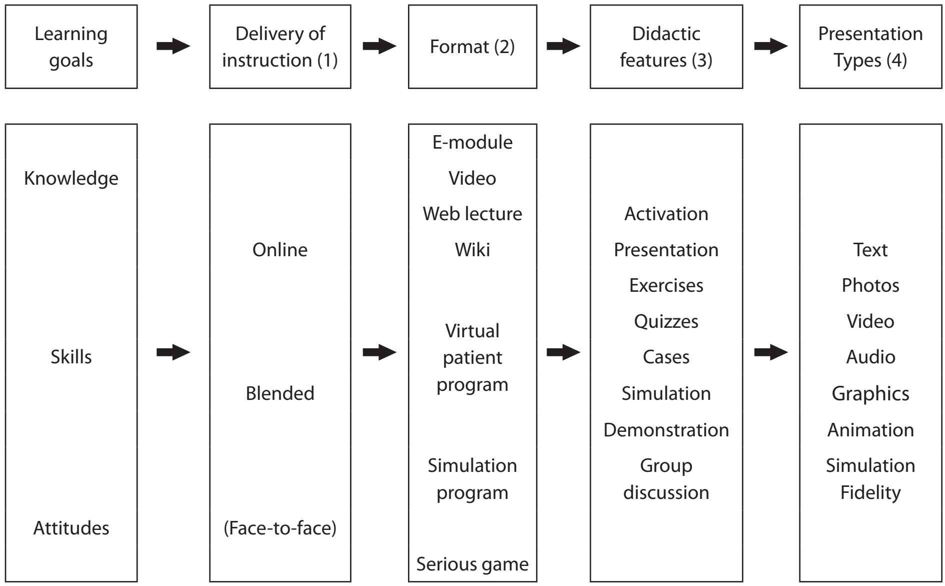
In the next section, a multi-level framework to structure design-choices in creating online learning is described. To avoid confounding, design-based research is best done within rather than between these design-levels, e.g, comparing the learning effects of didactic features[19]. The main research outcomes and open issues related to design choices will be briefly reviewed and mapped on this design framework. The research questions will then be described in relation to this framework.

Design framework

The framework integrates concepts of instructional design from Gagné and Merrill[20,21] with levels for designing online learning described by Cook[19]. An instructional design begins with establishing the learning setting and *learning goals*, preferably integrated into a comprehensive task which makes an appeal on competencies. Categories of learning objectives such as knowledge or cognitive or motor skills are identified within these integrative goals; different designs are needed for different learning goals or desired outcomes[21]. Many recent models of learning put real-world tasks and problems at the center of instruction, to support knowledge application and transfer[22]. Next, the delivery mode (1), the formats (2), didactic features (3) and presentation types (4) are chosen in a design (Figure 1).

Instruction delivery mode (step 1)

The first design-step is to choose a specific *delivery* mode for instruction: online or face-to-face instruction, or a mix of both, depending on the learning goals, the context and the student characteristics[4]. In the one extreme of the spectrum a small online



Based on Gagné & Merrill (1990) and Cook (2010)
Figure 1: Framework for Instructional Design choices for online or blended learning

supplement to a traditional course can be provided; in the other extreme online learning is the primary method of instruction, with a single classroom session or exam. In a blended design, the online learning goals tend to focus on acquisition of knowledge (e.g., through an e-module or web lecture), but skills and attitudes can also be taught online (e.g., through a simulation program or serious game)[23]. Classroom time can then be dedicated to training cognitive or motor skills on a higher level[6].

Despite the growing popularity of blended learning, little is known about the question whether blended learning is equally or more effective as classroom-based learning alone. Is it possible to reduce classroom-based training time, while increasing online self-study and maintaining learning outcomes? Especially for postgraduate medical education this is an important question, as online learning usually is more flexible and cost-effective than classroom-based learning. In some studies blended designs were reported to be more effective[7,8,24], while in others no differences in outcomes were reported compared to traditional learning[5,9,25]. Therefore our **first research question (Q1)** is: **What is the effectiveness of a blended training design compared to a conventional classroom-based training design in a postgraduate training program?**

When designing online learning, as a component in a blended design or as stand-alone, the next step is to choose *formats* (step 2). Formats (or configurations) define the way instruction looks and the functional and didactic possibilities it offers[26]. Selection of *didactic features* or methods (step 3) supports learning goals in more detail. They determine to a large extent the didactic effectiveness of a format. *Presentation types* (step 4) include text, audio, video, animations, simulations with particular levels of fidelity; they can enhance the didactic quality of a design. Figure 1 presents examples.

Formats (step 2)

In a course design a variety of instructional formats such as *e-modules, web lectures and wikis* can be used, with similar learning outcomes, but at different costs and access[27]. E-modules and web lectures are the online counterparts of traditional lectures; simulation programs and serious games are the counterparts of simulation training and role play. There is extensive research available on the effectiveness of instructional formats such as e-modules and web lectures. Studies have consistently demonstrated that they can achieve similar learning results compared to traditional instruction. This applies to a wide variety of learners, learning contexts (medical and non-medical), topics and learning outcomes such as knowledge, skills and attitudes[1,10,28,29]. Online instruction has a large effect compared with no intervention and has an effectiveness similar to traditional methods[30].

Virtual patient programs are interactive, computer-based clinical cases for developing clinical reasoning skills. A meta-analysis on the effectiveness of virtual patient programs has demonstrated that these programs are as effective as traditional non-computer teaching methods (standardized patients, clinical activities, a responsive mannequin) for knowledge, reasoning outcomes or satisfaction[31]. The most unique and cost-effective function of virtual patient programs is to facilitate and assess the development of clinical reasoning, using multiple and varied clinical cases[32].

Simulation programs can range from full-scale computer-based simulators or high fidelity mannequins to simple simulation programs or low fidelity mannequins. They provide learning opportunities for controlled skills practice, without risks to the patient[33]. Effectiveness of simulation programs has been well established. In comparison with no intervention, simulation programs in medical education are consistently associated with large effects for knowledge, skills and attitudes[33–35]. Full-scale computer-based simulators often are expensive, both in terms of initial purchase and running costs[36]. Furthermore, often little attention is paid to student motivation in simulation programs[37]; as a result learners tend to use them to reach certain learning outcomes but avoid systematic practice afterwards[38].

Serious games offer a challenging, experiential learning environment in which game characteristics such as a story-line, application of rules and principles and competition

are combined with educational goals[11,12,39–41]. An important category of serious games in education are simulation games; they offer learning tasks in a realistic, engaging online environment, where learners directly experience the consequences of their decisions[11,40,42]. In addition to games for training medical professionals, games have also been developed for patient treatment and health promotion. Several studies have demonstrated that games can positively affect patient health behaviors and outcomes; in comparison to health games, games for medical education are still in their infancy[43].
Experiential or task-centered learning in games can be associated with discovery learning[11]. The rationale for putting real-world tasks at the basis of a learning environment is to promote application of knowledge and transfer to practice[22]. Pure discovery learning, where learners actively work on tasks with little or no guidance, may be less effective than guided discovery learning, especially for novices[44,45]. The integration of fun and challenge can reduce stress and enhance motivation[46]. In addition, self-directed learning appears to enhance intrinsic motivation, as it affords a greater sense of autonomy[47]. If students are intrinsically motivated to learn, they are expected to spend more time on learning and feel more positive towards what they learn[48].
Because of the scalability of simulation games and simulation programs, they have, once developed, the potential to teach knowledge and skills that are typically acquired in simulation settings at a fraction of the costs[36].
In the last decade, several review-studies have been performed on the effects of serious games on learning outcomes and motivation; however they have shown mixed and ambiguous results, partly due to methodological flaws in the studies[12,18,42,49–52]. Hence justification studies, focused on the question of whether a game is effective in reaching its learning goals, are still needed (before clarification studies –how did it work- can be performed[53]). The serious game we choose to investigate on its effectiveness is designed to train complex cognitive emergency care skills for residents. Although worldwide over 1.5 million health care professionals attend certified courses in emergency care each year, little evidence exists on the validity and reliability of assessment instruments used in these courses[54]. As it is essential to use validated assessment instruments to assess the game efficacy, we first performed a validation study to answer the following, **second research question (Q2)**: **What is the validity and reliability of commonly used formats in the assessment of emergency care skills for residents?** Our **third research question is (Q3)**: **Do residents who use a serious game as an additional preparation for classroom training, show better complex cognitive skills before training than residents who only used a course manual as a preparation?**

Didactic features (step 3)

Didactic features are the critical elements in a format to match technology with learning goals[55]. Several features can be used in different formats. For instance, a presenter can explain concepts in a web lecture; with additional exercises and feedback learners are stimulated to actively process the information and integrate it with existing knowledge. A number of review studies have been done on the effectiveness of features (for different formats). Studies on designing instructional formats (e.g., e-modules) for knowledge, skills and attitudes show that effective features are: *feedback*, *practice exercises* (with medium to large effect sizes) and *interactivity*, *repetition* (with small effect sizes)[56,57]. Inconsistency among studies however was noticed and tempers conclusions; clear definitions of design features are essential for future comparative research[56].

For virtual patient programs, a review study has showed that *feedback, repetition* (until mastery), *advance organizers* and *contrasting cases* can improve learning outcomes[58]. For training skills with simulation programs, features such as *interactivity*, *repetitive practice and cases* with a *range of task difficulty and clinical variation* have proven to be effective[33,59].

For serious games, one meta-analysis has shown that games that stimulate learners to learn *actively*, which could be *accessed as many times* as desired and which were *supplemented* with face-to-face training were most effective for learning outcomes[42], another meta-analysis reported that *instructional support*, *modeling and feedback* can improve learning (especially for novices)[60]. *Spacing* (distributed learning) has been shown to positively influence students' learning outcomes in a number of studies[61,62]. *Adaptation* to learner's prior knowledge (e.g., through tests) improves achievement and decreases time required to learn material[63]. Learner-driven adaptivity, in which the learner is in control of the pace and sequence of instruction, is inherent to many online learning formats. In contrast, there is no evidence that learning styles exist and that there is a benefit of computer-driven adapted instruction to these 'styles'[64–66].

The effective didactic features we described for different formats overlap largely; they can be summarized as: stimulate active learning with practice exercises and cases, demonstration and feedback and repeated, spaced learning. Extensive evidence is available on the use of these features in longer existing formats such as e-modules, but less specific and less consistent research is available for newer formats such as serious games. For instance, evidence on the impact of features such as high levels of interactivity, immediate or postponed feedback and scoring systems on learning with serious games is needed. The fact that game-designs vary strongly is complicating the generalizability of research. Insight into the comparative effectiveness of a highly interactive serious game and a limited interactive, easier to develop e-module for motivation and performance would benefit our knowledge on design choices. The serious game we will investigate was designed to develop patient safety knowledge and awareness. It included video lectures, interactive biofeedback exercises and interactive patient cases. The e-module included text-based lectures with limited interactivity and information on the same topics. Therefore our **research question is (Q5): Do students develop better patient**

safety knowledge and awareness and are they more motivated after using a serious game than after using an e-module on the same topics?

Presentation types (level 4)

Presentation types include choices for multimedia, animations and simulation fidelity. For instance an animation may be used to demonstrate a physiological phenomenon. Regarding presentation types, a large body of research is available on *multimedia design principles* for learning, as multimedia (e.g., text and pictures) were one of the first components of computer-based learning. The studies on multimedia are often based on cognitive load theory[27] and the notion of a limited working memory. There is consistent evidence for a number of guiding principles which are especially important for novice learners in a certain domain. For example audio narration is superior to text when explaining a complex graphic, as text and graphics both use the visual modality of working memory (*modality principle*). Effect sizes reported in using the multimedia design principles with medical students were large for immediate and delayed tests[67]. These principles are summarized in Table 1[68,69].

Realistic cases in simulation have often shown to facilitate transfer to practice[70,71]. Simulation games offer learning tasks in a realistic, engaging computer-based environment[42]. How realistic should cases be presented? *Fidelity* is a multi-factorial concept. Functional fidelity pertains to the degree to which a simulated task environment reacts to the tasks executed by the learner in a similar way as the real task environment. Depending on the expertise of the learner, the properties of the simulation should be functionally aligned with the criterion task[72]. Physical fidelity pertains to the degree to which the simulated task environment looks, sounds and feels like the real environment[73]. *Higher-fidelity simulations* in emergency care seem to provide greater benefit than lower-fidelity simulation,

Name	**Multimedia design principle**
multimedia principle	The combination of graphics and words is preferable above words alone
contiguity principle	Align words closely (in place or time) to corresponding graphics, so they are in working memory at the same time. Avoid separation of feedback and responses
modality principle	Use audio narration rather than text to describe complex graphics. Presenting words to explain a graphic or table can overload the visual sub processor of working memory
redundancy principle	Explain visuals with audio narration or with text, not both
coherence principle	Adding extra material can inhibit learning. Avoid words, graphics that are not central to the goal. Simpler visuals lead to better learning than complex, realistic visuals
segmenting & pre-training principle	Manage complexity by breaking the material in parts. If students learn key concepts in pretraining, the training is more effective

Table 1: Multimedia design principles for presentation forms (Mayer, 2005)

although definitions of fidelity tend to vary[74]. However, in a review study by Norman et al. on the relationship between simulation fidelity and transfer of learning, four out of five studies have shown no superior outcomes for high-fidelity training for simple tasks (e.g., recognizing heart sounds or basic surgical skills); one study on a complex task (on critical care) did show a positive effect[75]. Motivational outcomes were not reported in this review. Considering the costs of creating high-fidelity training (e.g., highly realistic patient cases versus paper-based cases), more research into the added value of fidelity in learning cognitive skills, related to expertise levels of students, is needed[45]. The expertise level of the learner is relevant as high-fidelity cases may contain too many seductive details for novice learners and cause cognitive overload in combination with a complex task[73]. Therefore our **fourth research question is (Q4)**: **What are the effects of adding high-fidelity patient cases (a simulation game) compared to adding low-fidelity (text-based) cases to an instructional e-module on cognitive skills and motivation of medical students?** We will use the serious game on emergency care as the high-fidelity intervention in this study.

Serious games can be designed to accomplish different learning and behavioral outcomes, from supporting rehabilitation in disabled patients to training medical professionals. Currently, there is no systematic framework for classifying games, which complicates selection for medical education[18]. Furthermore, costs of performing an empirical effectiveness study following game production can be challenging for developers[43]. Therefore, we developed a framework to evaluate serious games in health care and education, in order to answer our **sixth research question (Q6)**: **How can serious games applied in health care and medical education be systematically evaluated?**
The context of our research study was medical education and training. The objects of our research studies were a post-graduate course for nurses on emergency care (Q1) and two serious games, both originally developed for medical residents. One game on emergency care, abcde*SIM*, developed by the Erasmus University Medical Center Rotterdam and Schola Medica[1], was used by residents (Q3) and medical students (Q4). Another game on patient safety, Air Medic Sky-1, developed by the University Medical Center Utrecht, was used by medical students (Q5).

In summary, this thesis focuses on the following aims and research questions (in the sequence of the chapters). The first aim was to investigate whether a blended training design is more effective or equally effective as face-to-face training in the acquisition of knowledge. Research question one is dedicated to this aim. The second aim was to investigate the effectiveness of serious games and the critical design choices for per-

[1] Training institution for family practitioners Utrecht, the Netherlands (former SBOH)

formance, specifically training complex cognitive skills and attitudes, while enhancing motivation. Research questions three through six are dedicated to this aim. Research question two is a supporting (validation) study.

1. What is the effectiveness of a blended training design versus a conventional classroom-based training design in a postgraduate training program? (Q1)
2. What is the validity and reliability of commonly used formats in the assessment of emergency care skills for residents? (Q2)
3. Do residents, who use a serious game as an additional preparation for classroom training show better complex cognitive skills before training than residents who only used a course manual as a preparation? (Q3)
4. What are the effects of adding high-fidelity patient cases (a simulation game) compared to adding low fidelity (text-based) cases to an instructional e-module on cognitive skills and motivation of medical students? (Q4)
5. Do students develop better patient safety knowledge and awareness and are they more motivated after using a serious game than after using an e-module on the same topics? (Q5)
6. How can serious games, applied to health care and medical education, be systematically evaluated? (Q6)

OUTLINE OF THE THESIS

Chapter 2 describes the study on the effectiveness of a blended training design versus a classroom- based training design in a postgraduate training program. In a retrospective study, one group of nurses took a traditional classroom-based course and another group took a blended course (one-third online and two-third classroom training) on emergency care. Examination and evaluation results were compared for both groups and cost-effectiveness was calculated.

Chapter 3 focuses on answering the research question: what is the validity and reliability of commonly used formats in the assessment of emergency care skills for residents? In order to answer this question, we conducted psychometric analyses of a checklist, a competency scale and a global performance scale; all commonly used in the assessment of training courses in emergency care. The validity and inter-rater reliability of the assessment instrument in a certified residents training were evaluated, using video-taped assessments.

In the study presented in **Chapter 4**, we answered the research question: do residents, who use a serious game as an additional preparation for classroom training show better cognitive skills before training than residents who only used a course manual as a preparation? We used the serious game abcde*SIM*, designed to train cognitive emergency care skills for residents as a base for this study. In a quasi-experimental design, with residents preparing for a rotation in the emergency department, one group received a course manual as preparation for face-to-face training and another group additionally received access to a serious game (on emergency skills) as a preparation for the same training. Emergency skills were assessed before training, using the validated assessment instruments described in Chapter 3.

The main objective of **Chapter 5** was to answer the question: what are the effects of adding high-fidelity cases (a simulation game) compared to low fidelity (text-based) cases to an instructional e-module on the cognitive skills and motivation of fourth-year medical students? We set up a three-group randomized design: a control group working on an e-module; a cases group, combining the e-module with low-fidelity text-based patient cases; and a game group, combining the e-module with a high-fidelity simulation game, with the same cases. Cognitive load and motivation were evaluated. After the study period, assessors rated students' cognitive emergency care skills in two mannequin-based scenarios.

Chapter 6 answers the question: do undergraduate medical students develop better patient safety knowledge and awareness and are they more motivated after using a (highly interactive) serious game or a (limited interactive) e-module on the same patient safety topics? Fourth-year medical students were randomly assigned to either a serious game (including video lectures and patient missions) or an e-module (including text-based lectures on the same topics). A third group acted as a historical control-group without extra instruction. Students were tested on knowledge, self-efficacy and motivation. During their clinical rotation they reported perceived stress and patient-safety awareness on a weekly basis.

In **Chapter 7**, answering the question how to systematically evaluate serious games applied to health care and medical education, we describe a general framework to evaluate serious games. The framework was developed together with members of the Dutch Society for Simulation in Healthcare with varying backgrounds in designing, applying and researching serious games. It provides a set of standardized criteria for developers, aimed at supporting end-users (educators and clinicians) in evaluating relevance and effectiveness of serious games.

In **Chapter 8** a general discussion of the findings of this thesis is provided, starting with the main conclusions from the six studies and the implications for designing blended learning and serious games. Next, the strengths and limitations of this thesis and the recommendations for future research are stated. We will conclude with the key points from this thesis.

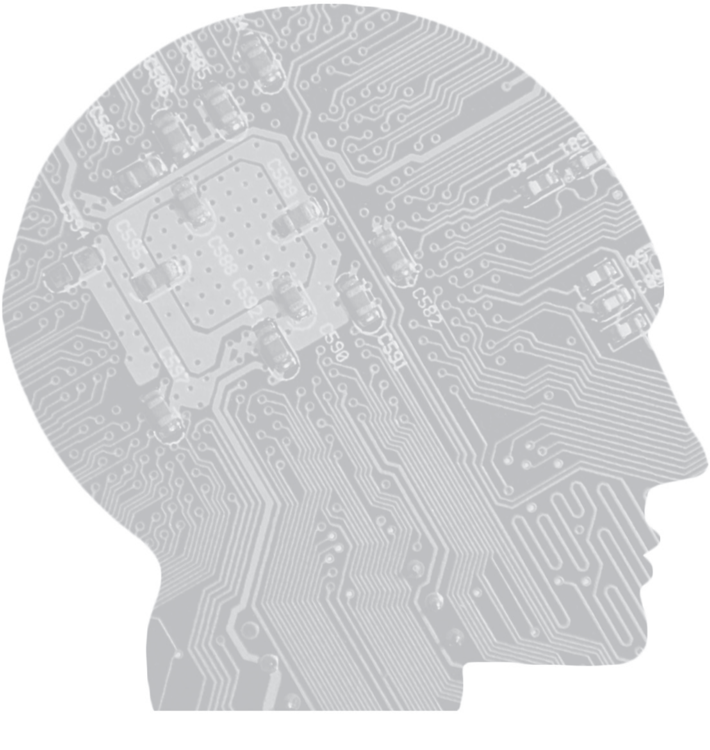
REFERENCES

1. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med. 2006 Mar;81(3):207–12.
2. Rosenberg M. E-learning: strategies for delivering knowledge in the digital age. New York: Mc Graw Hill; 2001. p. 28 p.
3. Moskal P, Dziuban C, Hartman J. Blended learning: A dangerous idea? Internet High Educ. Elsevier Inc.; 2013;18:15–23.
4. Bates TW. Teaching in a Digital Age. Bates, editor. www.tonybates.ca; 2015.
5. Bains M, Reynolds P, McDonald F, Sherriff M. Effectiveness and acceptability of face-to-face, blended and e-learning: a randomised trial of orthodontic undergraduates. Eur J Dent Educ. 2011 May;15(2):110–7.
6. McLaughlin JE, Roth MT, Glatt DM, Gharkholonarehe N, Davidson CA, Griffin LM, Esserman DA, Mumper RJ. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. Acad Med. 2014 Feb;89(2):236–43.
7. Abdelhai R, Yassin S, Ahmad MF, Fors UGH. An e-learning reproductive health module to support improved student learning and interaction: a prospective interventional study at a medical school in Egypt. BMC Med Educ. BioMed Central Ltd; 2012 Jan;12(1):11.
8. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ. American Physiological Society; 2005 Mar 1;29(1):35–9.
9. Beale EG, Tarwater PM, Lee VH. A retrospective look at replacing face-to-face embryology instruction with online lectures in a human anatomy course. Anat Sci Educ. 2013 Aug 19;
10. Bates T. Managing technological change. San Fransisco: Jossey-Bass; 2000.
11. Kiili K. Digital game-based learning: Towards an experiential gaming model. Internet High Educ. 2005;8:13–24.
12. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, Schünemann HJ. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach. 2010 Jan;32(1):16–27.
13. Wikipedia. History of the Internet [Internet]. accessed April 18, 2015. 2015. Available from: http://en.wikipedia.org/wiki/History_of_the_Internet
14. Prensky M. Digital Natives, Digital Immigrants Part 1 [Internet]. On the Horizon. 2001. p. 1–6. Available from: http://www.emeraldinsight.com/10.1108/10748120110424816
15. Helles R, Ormen J, Radil C, Jensen K. The Media Landscapes of European Audiences. Int J Commun. 2015;9:299–320.
16. Jones C, Shao B. The Net Generation and Digital Natives Implications for Higher Education. Milton Keynes; 2011.
17. Wetenschappelijke Raad voor het Regeringsbeleid. Naar een lerende economie (towards a learning economy). www.wrr.nl; 2013.
18. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. Elsevier Ltd; 2012 Sep;59(2): 661–86.

19. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med. 2005 Jun;80(6):541–8.
20. Gagne R, Briggs L. Principles of instructional design. USA: Holt, Rinehart and Winston; 1974.
21. Gagne RM, Merrill MD. Integrative Goals for Instructional-Design. Etr&D-Educational Technol Res Dev. 1990;38(1):23–30.
22. Francom GM, Gardner J. What is task-centered learning? TechTrends. 2014;58(5).
23. Cook DA, Dupras DM. A practical guide to developing effective web-based learning. J Gen Intern Med. 2004 Jun;19(6):698–707.
24. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. Arch Dis Child Fetal Neonatal Ed. 2013 Mar;98(2):F141–4.
25. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning: outcomes for staff development. J Nurses Staff Dev. 2012 Jul;28(4):186–90.
26. Cook DA, Garside S, Levinson AJ, Dupras DM, Montori VM. What do we mean by web-based learning? A systematic review of the variability of interventions. Med Educ. 2010 Aug;44(8):765–74.
27. Mayer RE. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005. 616 p.
28. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008 Sep 10;300(10):1181–96.
29. Sitzmann T, Kraiger K, Stewart D, Wisher R. The comparative effectiveness of web-based and classsroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.
30. Dankbaar MEW. De effectiviteit van e-learning en de implementatie in het medisch onderwijs. Tijdschr voor Med Onderwijs. 2009 Oct;28(5):212–22.
31. Cook D a, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010 Oct;85(10):1589–602.
32. Cook DA, Triola MM. Virtual patients: A critical literature review and proposed next steps. Med Educ. 2009;43(4):303–11.
33. Issenberg S, McGaghie W, Petrusa E, Gordon D, Scallese J. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;(4).
34. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ. 2010 Jan;44(1):50–63.
35. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek J, Wang A, Erwin P, Hamstra S. Technology-Enhanced Simulation for Health Professions Education; A systematic review and meta-analysis. JAMA. 2011;55905(September):978–88.
36. Kalkman CJ. Serious play in the virtual world: can we use games to train young doctors? J Grad Med Educ. 2012;4(March):11–3.
37. Kneebone R. Evaluating Clinical Simulations for Learning Procedural Skills: A Theory-Based Approach. Acad Med. 2005 Jun;80(6):549–53.
38. Jalink M, Goris J, Heineman E, Pierie J, Ten Cate Hoedemaker H. Construct and concurrent validity of a Nintendo Wii video game made for training basic laparoscopic skills. Surg Endosc Other Interv Tech. 2014;28(2):537–42.

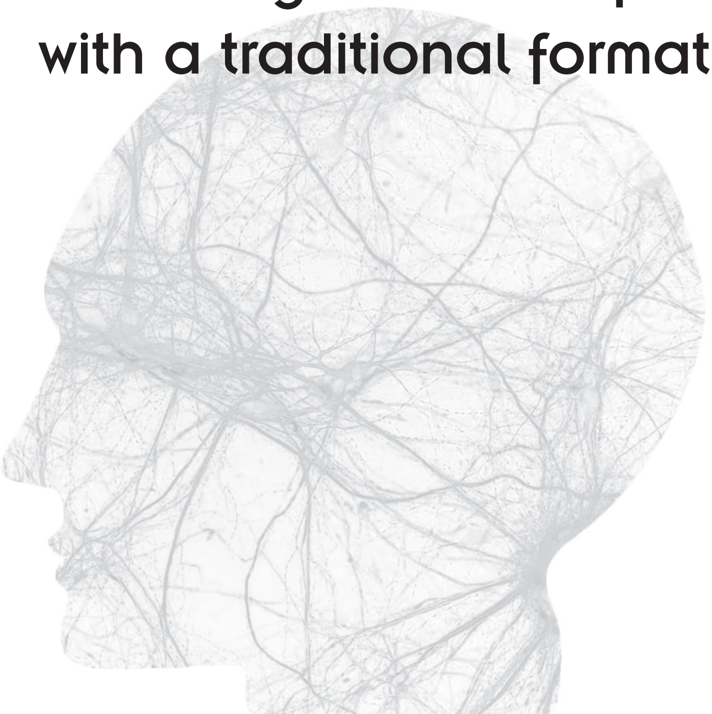
39. Van Staalduinen JP, de Freitas S. A First Step towards integrating educational theory and game design. In: Felicia P, editor. Research on improving learning and motivation through educational games: multidisciplinairy approaches. Hershey, PA: IGI Global; 2010. p. 28.
40. Clark RC, Mayer RE. E-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning. Elearning. San Fransisco: Pfeiffer; 2008. 476 p.
41. Huang WH, Huang WY, Tschopp J. Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing. Comput Educ. Elsevier Ltd; 2010;55(2):789–97.
42. Sitzmann T. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. Pers Psychol. 2011 Jun 27;64(2):489–528.
43. Kato PM. Video games in health care: Closing the gap. Rev Gen Psychol. 2010;14(2):113–21.
44. Mayer RE. Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. Am Psychol. 2004;59(1):14–9.
45. Baxter P, Akhtar N, Landeen J, Norman G. Teaching critical management skills to senior nursing students: videotaped or inter-active hands-on instruction? Nurs Educ Perspect. 2012;33(2): 106–10.
46. Allery LA. Educational games and structured experiences. Med Teach. 2004 Sep;26(6):504–5.
47. Ryan R, Deci E. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol. 2000;25:54–67.
48. Malone TW. Towards a Theory of Intrinsically Motivating Instruction. Cogn Sci. 1981;4:333–69.
49. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012 Oct;99(10):1322–30.
50. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol. 2013;105(2):249–65.
51. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, Simeoni Z, Tran M, Yukhymenko M. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. Rev Educ Res. 2012 Feb 1;82(1):61–89.
52. Vogel J, Vogel D, Canon-Bowers J, Bowers C, Muse K, Wright M. Computer gaming and interactive simulations for learning: a meta-analysis. J Educ Comput Res. 2006;34:229–43.
53. Cook DA, Bordage G, Schmidt HG. Description, justification and clarification: a framework for classifying the purposes of research in medical education. Med Educ. 2008 Feb;42(2):128–33.
54. Perkins GD, Hulme J, Tweed MJ. Variability in the assessment of advanced life support skills. Resuscitation. 2001 Sep;50(3):281–6.
55. Norman G. Research challenges in digital education. Perspect Med Educ. 2014 Sep;3(4):260–5.
56. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med. 2010 May;85(5):909–22.
57. Sitzmann, Tracy, Kraiger, Kurt, Stewart, David, Wisher R. The comparative effectiveness of web-based and clsssroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.
58. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010 Oct;85(10):1589–602.

59. Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Med Teach. 2013 Jan;35(1):e867–98.
60. Wouters P, van Oostendorp H. A meta-analytic review of the role of instructional support in game-based learning. Comput Educ. Elsevier Ltd; 2013 Jan;60(1):412–25.
61. Bude L, Imbos T, van de Wiel M, Berger M. The effect of distributed practice on students' conceptual understanding of statistics. High Educ. 2011;62(1):69–79.
62. Dempster F. Spacing Effects and Their Implications for Theory and Practice. Educ Psychol Rev. 1989;1(4):309–30.
63. Cook DA, Beckman TJ, Thomas KG, Thompson WG. Adapting web-based instruction to residents' knowledge improves learning efficiency: A randomized controlled trial. J Gen Intern Med. 2008;23(7):985–90.
64. Kirschner PA, van Merriënboer JJG. Do Learners Really Know Best? Urban Legends in Education. Educ Psychol. 2013 Jul;48(3):169–83.
65. Cook DA. Revisiting Cognitive and Learning Styles in Computer-Assisted Instruction. Acad Med. 2012;87(6):778–84.
66. Cook DA, McDonald FS. E-learning: is there anything special about the "E"? Perspect Biol Med. 2008;51(1):5–21.
67. Issa N, Schuller M, Santacaterina S, Shapiro M, Wang E, Mayer RE, DaRosa D a. Applying multimedia design principles enhances learning in medical education. Med Educ. 2011 Aug;45(8):818–26.
68. Mayer RE. The Cambridge handbook of multimedia learning. Cambridge University Press; 2005.
69. Clark, Ruth C Mayer RE. E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. 3rd ed. San Fransisco: Pfeiffer; 2008. 424 p.
70. Bransford J, Brown A, Cocking R. How people learn: Brain, Mind, Experience, and School. Washinton DC: National Academic Press; 2000. 3-127 p.
71. Schuwirth LWT, van der Vleuten CPM. The use of clinical simulations in assessment. Med Educ. 2003 Nov;37 Suppl 1:65–71.
72. Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? Adv Health Sci Educ Theory Pract. 2012 Mar;17(1):137–44.
73. Van Merrienboer JJ, Kirschner PA. Ten steps to complex learning. A systematic approach to four-component instructional design. 2nd Editio. Routledge; 2012. 49-77 (Chapter 4) p.
74. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ. 2015;49:161–73.
75. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. Med Educ. 2012 Jul;46(7):636–47.



Chapter 2

A blended design in acute care training: similar learning results, less training costs compared with a traditional format



Mary EW Dankbaar
Diana J Storm
Irene C Teeuwen
Stephanie CE Schuit

Perspectives on Medical Education, 2014 January

ABSTRACT

Introduction

There is a demand for more attractive and efficient training programmes in postgraduate health care training. This retrospective study aims to show the effectiveness of a blended versus traditional face-to-face training design. For nurses in postgraduate Acute and Intensive Care training, the effectiveness of a blended course design was compared with a traditional design.

Methods

In a first pilot study 57 students took a traditional course (2-h lecture and 2-h workshop) and 46 students took a blended course (2-h lecture and 2-h online self-study material). Test results were compared for both groups. After positive results in the pilot study, the design was replicated for the complete programme in Acute and Intensive Care. Now 16 students followed the traditional programme (11 days face-to-face education) and 31 students did the blended programme (7 days face-to-face and 40 h online self-study). An evaluation was done after the pilot and course costs were calculated.

Results

Results show that the traditional and blended groups were similar regarding the main characteristics and did not differ in learning results for both the pilot and the complete programme. Student evaluations of both designs were positive; however, the blended group were more confident that they had achieved the learning objectives. Training costs were reduced substantially.

Conclusion

The blended training design offers an effective and attractive training solution, leading to a significant reduction in costs.

INTRODUCTION

During the last decades, a large number of studies have been performed on the effectiveness of online learning, or technology-enhanced learning. From these studies with students and professionals, both inside and outside the medical domain, the conclusion can be drawn that online learning is at least as effective as more traditional forms of learning, for knowledge and skills and students are equally satisfied with both forms[1–5]. In some studies the combination of online and face- to-face learning (blended learning) has been found to be more effective than traditional learning alone[2,4,6] while other studies have shown the same results[7]. There is, however, still little research on the optimal mix of online and instructor- led learning[8].

In this article, we will describe a retrospective effectiveness study on the blended and the face-to-face design of a programme in Acute and Intensive Care for nurses. We will describe the methods we used to analyze its effectiveness and the learning and evaluation results. We will end with a discussion on the cost- effectiveness and implications for health care practice.

The training centre for health professionals of the Erasmus University Medical Center provides postgraduate, registered nurse education programmes for Intensive, Emergency and Cardiac Care. The basic education course in Acute and Intensive Care forms an important baseline for all specialized postgraduate nurse training programmes in Intensive, Emergency and Cardiac Care. These continuous education programmes take 6–18 months and are taken by nurses in combination with their work in hospitals. The programme in Acute and Intensive Care consists of three parts: a respiration, circulation and central nervous system course. The respiration course includes a section on acid–base balance. The students generally experience this specific part of the programme as difficult and were having problems in applying the principles in practice. We looked for improvements in the quality and efficiency of this programme, by applying a blended learning concept, using design features such as practice cases, multimedia, feedback and repetition, which have proven to be effective in online learning[9,10]. This online study material was offered through a permanently accessible Learning Management System (It's Learning). The students can decide when to study, (re)do the exercises and whether to collaborate with colleagues or do them alone.

This new design was first launched as a pilot, in order to be able to evaluate the results before implementing it for the rest of the programme. After analyzing the effectiveness of and student reactions to the new course design, the other components of the programme in Acute and Intensive Care were also redesigned to a blended concept. The length of the face-to-face part of the programme was reduced from 11 to 7 days; an online component was added which took ± 40 h of self-study.

If this new, blended course design is effective and appreciated by students, an improvement in course quality and efficiency in training time (and thus cost reduction) can be realized. In this time of reduced budgets and growing demands on knowledge and skills of health care professionals, this is an important asset for health care organizations. For students, blended learning not only offers the opportunity of flexible, 'anytime, anywhere' learning, adaptable to work pressure and personal conditions. It also offers the opportunity to personalize learning: specific, complex parts of the content can be exercised as often as desired, until they are profoundly understood and can be applied in practice, without risks for patient care[3]. Although there are a large number of articles on online learning, evaluation studies remain less common[11]. Comparing two instructional methods or formats in terms of learning outcomes is relevant to bring the field of online learning further[12].

METHODS

Traditional and blended design of the pilot program

The traditional design of the pilot programme ('acid–base balance', part of the course on respiration) consists of a 2-h face-to-face (f2f) lecture and a 2-h f2f workshop (practice exercises on acid–base balance, blood gas assessment, etc.). The new 'blended' design of this programme consists of the same 2-h lecture; the workshop is replaced by 2-h online self-study material. The material includes: short web lectures (explaining essentials), a range of exercises and examples with feedback.

Participants of the pilot

The test results of the students (n = 57) of the traditional 'acid base' part of the spring 2011 programme were compared with the test results of the blended design group in autumn 2011 (n = 46). In order to evaluate the appreciation by students, we used a survey which was sent to both groups after the course.

Traditional and blended design of the program

After analysis of the results and the conclusion that the new approach turned out to be successful, the blended design was applied to the rest of the programme on Acute and Intensive Care for all students.

The three main parts—respiration, circulation and central nervous system—were redesigned, reducing the number of face-to-face training days from 11 to 7, spread over 2 weeks. The seven training days mainly consisted of lectures, explaining the course material. The (original) 4 days of workshops were replaced by online study material with the same format as in the pilot (short web lectures, examples and exercises with feedback). Several exercises now had a game format. This takes about 40 h of self-study and can be used to prepare for the lectures and for self-study.

Participants of the program

Students from the basic programme on Acute and Intensive Care in January 2012 followed a traditional design programme (n = 13–16, depending on the specific test on respiration, circulation or the nervous system) and students from the same programme in May 2012 followed a blended design (n = 27–31, depending on the test). All students have experience as a nurse and are currently working as trainees on a special care unit. They all started with the basic programme on Acute and Intensive Care and continued with a specialized training programme. In Table 1 the characteristics of the two groups are compared.

Knowledge tests

The knowledge test on 'acid–base balance' includes 15 questions, as part of 60 multiple choice questions on respiration (2–4 alternatives). It is part of a summative exam in which the pass/fail cut-off is defined in a test matrix beforehand. The 15 test questions from both pilot groups are drawn from the same question pool, in line with the test matrix. They are equivalent in difficulty and in the number of alternatives.

The Acute and Intensive Care programme has three separate knowledge tests, in time sequence: on the central nervous system (45 questions), on circulation (53 questions) and on respiration (60 questions). The tests are summative exams and similar to the pilot test (multiple choice answers with 2–4 alternatives, pass/fail cut- off defined beforehand, questions are equivalent in difficulty and in the number of alternatives). All tests were taken at the end of each programme part. The test results of participants of the traditional (group 1) and blended group (group 2) of the pilot (2011) were compared and, a year later, the test results of the participants of the complete programme (2012) were again compared for group 1 and 2.

Evaluation

A short evaluation was done after the pilot programme; the survey was sent to the participants, including a number of statements (4-point scale) and open questions.

Statistics

We did a reliability analysis of the knowledge tests (Cronbach's alpha), T-tests to compare means of the knowledge test results and a Mann–Whitney test to compare the evaluation results, using SPSS version 20.

Calculation of training costs

We calculated the training costs for a hospital to have employees follow the traditional or blended design, by comparing the direct and indirect costs, assuming: (a) self-study time is not paid; (b) travel costs are €20– per day on average; (c) indirect costs (costs for the absence of an employee) for nurses are €50– per hour on average, €400– per day.

RESULTS

Comparability of groups

In Table 1 the characteristics of the participants in the pilot group and the complete programme on Acute and Intensive Care are described for group 1 (traditional) and 2 (blended). Although students in the traditional pilot group [1] were slightly younger and

	Pilot group 1 Traditional 2011 (n=57)	Pilot group 2 Blended 2011 (n=46)	P-value	Course group 1 Traditional 2012 (n=16)	Course group 2 Blended 2012 (n=31)	P-value
Age			0.92			0.99
Age < 30	40 (70%)	29 (63%)		9 (56%)	17 (55%)	
Age ≥30	17 (30%)	17 (37%)		7 (44%)	14 (45%)	
Sex			0.93			0.62
Female	46 (80%)	39 (85%)		8 (50%)	24 (77%)	
Male	11 (20%)	7 (15%)		8 (50%)	7 (23%)	
Previous education			0.88			0.86
Higher education	28 (50%)	18 (39%)		7 (44%)	10 (32%)	
Intermediate education	29 (50%)	28 (61%)		9 (56%)	21 (68%)	
Type Hospital			0.89			0.95
University hospital	21 (36%)	21 (46%)		6 (38%)	13 (42%)	
General hospital	36 (64%)	25 (54%)		10 (62%)	18 (58%)	
Specialisation			0.99			0.99
Intensive Care	18 (32%)	8 (17%)		6 (37,5%)	5 (16%)	
Emergency Care	10 (18%)	7 (15%)		4 (25%)	10 (32%)	
Cardiac Care	6 (10%)	6 (13%)		0	1 (3%)	
Others	23 (40%)	25 (54%)		6 (37,5%)	15 (48%)	

Table 1: Characteristics of participants in the research groups

more often followed higher education compared with the blended pilot group [2], there were no significant differences between the two groups for the pilot or complete course.

Reliability of knowledge tests

The reliability (Cronbach's alpha) for the different test versions on 'acid–base balance' was between 0.37 and 0.67 (for group 1 the average α of the test versions was 0.57; for group 2 the average α was 0.47). This reliability is relatively poor, probably because of the small number of questions [15]. The Cronbach's alpha for the different test versions of the tests on respiration, circulation and the central nervous system was between 0.35 and 0.77 (the average α of the test versions for group 1 was 0.63, for group 2 it was 0.65); this reliability is moderate. The tests on circulation for both groups had a good reliability (0.77 and 0.68).

Pilot group: test results

Group 1 and 2 both answered 76 % questions about the acid–base balance correctly: the blended group had the same results on the knowledge test as the traditional group.

Question	**Students group 1 (traditional, n=22)**	**Students group 2 (blended, n=31)**	**Mann–Whitney scores** (U and P value)
I have achieved the course objectives			
not at all	0%	0%	U = 390
somewhat	17%	3%	P = 0.023
reasonable well	22%	13%	
very well	50%	82%	
no answer	11%	2%	
This education format suits my learning style			
not at all	0%	0%	U = 393
somewhat	0%	0%	P = 0.100
reasonable well	67%	41%	
very well	33%	53%	
no answer	-	6%	
I can now understand the clinical characteristics of a patient from the blood gas analyses			
not at all	0%	3%	
somewhat	40%	9%	U = 364
reasonable well	33%	73%	P = 0.477
very well	27%	12%	
no answer	-	3%	

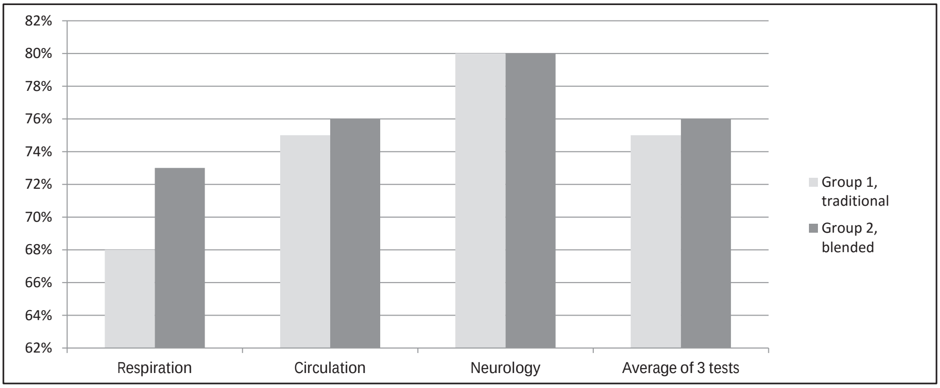
Table 2: Opinions of students in both groups (n=53)

Pilot group: evaluation of the new design

The participants of the traditional and blended group evaluated the pilot programme roughly equally positively, except for the statement 'I have achieved the course objectives'. The blended group was more self-confident in this statement (U = 390, P = 0.023, Table 2). Comments on open questions from students were: 'The explanation of the teacher during the workshop was good' (group 1) and 'It's pleasant to be able to practise independently', 'frequently repeating the exercises is useful' (group 2).

Acute and Intensive Care program: test results

As Graph 1 shows, the test results of the blended group [2] on the respiration test were slightly better compared with group 1 although not significant (P > 0.10). The respiration test (60 items) contains more abstract content compared with the other tests. The results on the circulation (53 items) and neurology (45 items) tests were similar for both groups. Both groups scored better on the neurology test, probably because of the smaller number of items.



Respiration	Circulation	Neurology
Gr1: 68%, SE = .02	Gr1: 75%, SE = .03	Gr1: 80%, SE = .03
Gr2: 73%, SE = .02	Gr2: 76%, SE = .02	Gr2: 80%, SE = .02
p > .10	p > .50	p > .40

Graph 1: Test results for the total course, group 1 and 2: percentage right answered questions

Training costs

The costs of training health personnel include direct training costs (course fee and travel expenses) and indirect costs (costs for the absence of an employee). The blended Acute and Intensive Care course has been reduced in price by the Erasmus MC training centre from € 1,350 to € 1,270 compared with the traditional format. As contact time is reduced by 4 days, travel costs are reduced by € 80 per employee on average. Direct training costs for a health organization are therefore reduced to € 160 per employee. Absence from work is reduced from 11 to 7 days, for 32 h the indirect cost saving is € 1,600 for nurses on average.

For hospitals that send their postgraduate nurses to training, direct and indirect costs of the training can be reduced by € 1,760 per person (€ 5,970 for the traditional model to € 4,210 for the blended model) by using this blended training model for the course on Acute and Intensive Care.

DISCUSSION

We found similar learning results for a blended course design (one-third less contact time and more self-study) compared with a traditional design (more contact time and less self-study). Looking at the background characteristics of the two groups, a higher score for the traditional group was to be expected, because they are younger and more highly educated. Reducing the face-to-face (f2f) training time by one-third (from 11 to 7 days in 2 weeks' time) and adding 40 h of online education led to equally effective learning. Participants were satisfied with the more interactive way of learning (although they needed to spend more free time studying) and were more confident that they had achieved the goals of the course. On the respiratory test (experienced as the most 'difficult' one), they performed somewhat, but not significantly better (although the number of participants was small). Particularly with complex subjects, the blended design offers the possibility for customized, frequent exercise.

Although we did not randomize the two student groups and cannot be sure what their knowledge levels were before the training, we were able to confirm that they are comparable on a number of important characteristics. These findings were also confirmed in other small-scale studies[13]. Further research with larger groups is necessary to validate these outcomes, preferably randomized studies with a posttest- only design[14]. One study with a randomized trial on traditional lecturing versus additional online learning (blended) reported higher levels of newborn examination skills in the blended learning group[15]; however, this may be caused by a longer learning time. In another randomized trial learning time was fixed, and the same results for the f2f and blended design were reported[16]. A limitation of this study is the fact we do not have data on the learning time (the response from students was too low to make reliable conclusions). Very few studies report on learning time in blended design studies. One qualitative study reported that students welcome the flexibility, but some feel the online component is invasive in their everyday life, specifically following a day at work[17].

In addition, it would be interesting to find out more about the most optimal mix and conditions for implementation of a blended design. The group in this research was motivated, as they work in a clinical setting where knowledge on acute care subjects is essential. For other groups who are not yet working, the results might be different and another blended mix might be advisable.

As the total training costs are significantly reduced by 30 %, this training design is an attractive model for health organizations that want to offer efficient and cost- effective training. For the delivering organization, redesigning a course requires both time and expertise in the development of effective and attractive e-learning material. Although the initial investment is high, the investment is worth the costs for training institutions with a large number of participants, because teacher costs are also reduced. In addition, it stimulates evaluation of the didactic quality of the course. The Erasmus University MC training centre has decided to implement the blended design for all specialized training programmes for nurses.

Conclusions

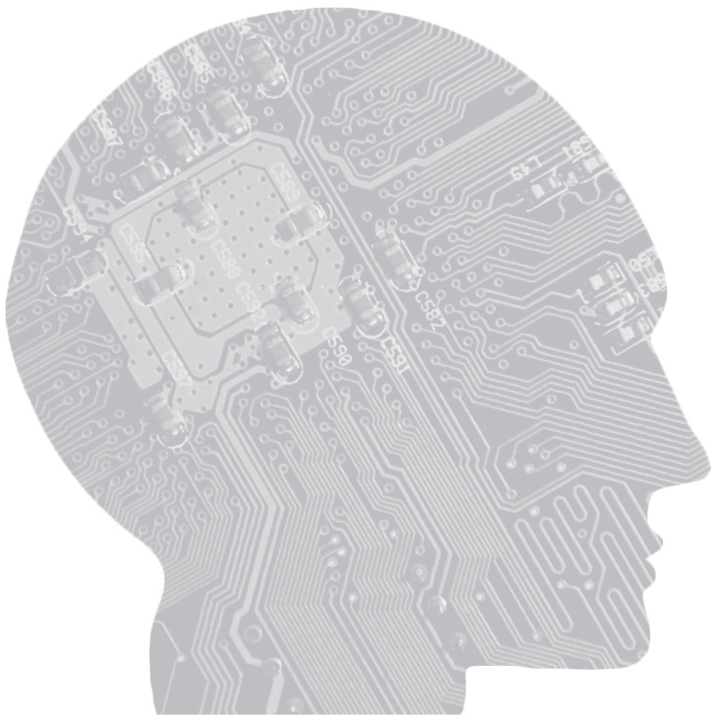
For the basic Acute and Intensive Care programme at Erasmus MC we have shown that when a course is redesigned from 11 days face-to-face training (in 2 weeks) to a blended design of 7 days f2f training, complemented by online self-study material, learning outcomes remain the same. Students are satisfied with this new design, as they can tailor learning to their own needs and it offers flexibility. For the affiliated hospitals it is important that the training costs are reduced and nurses are more available for patient care.

Essentials

- Blended learning, with one-third less face-to-face training time and more online self-study is equally effective in learning outcomes for a postgraduate course on Acute and Intensive Care.
- Highly interactive online self-study material enables students to study complex subjects in a customized and flexible way, being able to exercise as frequently as they wish.
- Students are satisfied with the new blended design, although it implies more self-study time.
- For health organizations a blended design leads to a significant reduction of costs, mainly because of savings in indirect training costs. Health professionals are more available for patient care.
- More research is needed with larger and different groups of participants and courses to learn more about the optimal blended mix and implementation conditions.

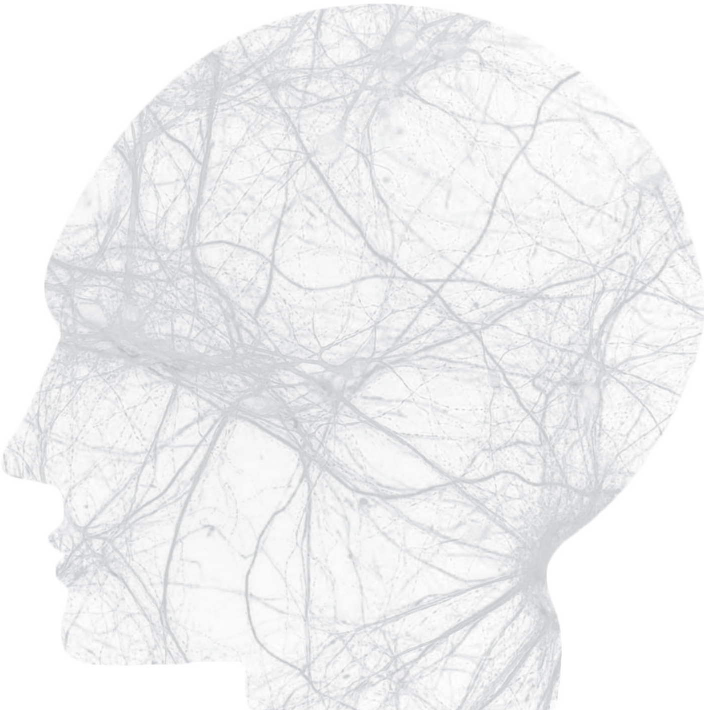
REFERENCES

1. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008 Sep 10;300(10):1181–96.
2. Sitzmann T, Kraiger K, Stewart D, Wisher R. The comparative effectiveness of web-based and classsroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.
3. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med. 2006 Mar;81(3):207–12.
4. Abdelhai R, Yassin S, Ahmad MF, Fors UGH. An e-learning reproductive health module to support improved student learning and interaction: a prospective interventional study at a medical school in Egypt. BMC Med Educ. BioMed Central Ltd; 2012 Jan;12(1):11.
5. Sung YH, Kwon IG, Ryu E. Blended learning on medication administration for new nurses: integration of e-learning and face-to-face instruction in the classroom. Nurse Educ Today. Elsevier Ltd; 2008 Nov;28(8):943–52.
6. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ. American Physiological Society; 2005 Mar 1;29(1):35–9.
7. Beale EG, Tarwater PM, Lee VH. A retrospective look at replacing face-to-face embryology instruction with online lectures in a human anatomy course. Anat Sci Educ. 2013 Aug 19;
8. Dankbaar MEW. Online leren: werkt dat? Opleid & Ontwikk. 2012;(3):42–7.
9. Clark RC, Mayer RE. E-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning. Elearning. San Fransisco: Pfeiffer; 2008. 476 p.
10. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. Acad Med. 2010 May;85(5):909–22.
11. Adler MD, Johnson KB. Quantifying the literature of computer-aided instruction in medical education. Acad Med. 2000 Oct;75(10):1025–8.
12. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med. 2005 Jun;80(6):541–8.
13. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning: outcomes for staff development. J Nurses Staff Dev. 2012 Jul;28(4):186–90.
14. Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract. 2010 Aug;15(3):455–64.
15. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. Arch Dis Child Fetal Neonatal Ed. 2013 Mar;98(2):F141–4.
16. Bains M, Reynolds P, McDonald F, Sherriff M. Effectiveness and acceptability of face-to-face, blended and e-learning: a randomised trial of orthodontic undergraduates. Eur J Dent Educ. 2011 May;15(2):110–7.
17. Smyth S, Houghton C, Cooney A, Casey D. Students' experiences of blended learning across a range of postgraduate programmes. Nurse Educ Today. 2012 May;32(4):464–8.



Chapter 3

Assessing the assessment in emergency care training



Mary EW Dankbaar
Karen M Stegers-Jager
Frank Baarveld
Jeroen JG van Merrienboer
Geoff R Norman
Frans L Rutten
Jan CM van Saase
Stephanie CE Schuit

Plos-One 2014 ***9(12):*** e114663

ABSTRACT

Objective

Each year over 1.5 million health care professionals attend emergency care courses. Despite high stakes for patients and extensive resources involved, little evidence exists on the quality of assessment. The aim of this study was to evaluate the validity and reliability of commonly used formats in assessing emergency care skills.

Methods

Residents were assessed at the end of a 2-week emergency course; a subgroup was videotaped. Psychometric analyses were conducted to assess the validity and inter-rater reliability of the assessment instrument, which included a checklist, a 9-item competency scale and a global performance scale.

Results

A group of 144 residents and 12 raters participated in the study; 22 residents were videotaped and re-assessed by 8 raters. The checklists showed limited validity and poor inter-rater reliability for the dimensions "correct" and "timely" (ICC= .30 and .39 resp.). The competency scale had good construct validity, consisting of a clinical and a communication subscale. The internal consistency of the (sub)scales was high (α=.93/.91/.86). The inter-rater reliability was moderate for the clinical competency subscale (.49) and the global performance scale (.50), but poor for the communication subscale (.27). A generalizability study showed that for a reliable assessment 5-13 raters are needed when using checklists, four when using the clinical competency scale or the global performance scale.

Conclusions

This study shows poor validity and reliability for assessing emergency skills with checklists but good validity and moderate reliability with clinical competency or global performance scales. Involving more raters can improve the reliability substantially. Recommendations are made to improve this high stakes skill assessment.

INTRODUCTION

Background

Worldwide, over 1.5 million health care professionals attend a variety of certified emergency care courses each year, in which they are trained for the initial assessment and management of seriously ill patients, using a standardized approach[1]. At the end of most of these advanced emergency care courses, participants are assessed on their knowledge, typically using a multiple-choice test, and on their skills, typically using a single scenario test with a simulated patient and 1 or 2 raters, depending on the course, the country and the institution[2,3].

The development of competencies in emergency care is a core component in both undergraduate and postgraduate medical curricula[1]. As emergency care skills are critical for doctors in emergency departments and hospitals, as well as for family practitioners, the quality of assessment is essential. It is vital that certification of such skills is based on robust testing, giving the certificate credibility[4]. Considering the widespread use of emergency care courses and assessment, the implications for patient safety and the high costs involved, it is surprising how little research is available on the quality of these assessments; in particular, their reliability and validity. A small number of studies have been done on advanced life support courses, sometimes showing poor inter-rater agreement[5] and sometimes showing moderate to good inter-rater agreement with small selected samples of participants[3]. Most studies recommend further research, but few provide specific recommendations. In the current study, we conduct psychometric analyses on a checklist, a 9-item competency scale and a global performance scale, used in a certified emergency care course. These assessment formats are commonly used in emergency care and other settings[6,7].

Purpose of this study

The purpose of this study was to assess the validity and reliability of commonly used formats (checklists, competency and global performance scales) in the assessment of emergency care skills for residents, and to make specific recommendations for further validation of these instruments.

METHODS

Participants

Since 2009, emergency care courses are mandatory for emergency staff in all Dutch hospitals. Medical graduates are trained in standard emergency care methods. All Dutch family-practice residents are required to do a 6 months traineeship in an emergency department. Prior to the start all residents must complete a 2 week general emergency care preparatory course according to specific guidelines[8]. After passing this course they are allowed to start their traineeship under supervision of certified attending physicians. This course includes emergency care subjects such as the ABCDE-approach to emergency resuscitation. Each year 500-600 family-practice residents are trained and assessed, using a scenario assessment and knowledge test. The scenario assessment takes 15 minutes and is performed with one scenario, a standardized (trained) patient and one trained rater. This is a high stakes assessment for the residents: if they fail, one resit is offered with another scenario and two raters (including the course director). If a resident fails again, he/she is not allowed to start the emergency department traineeship.

All trainees in the present study were 2nd year residents in the December 2012 or March 2013 course (n=179). They were asked to consent to participate in the study and to being video-taped during their assessment. Fifteen raters with different medical backgrounds assessed the participants (one rater per candidate).

Materials

The assessment instrument under evaluation aims to measure the skills and competencies of a doctor in an emergency situation: the ability to perform primary assessment of a seriously ill or injured patient, to determine the priority of the necessary actions, and to start treatment of all immediately life-threatening conditions[8]. The instrument contains three parts (formats). A case specific *checklist* with 8-11 critical decisions ('recognizes airway obstruction, 'supplies oxygen (NRM 12-15L/min.)', scored yes/no for 'correct' and 'timely'. Secondly, a *Competency Scale* (9 items, 6 on the ABCDE method ('uses ABCDE approach on initial assessment') and diagnostics ('requests additional diagnostic studies'), 3 on communication ('communicates with patient effectively'), rated on a 7-point scale, 1=very weak, 7=excellent). Thirdly, a *Global Performance Scale* using a single 10-point scale to rate 'independent functioning in caring for acutely ill patients in the Emergency Department' (10=perfect). Appendix 1 provides an example. The pass/fail cut point is based on the Global Performance Scale (fail is <6). Cases were developed by experts and covered varying emergency care situations (case 1 Diabetic Ketoacidosis, case 2 Cerebrovascular accident, case 3 Hemorrhagic shock due to trauma, case 4 Burn victim).

The knowledge test contained 100 Multiple Choice (MC) items; the pass mark was 60%.

Inter-rater reliability was assessed using videotaped encounters, as has been done satisfactorily in previous investigations[5]. Using stratified random sampling, we selected 22 videos from the available 40 videos on two cases (1 and 4) for video assessments.

The participants were stratified in 3 groups for both cases, based on their global performance: <7, between 7-8, > 8. The amount of videos selected from the stratified groups was proportional to that in the group of participants (n=144).

Eight raters from the regular assessor group viewed the 22 video registrations and assessed the candidates independently at a self-chosen time. All live and video based raters had different medical backgrounds (emergency medicine physician, surgeon, internist-intensivist or anaesthesiologist), as is typical in emergency medicine in clinical settings and for most emergency courses. They were all qualified instructors in internationally certified emergency medicine courses and trained according to international standards. The design and assessment of this type of emergency care training is comparable with ATLS (Advanced Trauma Life Support) and ALS (Advanced Life Support) courses.

Measurements

Reliability of a test indicates the extent to which test results are reproducible and consistent under different conditions[9–11]. For validation, evidence was collected on different dimensions of construct validity: content, response process, internal structure (including reliability across items, stations, raters), relations with other variables and consequences[12]. We have measured: *content validity* (expert panel review of the assessment instrument), *response process* (debriefing sessions with raters), *internal structure* (factor analysis on the competency scale, analysis of internal consistency, calculation of agreement between 'live' and 'videobased' raters and a generalizability study to explore sources of error) and *relations with other variables* (associations between assessment formats).

Statistical analysis

We used Cronbach's alpha to calculate internal consistency of scales. Inter-rater agreement was computed using the Intra Class Correlation (ICC)[11],[13]. We used the two-way random effects model of the ICC, as raters and candidates are random factors; 'absolute agreement' because this is a criterion-referenced test and 'single measures' because the assessment is normally done by a single rater. For the generalizability study, we used an absolute G-coefficient as both raters and cases can differ between candidates[14]. In order to analyze the reliability of the checklist, with a different number of critical decisions in both cases (8 and 7 items respectively for case 1 and 4), we used the mean summed score of both cases. Each candidate is assessed with one case, therefore case error could not be analyzed. The association between formats was calculated using Pearson's correlation coefficient.

The videotaped and non-videotaped groups were compared on main characteristics and results using the Student's t-test (independent and paired); one-way ANOVA was used to compare means of cases. We used SPSS version 22. The generalizability study was done using G String IV software.

Ethics

All participants signed a consent form; the study was approved by the Dutch ethical board for research in (medical) education (NVMO, the Netherlands Association for Medical Education, approval no 210). The NVMO is an independent association that carries out activities for anyone involved in medical and health care education in the Netherlands and Flanders (Belgium).

RESULTS

Characteristics of study subjects

144 residents, 80% of 179 course trainees, agreed to participate in this study. Of this group, 86 agreed to have their assessment videotaped (60%). Participants' average age was 29.3 years, 82% women, with an average of 5.6 months of experience with acute patients and an average score of 48% on the national knowledge test for family practitioners. The 22 residents participating in the videotaped assessments were on average 30 years, 68% women, had 4.6 months of experience with acute patients and a score of 47% on the national test. Differences between the sample and population were not statistically significant.

Assessment results of the total group

The critical decisions on the checklist were scored as correct and timely for 80% of the residents (Table 1); means between cases were significantly different for the checklist (F (3, 80), p<.0001). Scores on the Competency Scale and the Global Performance Scale did not differ significantly between cases (p=0.25/0.19). Thus, only the checklist showed variation between cases. Eight percent (n=12) of the residents failed the assessment. On average 66% of the knowledge test items were answered correctly (SD=9.78).

Assessment results of the video based assessments

The 22 residents were individually (live) assessed by four different raters and all re-assessed by eight raters. Comparing live and video based assessment showed that the eight video based raters were more stringent than the live raters for all three parts of the assessment instrument (Table 2).

Assessment format	Mean (SD)	Cases 1	2	3	4
n	**144**	44	26	36	38
Checklist critical decisions; Mean percentage correct and timely (7-11 items)	**80%**	92%	93%	40%	96%
(SD)	(29%)	(10%)	(9%)	(31%)	(13%)
Competency scale, 9 items, 7 pt (1=poor, 7=excellent) Mean	**5.61**	5.6	5.37	5.70	5.62
(SD)	(.70)	.56	.55	1.08	.50
Global Performance scale 1-10, 10=positive end; Mean	**7.35**	7.44	6.87	7.44	7.47
(SD)	(1.24)	1.176	1.25	1.54	.93

Table 1: Overview of the assessment results of the skill assessment instrument (all cases)

Assessment format	Live rating	Video based rating	P-value
Checklist; Mean percentage correct and timely (SD)	0.93 (0.12)	0.82 (0.11)	< 0.001
Competency scale, Mean 9 items (SD)	5.50 (0.52)	5.30 (0.55)	0.012
Global Performance scale, Mean (SD)	7.3 (1.17)	6.9 (1.13)	0.002

Table 2: Comparison of video-based and live assessment results of 22 residents (case 1 and 4)

Content and response process evidence

The assessment instrument was evaluated by medical and educational experts and found to be representative of the emergency care skill to be assessed. The assessment instrument was implemented from Fall 2012 by the family practice training centre in the Netherlands. At the beginning of each assessment day, raters were briefed on the different scenarios and assessment instrument. During the assessment, a course director was available for questions. At the end of each assessment day a debriefing session was organized to evaluate the assessment process. Standardized patients were trained regularly.

Internal structure

We found an indication of good coherence for the 9-item competency scale, as the factor analysis showed that 65% of the total variance in the items was explained by one factor. Inspection of the items loading on the first factor (1-6) indicated this factor could be summarized as 'clinical competency'. Further analyses indicated a second factor, explaining another 9% of the variance. The second factor (item 7-9) could be summarized as 'communication competency'. The two factors together explained 74% of the total variance (Factor analysis results are provided as Appendix S2). The internal consistency (Cronbach's alpha), for the 9-item competency scale was 0.93, for the clinical skill subscale (item 1-6) 0.91, for the communication subscale (item 7-9) 0.86. The alpha for the knowledge test was 0.74. All scales had good internal consistency.

Inter-rater reliability

The inter-rater agreement, indicated by the Intra Class Correlation (ICC), was poor for the dimensions 'correct' and 'timely' of the checklist (Table 3). The ICC on the clinical competency scale was moderate; it was poor for the communication competency scale. The ICC for the global Performance scale was moderate.

Assessment format	ICC score
Checklist (critical decisions)	
Correct (mean sum over 2 scenarios)	0.30
Timely (mean sum over 2 scenarios)	0.39
Clinical competency scale (item 1-6)	0.49
Communication competency scale (item 7-9)	0.27
Global Performance scale	0.50

Table 3: IntraClass Correlation scores for the Checklist, Competency Scales and Global Performance Scale. Two way random effects model, absolute agreement, single measures, 95% confidence interval. ICC values 0.21-0.40 are considered ' poor/fair', 0.40-0.60 'moderate', 0.60-0.80 'substantial', 0.80-1.00 'almost perfect'

Generalizability analysis showed that using the checklist, for the dimension 'correct' 13 raters would be required to achieve a reliability of 0.80 and for the dimension 'timely' 5 raters would be needed. For both the clinical competency scale and the global performance scale, 4 raters would be required to achieve an acceptable reliability (Figure 1). Using the communication competency scale 13 raters would be needed.

To summarize, the 9-item competency scale showed good overall coherence and consisted of two factors (constructs): a clinical and communication subscale. Internal

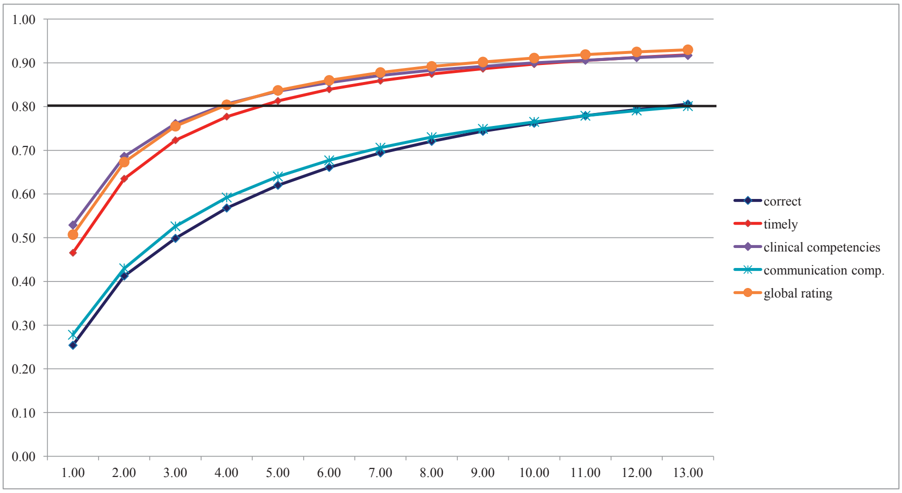


Figure 1: Absolute G-coefficients as a function of the number of raters per assessment format

consistency of (sub)scales was high. The inter-rater agreement was moderate for the clinical competency and the global performance scale and poor for the communication scale and for the checklist (raters agree more on timeliness than on correctness of critical decisions). For a reliable assessment, 4 raters are needed using the clinical competency or global performance scale, 5-13 for the two dimensions of the checklist and 13 using the communication competency scale.

Relation between assessment formats

The global performance scale was strongly correlated with the 9-item competency scale (0.85) and was less strongly, but still significantly correlated with the checklist (Table 4). The score on the knowledge test was correlated with the global performance scale and 9-item competency scale, but not with the checklist score.

		Competency Scale	**Global Performance Scale**	**Knowledge test score**
Checklist score	Correlation	.13	.19*	.14
	p	.12	.02	.10
	N	144	144	142
Competency Scale	Correlation		.85**	.24**
	p		.00	.004
	N		144	142
Global Performance Scale	Correlation			.27**
	p			.001
	N			142

Table 4: Pearson's Correlation coefficient between the 3 assessment formats and knowledge test (Correlation is significant at the 0.05 level* or the 0.01 level **, 2-tailed).

DISCUSSION

To the best of our knowledge, this is the first time a psychometric study has been done on three commonly used formats in emergency skills assessment, as part of a certified emergency care course.

It is clear that the weakest link in the assessment is the checklist, which has poor reliability. This may seem counter-intuitive, as case-specific, task oriented item lists were thought to minimize inter-rater variability, but various studies have shown this is not the case[15]. Our results are consistent with research in other domains spanning many years which has shown that checklists are not as objective as originally supposed[16,17], have inferior validity, and do not show better reliability than global rating scales[18–21]. The dimension 'timely' in the checklist showed a higher reliability than the dimension 'correct'. As timeliness is at the core of the emergency skill being assessed (the ABCDE approach is about the right sequence of actions), a higher level of agreement among raters on this dimension might be expected compared to agreement on the correctness of actions (e.g. whether the right tools were used). It is good to note that checklists are usually related to the correctness of clinical actions. Discussion with raters from our study on the poor inter-rater agreement with checklists made clear that they differed in their professional opinions on the relevance of the checklist items, probably a consequence of their varying medical background. The detailed character of checklists makes them more vulnerable to these differences than rating scales. Although we cannot completely rule out the possibility that our checklist was insufficiently validated, our findings are consistent with previous research and provide a legitimate base for our conclusion that checklists, in general, have insufficient reliability and validity and should be replaced by more global rating scales.

The use of checklists has grown particularly since the widespread implementation of Objective Structured Clinical Examinations (OSCEs), where clinical performance is scored at different stations with standardized problems[22,23]. Checklists however tend to capture only whether some action occurred or not, whereas rating scales require the interpretation of actions[24]. In addition, checklists often do not capture increasing levels of expertise whereas rating scales do[25,26]. An explanation might be that checklists reward thoroughness and mimic a novice's approach to problem solving, whereas expert clinicians are not necessarily thorough but highly accurate in problem solving[16,18]. The intuitive attractiveness of checklists was confirmed in a study where clinicians found checklists more appropriate than competency scales to score clinical performance, but consistency in pass/fail decisions was equally poor for both forms[7]. Checklists with long lists of critical items to be scored often are labour-intensive and we found they are sensitive to case-differences, but do not add any reliability or validity to the assessment.

The competency scale and global performance scale showed high correlations; they seem to be measuring largely the same underlying dimensions. Advantages of the clinical competency scale are its potential for providing specific feedback to participants, which is useful for remediation. The communication scale showed poor reliability. Apparently raters agree more on clinical than on communication skills, possibly because they are less trained in assessing communication. What would be the added value of the global performance scale? An advantage of this scale is the more holistic assessment of the candidate's performance, where competency scales are more focused on specific skills. Multi-case research will have to show which scales have higher inter-case reliability.
Video based raters appeared to be more stringent than live raters on all three parts of the assessment instrument. Raters assessing videos are more limited: they do not have the

opportunity to ask supplementary questions or to alter their position; they might miss subtle checks by the candidate. On the other hand, they could feel more psychological distance to the candidate and hence look more objectively. We cannot conclude from this study which assessment is more valid, but in general delayed videos assessment of emergency skills is found as reliable as live assessment[27].

Limitations

This study has some limitations. The assessment method studied was part of a specific emergency care course in The Netherlands. However, the course design and assessment method used are quite comparable to a number of internationally certified resuscitation courses, such as ATLS and ALS, and thus are recognizable for instructors from other courses. Another limitation of this study is the use of a single patient scenario per candidate. Extensive evidence in assessment research in the last decades shows content specificity is the main cause of unreliability and as such outweighs all other sources of bias(24,25). Various studies have shown that inter-rater reliability is a relatively small source of error compared with inter-case variability (which rarely exceeds a correlation of 0.1-0.3 across cases[24,29,30]). Thus, the reliability coefficients reported in this study may well be an upper limit of actual reliability when candidates see different cases. However, we have analyzed an internationally representative emergency care assessment situation where single patient scenario assessment is commonly used.

This study yields a number of practical implications for those involved in emergency care assessment. A first implication is that assessment may be improved by putting more resources into multiple raters and multiple cases using the competency scales. While it is clear that additional raters will increase reliability, the data suggest that many raters will be needed to achieve acceptable standards. In fact, raters would be better deployed doing independent assessments with different scenarios, as performance of a candidate in one case is a poor predictor on another case[9]. One study showed that with 6 short (5 min.) cases and 1 rater per case a substantial reliability in emergency assessment can be achieved[28]. A challenge in strengthening reliability is to maintain validity, since complex skills require an integrated judgement and sufficient time to assess them[16]. Another way to increase the number of cases is to combine formative and summative assessment. During training, opportunities for observation and feedback can be used to collect more information on the candidate's performance with different scenarios. The formative value of assessment has been given increasing attention lately, characterized as a shift from assessment *of* learning to assessment *for* learning[24].

A second way to improve assessment would be to supplement performance testing with 'know how' test formats, using the strong relationship between 'know how' (applied knowledge) and 'show how' (demonstrate skills) for a variety of clinical procedural skills[31]. The 'key feature' test, a written test of clinical decision-making, may be useful. It requires short uncued answers, focusing on the critical steps ('key features') in a range of scenarios[16,32]. 'Key feature' tests have modest correlation with clinical knowledge and competency tests[33]. Several studies have shown that these types of tests have good validity in predicting clinical performance in practice[34–36].

Although raters are typically experienced and trained, the low inter-rater agreement, also confirmed in other studies in general clinical skill assessment(6), shows there is room for improvement. Training has shown to have modest or no gains with experienced raters[37]; alignment sessions in which individual rating of videotaped performance is discussed may be more useful. Pass/fail decisions could be based on composite scores, using data from multiple tests and occasions[24].

Conclusions

This study analyzed the validity of commonly used formats in the assessment of emergency skills, using checklists, competency scales and global performance scales in a single scenario assessment. One conclusion from this study is that checklists have poor validity and reliability in assessing emergency skills. As they are still standard practice in most assessments of emergency care courses, this is an important conclusion. Instead, the study suggests better measurement will result from using a clinical competency or a global rating scale. Nevertheless, we observed a limited overall agreement between raters in the assessment of emergency skills.

A concern is that the limited number of scenarios and raters usually applied in emergency care assessment, results in low reliability and validity. While additional raters may help, research should be conducted to examine the use of multiple cases in this assessment. Improving validity of assessment is vital, considering the worldwide use of emergency care training, the implications for patients, and the high costs of training. More research into improving valid and feasible assessment of emergency skills is strongly recommended, using several (short) scenarios and raters.

Acknowledgement

The authors thank W.Henny MD, E.H.Jansen MD, M.G.Luiting MD, I.M.Maissan MD P.Pahlplatz MD, M.Peters MD, L.Reenalda MD and D.J.Versluis PhD MD MHA for their video-assessments.

REFERENCES

1. Perkins GD, Kimani PK, Bullock I, Clutton-brock T, Davies RP, Gale M, Lam J. Improving the Efficiency of Advanced Life Support Training. Ann Intern Med. 2012;157(1).
2. Preston JL, Currey J, Eastwood GM. Assessing advanced life support (ALS) competence: Victorian practices. Aust Crit Care. 2009 Nov;22(4):164–71.
3. Ringsted C, Lippert F, Hesselfeldt R, Rasmussen MB, Mogensen SS, Frost T, Jensen ML, Jensen MK, van der Vleuten C. Assessment of Advanced Life Support competence when combining different test methods - Reliability and validity. Resuscitation. 2007;75:153–60.
4. Napier F, Davies RP, Baldock C, Stevens H, Lockey AS, Bullock I, Perkins GD. Validation for a scoring system of the ALS cardiac arrest simulation test (CASTest). Resuscitation. 2009 Sep;80(9):1034–8.
5. Perkins GD, Hulme J, Tweed MJ. Variability in the assessment of advanced life support skills. Resuscitation. 2001 Sep;50(3):281–6.
6. Flowerdew L, Brown R, Vincent C, Woloshynowych M. Development and validation of a tool to assess emergency physicians' nontechnical skills. Ann Emerg Med. Elsevier Inc.; 2012 May;59(5):376–85.e4.
7. Ringsted C, Østergaard D, Ravn L, Pedersen J a, Berlac P a, van der Vleuten CPM. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. Med Teach. 2003 Nov;25(6):654–8.
8. Breedveld. Rapportage werkgroep Kwaliteitsindeling SEH Spoedeisende hulp: Vanuit een stevige basis. 2009;
9. Schuwirth LWT, Van der Vleuten CP. General overview of the theories used in assessment : AMEE Guide No. 57. Med Teach. 2011;33(57):783–97.
10. Downing SM. Reliability : on the reproducibility of assessment data. Med Educ. 2004;38:1006–12.
11. Streiner D, Norman G. Health measurement scales. Oxford University Press; 2008.
12. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract. 2013 May 2.
13. McGraw K, Wong S. Forming inferences about some intraclass correlations coefficients. Psychol Methods. 1996;1(4):390–390.
14. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach. 2012 Jan;34(11):960–92.
15. Schuwirth LWT, van der Vleuten CPM. The use of clinical simulations in assessment. Med Educ. 2003 Nov;37 Suppl 1:65–71.
16. Wass V, Vleuten C Van Der, Shatzer J, Jones R. Medical education quartet. Assessment of clinical competence. 2001;357:945–9.
17. Reznick R, Regehr G, Yee G, Rothmann A, Blackmore D, Dauphinee D. Process-rating forms versus task-specific checklists in an OSCE for medical licensure. Acad Med. 1998;73(10):S97–9.
18. Regehr G, MacRae H, Reznick RK SD. Comparing the psychometric properties of checklists and global rating scales for assessing performance of an OSCE-format examination. Acad Med. 1998;73(9):993–8.

19. Cohen DS, Colliver JA, Marcy MS FE, Swartz M. Standard Setting; Psychometric proporties of a standardized-patient checklist and rating-scale form used to access interpersonal and communication skills. Acad Med. 1996;71(1):S87–9.
20. Cohen DS, Colliver JA, Robbs RS, Swartz MH. A large-scale study of the reliabilities of checklist scores and ratings of interpersonal and communication skills evaluated on a standardized-patient examination. Adv Health Sci Educ Theory Pract. 1996 Jan;1(3):209–13.
21. Govaerts MJB, van der Vleuten CPM, Schuwirth LWT. Optimising the reproducibility of a performance-based assessment test in midwifery education. Adv Health Sci Educ Theory Pract. 2002 Jan;7(2):133–45.
22. Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. Med Teach. 2010 Jan;32(10):802–11.
23. Patrício MF, Julião M, Fareleira F, Carneiro AV. Is the OSCE a feasible tool to assess competencies in undergraduate medical education? Med Teach. 2013 Jun;35(6):503–14.
24. Swanson DB, van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art revisited. Teach Learn Med. 2013 Jan;25 Suppl 1(December):S17–25.
25. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med. 1999 Oct;74(10):1129–34.
26. Hodges B, Regehr G, Hanson M MN. Validation of an objective structured clinical examination in psychiatry. Acad Med. 1998;73(8):910–2.
27. House JB, Dooley-Hash S, Kowalenko T, Sikavitsas A, Seeyave DM, Younger JG, Hamstra SJ, Nypaver MM. Prospective comparison of live evaluation and video review in the evaluation of operator performance in a pediatric emergency airway simulation. J Grad Med Educ. 2012 Sep;4(3):312–6.
28. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. Anesthesiology. 2003 Dec;99(6):1270–80.
29. Norman G. Research In Medical Education: Three decades of progress. BMJ. 2002;324(7353): 1560–2.
30. Elstein, AS, Shulman, LS, Sprafka S. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge MA: Harvard University Press; 1978.
31. Van Dalen J, Kerkhofs E, Verwijnen G, van Knippenberg-van den Berg B, van den Hout H, Scherpbier A, van der Vleuten C. Predicting communication skills with a paper-and-pencil test. Med Educ. 2002 Feb;36(2):148–53.
32. Farmer EA, Hinchy J. Assessing general practice clinical decision making skills. Aust Fam physician. 2005;34(12):1059–61.
33. Hatala R, Norman GR. Adapting the Key Features Examination for a clinical clerkship. Med Educ. 2002 Feb;36(2):160–5.
34. Tamblyn R, Abrahamowicz M, Wenghofer E, Blackmore D, Winslade N. Physician Scores on a National Clinical Skills Examination as Predictors of Complaints. JAMA. 2013;298(9):993–1001.
35. Wenghofer E, Klass D, Abrahamowicz M, Dauphinee D, Jacques A, Smee S, Blackmore D, Winslade N, Reidel K, Bartman I, Tamblyn R. Doctor scores on national qualifying examinations predict quality of care in future practice. Med Educ. 2009 Dec;43(12):1166–73.

36. Ram P, van der Vleuten C, Rethans JJ, Schouten B, Hobma S, Grol R. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. Med Educ. 1999 Mar;33(3):197–203.

37. Van Der Vleuten, CPM, Van Luyk, SJ, Van Ballegooijen, AMJ, Swanson D. Training and experience of examiners. Med Educ. 1989;23(3):290–6.

APPENDIX 1: ASSESSMENT INSTRUMENT EMERGENCY CARE SKILL (FOR SCENARIO 1: DIABETIC KETOACIDOSIS)

Name candidate:....................................... Number: ....

Date: .... Rater:...

Checklist (items on critical decisions)		Executed	
		correct	timely
A	Recognizes airway obstruction (snoring breathing)		
	Treats airway obstruction (manual maneuver)		
B	Measures breathing frequency		
	Supplies oxygen (NRM 12–15 L/min.)		
C	Recognizes shock (pulse 104/min and BP 90/60 mmHg)		
	Starts intravenous fluid bolus (Normal saline 0.9% of Lactated Ringer's)		
D	Measures glucose		
	Assesses state of consciousness using GCS parameters		
E			

	Competence Scale	1 very weak	2 weak	3 insufficient	4 questionable	5 sufficient	6 good	7 excellent
	ABCDE approach							
1	Uses ABCDE approach on initial assessment							
2	Uses ABCDE approach on initial treatment							
3	Uses re-assessment properly							
	Additional actions							
4	Requests additional diagnostic studies							
5	Proposes a working diagnoses							
6	Consults specialist when needed							
	Communication							
7	Communicates with patient effectively							
8	Gives clear instructions to nurse							
9	Radiates a calm and confident attitude							

Global Performance Scale
Assessment of independent function in caring for acutely ill patients in the Emergency Department:... (1–10)

Passed/Failed

APPENDIX 2. PRINCIPAL COMPONENT ANALYSES OF THE COMPETENCY SCALE

Communalities

Competency Scale	Initial	Extraction
1 uses ABCDE approach on initial assessment	1.000	.637
2 uses ABCDE approach on initial treatment	1.000	.768
3 uses re-assessment properly	1.000	.790
4 requests additional diagnostics	1.000	.711
5 proposes a working diagnoses	1.000	.610
6 consults specialist when needed	1.000	.702
7 communicates with patient effectively	1.000	.869
8 gives clear requests to nurse	1.000	.814
9 radiates a calm and secure attitude	1.000	.749

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.845	64.948	64.948	5.845	64.948	64.948	3.875	43.061	43.061
2	.805	8.943	73.891	.805	8.943	73.891	2.775	30.830	73.891
3	.591	6.572	80.463						
4	.437	4.850	85.313						
5	.371	4.119	89.432						
6	.290	3.220	92.653						
7	.245	2.723	95.376						
8	.226	2.516	97.891						
9	.190	2.109	100.000						

Extraction Method: Principal Component Analysis.

3

Component Matrix[a]

Competency Scale	Component	
	1	2
1 uses ABCDE approach on initial assessment	.763	-.237
2 uses ABCDE approach on initial treatment	.824	-.297
3 uses re-assessment properly	.878	-.137
4 requests additional diagnostics	.830	-.151
5 proposes a working diagnoses	.778	-.062
6 consults specialist when needed	.782	-.300
7 communicates with patient effectively	.750	.554
8 gives clear requests to nurse	.824	.368
9 radiates a calm and secure attitude	.816	.289

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

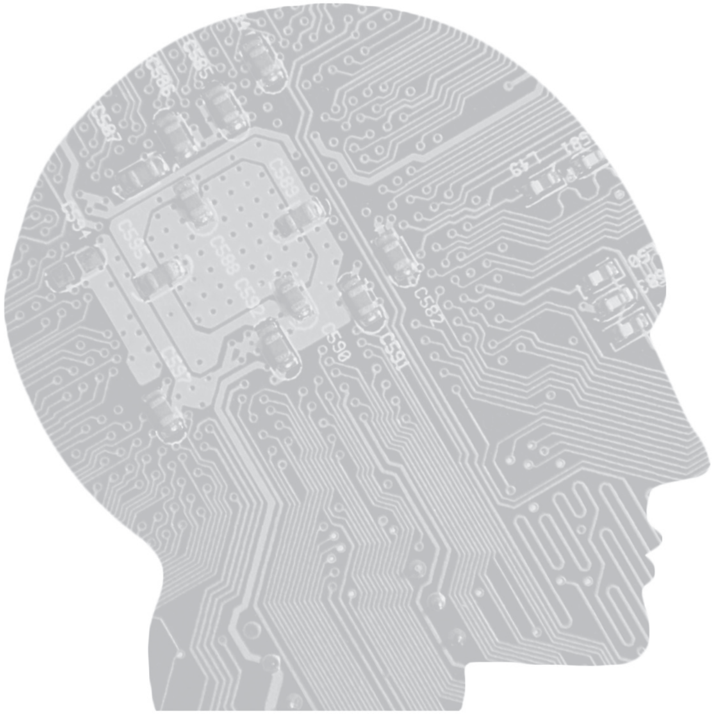
Rotated Component Matrix[a]

	Component	
	1	2
1 uses ABCDE approach on initial assessment	.743	.292
2 uses ABCDE approach on initial treatment	.829	.284
3 uses re-assessment properly	.771	.442
4 requests additional diagnostics	.742	.401
5 proposes a working diagnoses	.647	.438
6 consults specialist when needed	.798	.255
7 communicates with patient effectively	.239	.901
8 gives clear requests to nurse	.413	.802
9 radiates a calm and secure attitude	.456	.735

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 3 iterations.

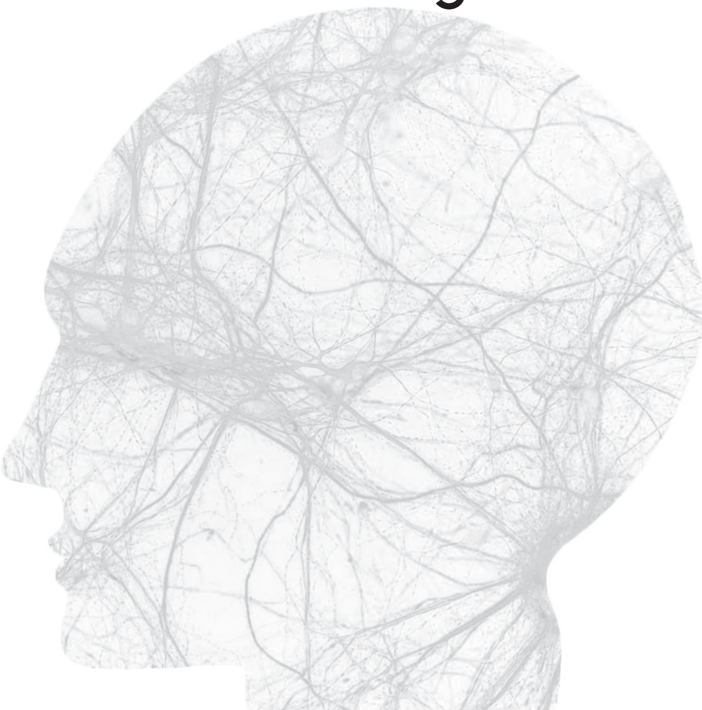
Component Transformation Matrix

Component	1	2
1	.781	.625
2	-.625	.781



Chapter 4

Preparing residents effectively for emergency skills training with a serious game



Mary EW Dankbaar
Maartje Bakhuys Roozeboom
Esther APB Oprins
Frans Rutten
Jan LCM van Saase
Jeroen JG van Merrienboer
Stephanie CE Schuit

Submitted for publication

ABSTRACT

Introduction

Training emergency care skills is critical for patient safety but cost-intensive. Serious games have been propagated as an engaging self-directed learning tool for complex skills. The objective of this study was to compare the cognitive skills and motivation of residents who only used a course manual as a preparation for classroom training on emergency care with residents who used an additional serious game.

Methods

In a quasi-experiment with residents preparing for a rotation in the emergency department, the 'reading' group (n=52) received a course manual before classroom training; the 'reading & game' group (n=107) received this manual plus the game as preparation for the same training. Emergency skills were assessed before training (with residents who agreed to participate in pre-training assessment, n=18 resp. n=24), using validated competency scales and a global performance scale. We also measured motivation.

Results

All groups were comparable on main characteristics. Before training, the reading & game group felt engaged with the game and spent more time on task (+2.5 h) than the reading group. Game playing residents showed higher scores on objectively measured and self-assessed clinical competencies, but equal scores on the global performance scale and were equally motivated for training, compared to the reading group. After the 2-week training, no differences between groups existed.

Conclusions

After preparing training with an additional serious game, residents showed improved clinical competencies, compared to residents who only studied course material. After 2-week training, this advantage disappeared. Future research should study the retention of game effect in blended designs.

INTRODUCTION

In the last decade, serious games have been propagated for education and training, combining the engaging aspects of video games with the serious goals of education. Serious games are educational formats requiring learners to participate in a competitive activity with preset rules and are built to help learners achieve specific learning objectives in a motivational environment[1,2]. An important category of serious games in education and health care are simulation games; they offer realistic, engaging learning environments where students can train decision making processes[3]. The role of motivation in learning has been much neglected; if students are intrinsically motivated to learn something, they are expected to spend more time on learning and feel more positive about what they learn[4]. Several studies have been performed regarding the effects of serious games on learning outcomes and motivation; most reviews however show mixed and ambiguous results, largely due to methodological flaws in the reviewed studies[1,3,5–7].

Training emergency care skills is critical for patient safety and an essential part of undergraduate and postgraduate medical education. Each year, worldwide more than 1.5 million health care professionals attend a variety of emergency care courses, using the internationally standardized “ABCDE” approach[8]. This method prioritizes initial resuscitation of critically ill patients. The mnemonic stands for Airway, Breathing, Circulation, Disability and Exposure. Substantial resources are involved in training these complex cognitive skills[9]. Complex skills integrate knowledge, several skills and a professional attitude, to be performed within a constrained time interval[10]. Considering the training challenges and tight budgets in health care organizations, more efficient training models are required. In 2010, the International Liaison Committee of Resuscitation recommended to combine online self-directed training with instructor lead skills training[11].

As a preparation for instructor-led emergency skills training, we have developed a serious game (abcde*SIM*), in which residents can stabilize patients in a virtual emergency department. Should games prove to be effective as a preparation for training complex acute care skills, then training time may be reduced. Once simulation games have been developed, they can be used for skills training for large numbers of trainees, with no extra costs for instructors or simulation patients, in contrast to simulation centers. Blended learning, combining e-learning with classroom training, has been shown to make emergency care training more efficient for different learning goals, while maintaining learning outcomes[8,12].

The research questions of this study are: Do residents, who use the abcde*SIM* game in addition to a course manual as a preparation for classroom training, show better emergency care skills than residents who exclusively use a course manual, and are they motivated to play the game? Are game-playing residents more motivated towards the course content than residents who only used the course material? Is there a difference between groups in skills level after two weeks of training?

METHODS

Participants, setting and study design

We performed this study with 2nd year family practice residents. All family-practice residents in the Netherlands are required to do a 6-month traineeship in an emergency department; prior to the start residents must complete a 2-week general emergency care course. After passing this course they are allowed to start their traineeship under supervision of certified attending physicians. This course includes emergency care subjects based on the ABCDE-approach to emergency resuscitation. Each year 500-600 Dutch family-practice residents are trained and assessed in this 2-week certified course. They are assigned in a timeline in training groups of 70-90 residents by the academic training organization; there are no systematic differences between groups.
Most residents have limited experience with emergency care. As the residents of a particular training group usually know each other (they are following the same courses), random assignment to the reading & game and reading group was not feasible; game accounts could easily be exchanged between groups. In addition, different resident groups are quite homogeneous in age and experience. Therefore we used a *quasi-experimental design with a historical control*: residents from the December training group were treated as the control (reading) group and received the course manual 6 weeks before the 2 week classroom training. Residents from March and September groups were treated as the intervention (reading & game) group and in addition received an account for the *abcdeSIM* game 6 weeks before training. They were advised to first study the course material and then play the game. The game group completed an online evaluation questionnaire after playing the game. Emergency care skills were assessed prior to the training and on the last training day. Participants completed a pre-training questionnaire during the first training day and their data on post-training assessment were recorded.

Selection of participants

Six weeks before the training, residents from the groups involved were asked to participate in the pre-training assessment 2 hours before the training; this assessment was not part of the regular program and had no consequences for the participants. On the first training-day, all residents were additionally asked to participate in the rest of the study (questionnaires, post training assessment). Participants signed a consent form for both study parts; the study was approved by the Dutch ethical board for research in medical education (NVMO, no 210).

Materials

Preparatory course manual

All residents received a course manual on emergency care skills as a preparation for classroom training. The manual contained *instructional* material (no cases) on the ABCDE-approach and the essentials of medical emergency and trauma care. Subjects that were covered in the manual were impaired vital functions, disturbances of consciousness, etc (Table 1).

	Learning goals	Content
Course material	Develop knowledge on ABCDE-approach and essentials of emergency and trauma care	Theory on emergency care approach and interventions, such as: impaired vital functions, breathing problems, chest pain, disturbances of consciousness, patient in shock, pediatric patient, acute abdomen, injuries, medication.
Serious game	Develop cognitive emergency care skills	Six high-fidelity cases with feedback on common emergency care syndromes: exacerbation COPD, pneumosepsis, subarachnoid hemorrhage, acute coronary syndrome, hemorrhagic shock due to gastrointestinal bleeding, deep venous thrombosis.

Table 1: Overview of the preparation of classroom training

Preparatory serious game

The *abcdeSIM simulation* game was developed by Erasmus University Medical Center Rotterdam and Schola Medica and provides an online simulated emergency department, where residents can apply and exercise their emergency care skills with virtual patients. All relevant tools for assessment and stabilization of acutely ill patients are virtually available (stethoscope, lab tests, infusion fluids, blood, medication) and a high-fidelity mathematical model of human physiology for respiration, circulation and neurological functioning was implemented on the virtual patients to create realism and give feedback on the patients' condition. The game contained six regular adult patient cases (cases in Table 1).

The game was primarily aimed at training clinical emergency care skills; communication skills were not addressed in the game. Each case took 15 minutes (a timer was presented) and could be done as often as desired. During playing, players received direct feedback on their actions through a monitor with data on the patient's condition and from the assisting virtual nurse. After playing a case residents got a score and narrative feedback.

Figure 1: Screenshots of the abcde*SIM* game

The game score was dependent on how many right decisions were performed according to the ABCDE method and how efficiently this was done (less minutes resulted in more points). There was a high score list to compare scores of players. The simulation game contained only (six) cases, no additional instructions (Screenshots of the game are provided as Figure 1). The reading & game group studied the course material as a preparation for game play.

Classroom training

The 2 week training course was aimed at training emergency care skills, using the ABCDE approach and basic and advanced life support techniques. The design and assessment of this course is comparable with A(T)LS (Advanced (Trauma) Life Support) courses. The training provided a combination of lectures, scenario training in small groups with standardized patients and mannequins and skills training. Each participant acted as the treating physician on three scenarios with a simulation patient, on one scenario with a pediatric mannequin and assisted as a nurse on six other scenarios. In addition s/he observed and critiqued peers on 15 scenarios. Each 20 minute scenario play is followed by a 20 minutes feedback session. To summarize: during training residents played an active role on ten scenarios and reviewed another fifteen scenarios.

Assessment and evaluation instruments

Skill assessment

Post training assessment was part of the regular certified emergency care course. Fifteen assessors with different medical specialty backgrounds, all qualified instructors in internationally certified emergency medicine courses and trained according to international standards (a standardized Generic Instructor Course), assessed the residents;

one rater per candidate. Assessments consisted of a one-case scenario test (15 minutes) with a standardized, trained simulation patient. Due to the design the assessors could not be blinded for the condition; they may have heard about the game as a preparation for training but they were not involved in the study, nor had any interest in a specific outcome. Before assessments, raters were briefed on the scenarios and instrument. During assessments, a course director was available for questions. This is a high stakes assessment for the residents: if they fail, one resit is offered with another scenario and two raters (including the course director). If a resident fails again, s/he is not allowed to start the emergency department traineeship. Participants for the pre and post training assessments were assigned at random to different assessment scenarios. The same scenarios were used for the reading group and reading & game group.

The assessment instrument was aimed to measure the ability to assess and treat seriously ill or injured patients and consisted of a clinical competency scale (6 items on the ABCDE method and diagnostics, e.g. 'uses ABCDE approach on initial assessment'), a communication competency scale (3 items on communication, e.g. 'communicates with patient effectively'), both rated on a 7-point scale (7=excellent). In addition, the assessor judged the candidate on a global performance scale (using a single 10-point scale to rate 'independent functioning in caring for acutely ill patients in the Emergency Department' (10=perfect, assessment instrument is Supplementary File 1). The pass/fail cut point was based on the global performance scale (fail if <6). The assessment instrument was validated in a separate study; the competency scales showed good construct validity and internal consistency. The clinical competency scale and global performance scale showed moderate interrater reliability (ICC=0.49 resp. 0.50); the communication competency scale had poor interrater reliability (ICC=0.27)[9].

Motivation questionnaire on the game

After working on all cases in the game, residents who participated in the reading & game group completed an online questionnaire evaluating the game and the motivation to play. The questionnaire consisted of 9 statements, including items such as: "I felt actively involved with the patient cases", to be scored on a 5-point Likert scale (5=fully agree), (Supplementary file 2).

Pre-training questionnaire on task value, self-efficacy and self-assessment

During the first training day, a questionnaire was completed by all participants on motivation towards the course, self-efficacy and self-assessment. The Motivated Strategies for Learning Questionnaire (MSLQ) has been used extensively in educational research projects to measure students' motivational orientations[13]. We used two subscales: Task value (9 items, e.g. "I think that what I am learning in this course is useful for me to know") and Self-efficacy (9 items, e.g. "I'm certain I can understand the ideas taught in this course". All items were rated on a 7-point Likert scale (7=very true of me). In addition participants completed a self-assessment instrument, based on the raters' assessment instrument.

Self-study time

On the last training day (before assessment), all residents completed an evaluation form on the course, including a question on how much self-study they had spent preparing the course in hours. The questionnaire was anonymous; data on self-study time was available on group-level only.

Statistical analysis

A reliability analysis (Cronbach's alpha) was calculated for the questionnaires and assessment instruments. Independent t-tests were performed to compare group characteristics, assessment data and motivation data from the reading and reading & game group. Unless the distribution of scores is severely skewed, data from rating scales can be analyzed as if they were interval without introducing bias[14]. Effect sizes (ES) were calculated using Glass's delta. Practical significance of research results can be quantified from being small (ES ≈ 0.20) to moderate (ES ≈ 0.50) to large (ES ≈ 0.80)[15]. Associations between game data were calculated using Pearson's coefficient. We treated missing data with pairwise deletion and used SPSS for the statistical analysis.

RESULTS

Characteristics of participants

Out of 210 eligible residents (60 from the reading group, 150 from reading & game group), 159 consented to participate in the study (76%); 52 in the reading and 107 in the game & reading group. A subgroup of 18 (reading) and 24 (reading & game) residents agreed to participate in the pre-training skill assessment. There were no differences between the reading group and reading & game group on main characteristics, such as experience with acute patients and score on national test for family practice residents (Table 2). The pre-training assessment group (n=42) did not differ from the rest of the research participants (n=117, data not presented).

Assessment results acute care skills

Reliability of the scales (Cronbach's alpha) was 0.92 for the 6 items clinical competency scale and 0.81 for the 3 items communication competency scale. Before training (after the game), the reading & game group performed better on clinical competency skills

	Before training assessment			After training assessment		
	Reading group	Reading & game group	p	Reading group	Reading & game group	p
n	18	24		52	107	
Women	13 (72%)	19 (79%)	0.61	42 (81%)	82 (77%)	0.56
Age	29.8	29.5	0.80	29.8	29.2	0.28
Experience acute care (months)	8.2	4.3	0.20	6.0	5.3	0.59
Score on national knowledge test	52.2	49.0	0.33	47.6	45.8	0.13

Table 2: Characteristics of reading group and reading & game group before and after training
**Independent t-test comparing reading and reading & game group; Confidence Interval is 95%.*

(p=0.03, Table 3) with a medium-large effect size (Glass's d=.62) than the reading group. Improvements occurred particularly in the items on initial assessment (d=0.82), treatment (d=0.72) and requests for additional diagnostics (d=0.50). The reading & game group also showed less variability in competency-levels (more homogeneity measured as smaller standard deviation scores; p=0.02). There were no differences in communication competency skills or on global performance scores between groups before training. There was an association between assessment scores on the global performance scale with the clinical competencies scale (r=0.74, p<0.001) and with the communication competencies scale (r=0.42, p=0.006).

	Before training			After training		
	Reading group	Reading & game group	p	Reading group	Reading & game group	p
N	18	24		52	107	
	Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)	
Clinical competencies (7 pt. scale)	3.46 (1.27)	4.25 (0.75)	0.03*	5.6 (0.54)	5.7 (0.91)	0.34
Communication competencies (7 pt. scale)	4.85 (0.76)	4.71 (0.92)	0.59	5.7 (0.56)	5.8 (0.84)	0.41
Global performance (10 pt. scale)	4.92 (1.19)	5.04 (1.08)	0.73	7.4 (0.93)	7.2 (1.36)	0.24

Table 3: Assessment results before and after training
* on two-tailed test

At the end of the 2 week training, scores on the competency and global performance scales were similar for both groups (Table 3). There was also an association between assessment scores on the global performance scale with the clinical competencies scale (r=0.80, p<0.001) and with the communication competencies scale (r=0.60, p<0.001).

Self-study time and game data

Mean self-study time (self-reported) before the course for the reading group was 9.9 hours (SD=5.9 h) and for the reading & game group it was 12.4 hours (SD=5.7 h); the game group spent 2.5 hours extra self-study time (p=0.007). Mean game playing time (logged) for the game group was 2.2 hours per resident (SD=0.95 h), which is consistent with the self-reported study time. The game group spent on average 17 minutes per case. There was an association between playing time and game score (r=0.49, p< 0.001), indicating a learning effect within the game.

Evaluation of the game

Reliability (Cronbach's alpha) of the 9-item game motivation scale was 0.80 (n=90). Mean score on the 5-point scale was 3.9 (SD=0.41). Above average scoring items were "I felt actively involved with the cases" (Mean=4.2,SD=0.53), "My attention was completely drawn to the cases" (M=4.2,SD=0.63), "I liked playing the game" (M=4.2, SD=0.69). Below average scoring item was "I regularly felt stressed during playing the cases" (M=3.3, SD=0.87) (Table with evaluation data is available as Supplementary file 3).

Motivation towards the course, self-efficacy and self-assessment

Reliability of the task value scale was 0.76 (n=159), the self-efficacy scale α=0.83 (n=159) and the self-assessment scales on clinical and communication competencies α=0.88 respectively α=0.71 (n=150). On comparison, both groups did not show different scores on task value (motivation for the course, M=6.2/6.2; SD=0.43/0.39 for the reading and reading & game group resp. on the 7-point scale) and self-efficacy before training (M=4.8/4.8; SD=0.66/0.61 for the reading and reading & game group resp.). The self-assessment scores on clinical competencies before training were higher for the reading & game group (n=104, M=4.7, SD=0.54) than the reading group (n=46, M=4.4, SD=0.74, p=0.01). There were no differences in self-assessed communication competences or global performance before training.

DISCUSSION

This study to compared the cognitive skills and motivation of residents who only used a course manual as a preparation for classroom training on emergency care (the 'reading group') with residents who used an additional serious game (the 'reading & game group'). The results showed a positive effect of an additional preparatory serious game on the clinical emergency care competencies of residents, compared to a group who exclusively read a course manual. The game had no effect on communication competencies or on global performance in caring for acutely ill patients.

With limited extra self-study time (2.5 hours, compared to 10 hours spent on the course manual by both groups), the serious game had a positive effect on clinical competencies, with a moderate-large effect size, and reduced the variability in clinical performance. For teachers, a more homogeneous student group is easier to train. The same communication competency levels were measured in both groups, which can be explained by the fact that the game did not address these competencies. Nor was there an effect of the game on global performance; apparently this more holistic, perhaps personal notion of global performance captures for assessors more than only clinical competencies. The fact that both skills are not perceived as the same by assessors was supported by the (only) partial correlation of the global performance and clinical competency scores. Moreover, the self-assessment of the reading & game group only improved for the clinical competencies (not for global performance), compared to the reading group.
Teachers' reactions indicated that the extended (game-based) training preparation had a positive impact on the residents' starting level and on the training. Before, there was a considerable heterogeneity in the group; some trainees had experience with scenario-based ABCDE training, while most residents started with the knowledge from the course manual.

Our study basically compared reading and reading in adjunct with game-play. As the 'reading and game group' spent more time-on-task, It remains open which characteristics of the game were important for skills-learning (the cases, high-fidelity character, the scoring system). The game added an opportunity to practice with the cases cases (virtual patients). The finding that this had a positive impact on residents' clinical skills is consistent with research on task-centered learning, showing that learning with real-world tasks facilitates skills development and transfer to practice[10,16]. In addition, reviews on technology enhanced simulation in health care show that in comparison with no intervention, simulation is associated with large effects on knowledge and skills[17,18]. Future design-based research which controls for time on task will have to clarify the specific features that support skills learning.

Considering the question of games and motivation, we found that residents were motivated to learn with the game and felt engaged with the game cases. This was supported by the fact that they played the patient cases several times and spent 2.5 hours extra self-study time on the game (as showed by logged game-time). This is an important result, as in self-guided, online training programs motivated trainees will put more effort in learning. It also suggests the simulation game could be used as a skills maintenance tool after training.
How do games motivate learners to spend extra time on task? Choice and the opportunity for self-direction appear to enhance intrinsic motivation, as they afford a greater

sense of autonomy[19]. The opportunities for self-directed learning and interest in the subject of the game probably created intrinsic motivation. More research is needed on the specific features that make games engaging for learning, compared to simulations and interactive cases.

Residents who used the game were just as motivated for the course as residents who did not use it. An explanation for this is that these residents already were very motivated for this training, as was shown by the high 'task values'. When residents know they are going to need certain skills in practice, they usually are quite motivated for the course. Both groups had the same (general) self-efficacy levels, but the reading & game group correctly self-assessed their clinical skills as superior to that of the reading group. These findings may be explained by the fact that the game does not easily change the residents' general sense of self-efficacy to be successful during the course, whereas it did make them more aware of their improved emergency skills. Learners often radically overestimate the level of skills they have achieved; the fact we have not found this *illusion of competence*[20] and our residents accurately assessed their clinical skills as improved, in contrast to their communication skills, is a relevant finding for self-directed learning.

After two weeks of training, we no longer found a positive effect of the preparatory game. An obvious explanation for this is learning time: the effect of the 2.5 hours game-time was overshadowed by the 2-week classroom training. Also, in terms of the number of different scenarios done and reviewed, classroom training dominated game training largely (25 versus 6). For complex cognitive skills like problem solving, offering training programs with a high variety of learning tasks is important to allow transfer to new tasks[21]. Considering the relatively high assessment scores at the end of training, there probably also was a ceiling effect in the residents' competencies after the 2-week training.

Given the higher starting level in clinical competencies, a relevant question is whether classroom training can be shortened in combination with the game, maintaining learning outcomes. This would make the blended training design more cost effective because online games are scalable to large numbers of health care professionals without extra costs (in contrast to simulation centers). A study on Advanced Life Support training, comparing assessment results at the end of a conventional 2-days course with a 1-day course supplemented with online interactive simulation cases, showed similar knowledge and skills[8]. Future research should study the retention of game effects on learning outcomes in relation to subsequent shorter classroom training (blended designs).

Limitations

One limitation of our study is the fact that groups were not randomized, thereby confounders may potentially have played a role. Our research groups were very homogeneous (2nd year family practice residents), and we believe we measured the most important possible confounders, such as experience with acute patients and general knowledge on national tests, and found no differences between groups. Another limitation is the relatively small number of participants for the pre-training assessment (n=42). Although the number of residents was limited, they do not differ from the total group on important characteristics and thus may be considered as representative for the groups they came from. In addition, self-assessment data from the total group (n=150) supports the measured improvement in clinical skills from the smaller group. Despite the small sample, we did find (practical) significant differences in skills between groups. Nevertheless, the small numbers limited our possibilities to analyze relationships between game-time, performance and motivation more specifically. It would be interesting to replicate the study with larger groups of residents and a shorter (blended) training design. Thirdly, we did not assess the effects on patient outcomes. Few emergency care courses have had patient outcomes as an end-point[22], but it would be worthwhile to investigate the transfer to clinical practice. Finally, this study is not conclusive on the question of whether learning time or the game format were responsible for the improved skills, as the game group spent more time on task. Many studies on serious games discuss proposed elements that are important for motivation and learning, but relatively few describe empirical evidence[23–26]. Comparative design-based research will have to show which features of simulation cases enhance learning outcomes[27,28]. This 'value-added' research approach (Mayer) will be an important next step in the young field of game-based learning[29].

Conclusion

In summary, this study showed that serious games can be used as an effective, motivating preparation for instructor-led emergency care courses to train residents' clinical competencies. Learning from doing, with a variety of realistic, interactive patient cases, through error and without harming patients are important potential benefits of games and simulation programs. Future research will have to show how this effect can be sustained and whether training time can be reduced in combination with online simulated cases, maintaining learning outcomes. More research is also needed on the game-features which are critical for engaged skills learning.

Acknowledgement

We would like to thank the Training institution for family practitioners (Schola Medica/ SBOH, the Netherlands) for facilitating this research study.

Funding and competing interest statement
External funding for this study was received from the Dutch Ministry of Economic Affairs and Schola Medica (training institution for family practitioners). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. AbcdeSIM is a spin-off company created by Erasmus University Medical Center to aid in selling licenses for the abcdeSIM serious game to other hospitals and institutions. By virtue of standard valorization practices at Erasmus MC Dr. S. Schuit (author) has a financial interest in abcdeSIM. We neutralized this interest by asking TNO, an independent organization with expertise in research on game effectiveness, to participate in the data-collection and analysis. Dr. S.Schuit was in no way involved in either. Data is available.

REFERENCES

1. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, Schünemann HJ. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach. 2010 Jan;32(1):16–27.
2. Clark RC, Mayer RE. E-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning. Elearning. San Fransisco: Pfeiffer; 2008. 476 p.
3. Sitzmann T. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. Pers Psychol. 2011 Jun 27;64(2):489–528.
4. Malone TW. Towards a Theory of Intrinsically Motivating Instruction. Cogn Sci. 1981;4:333–69.
5. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. Elsevier Ltd; 2012 Sep;59(2): 661–86.
6. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012 Oct;99(10):1322–30.
7. Vogel J, Vogel D, Canon-Bowers J, Bowers C, Muse K, Wright M. Computer gaming and interactive simulations for learning: a meta-analysis. J Educ Comput Res. 2006;34:229–43.
8. Perkins GD, Kimani PK, Bullock I, Clutton-brock T, Davies RP, Gale M, Lam J. Improving the Efficiency of Advanced Life Support Training. Ann Intern Med. 2012;157(1).
9. Dankbaar ME, Stegers-Jager KM, Baarveld F, Van Merrienboer JJ, Norman GR, Rutten FL, Van Saase JL, Schuit SC. Assessing the Assessment in Emergency Care Training. PLoS One. 2014;9(12): e114663.
10. Van Merrienboer JJ, Kirschner PA. Ten steps to complex learning. A systematic approach to four-component instructional design. 2nd Editio. Routledge; 2012. 49-77 (Chapter 4) p.
11. Mancini ME, Soar J, Bhanji F, Billi JE, Dennett J, Finn J, Ma MH-M, Perkins GD, Rodgers DL, Hazinski MF, Jacobs I, Morley PT. Part 12: Education, implementation, and teams: 2010 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. Circulation. 2010 Oct 19;122(16 Suppl 2):S539–81.
12. Dankbaar ME, Storm DJ, Teeuwen IC, Schuit SC. A blended design in acute care training: similar learning results, less training costs compared with a traditional format. Perspect Med Educ. 2014 Sep;3(4):289–99.
13. Pintrich PR, de Groot E. Motivational and self-regulated learning components of classroom academic performance. J Educ Psychol. 1990;82(1):33–40.
14. Streiner D, Norman G. Health measurement scales. Oxford University Press; 2008.
15. Hojat M, Xu G. A Visitor's Guide to Effect Sizes. Adv Health Sci Educ Theory Pract. 2004;9:241–9.
16. Francom GM, Gardner J. What is task-centered learning? TechTrends. 2014;58(5).
17. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-Enhanced Simulation to Assess Health Professionals: A Systematic Review of Validity Evidence, Research Methods, and Reporting Quality. Acad Med. 2013 Jul;88(6):872–83.
18. Ilgen JS, Sherbino J, Cook DA. Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. Acad Emerg Med. 2013 Mar;20(2):117–27.

19. Ryan R, Deci E. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol. 2000;25:54–67.
20. Brydges R, Dubrowski A, Regehr G. A new concept of unsupervised learning: directed self-guided learning in the health professions. Acad Med. 2010 Oct;85(10 Suppl):S49–55.
21. Kulasegaram K, Min C, Ames K, Howey E, Neville A, Norman G. The effect of conceptual and contextual familiarity on transfer performance. Adv Health Sci Educ Theory Pract. 2012 Oct;17(4):489–99.
22. Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? Adv Health Sci Educ Theory Pract. 2012 Mar;17(1):137–44.
23. Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simul Gaming. 2002 Dec 1;33(4):441–67.
24. Huang WH. Evaluating learners' motivational and cognitive processing in an online game-based learning environment. Comput Human Behav. Elsevier Ltd; 2011;27(2):694–704.
25. Bedwell WL, Pavlas D, Heyne K, Lazzara EH, Salas E. Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study. Simul Gaming. 2012 May 14;43(6):729–60.
26. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, Simeoni Z, Tran M, Yukhymenko M. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. Rev Educ Res. 2012 Feb 1;82(1):61–89.
27. Norman G. Simulation comes of age. Adv Health Sci Educ Theory Pract. 2014 May;19(2):143–6.
28. Wouters P, van Oostendorp H. A meta-analytic review of the role of instructional support in game-based learning. Comput Educ. Elsevier Ltd; 2013 Jan;60(1):412–25.
29. Mayer RE. Computer games for learning: an evidence-based approach. E-Pub. Cambridge, MA: MIT press; 2014. 264 p.

APPENDIX 1: ASSESSMENT INSTRUMENT EMERGENCY CARE SKILL

Name candidate: ………………………………. Number: ……..

Date: …………………………………………… Rater:…………

	Competency Scale	1 very weak	2 weak	3 insufficient	4 questionable	5 sufficient	6 good	7 excellent
	ABCDE approach							
1	Uses ABCDE approach on initial assessment							
2	Uses ABCDE approach on initial treatment							
3	Uses re-assessment properly							
	Additional actions							
4	Requests additional diagnostic studies							
5	Proposes a working diagnoses							
6	Consults specialist when needed							
	Communication							
7	Communicates with patient effectively							
8	Gives clear instructions to nurse							
9	Radiates a calm and confident attitude							

Global Performance Scale
Assessment of independent function in caring for acutely ill patients in the Emergency Department:… (1-10)

Passed/Failed:

APPENDIX 2: MOTIVATION QUESTIONNAIRE ON THE ABCDE*SIM* GAME

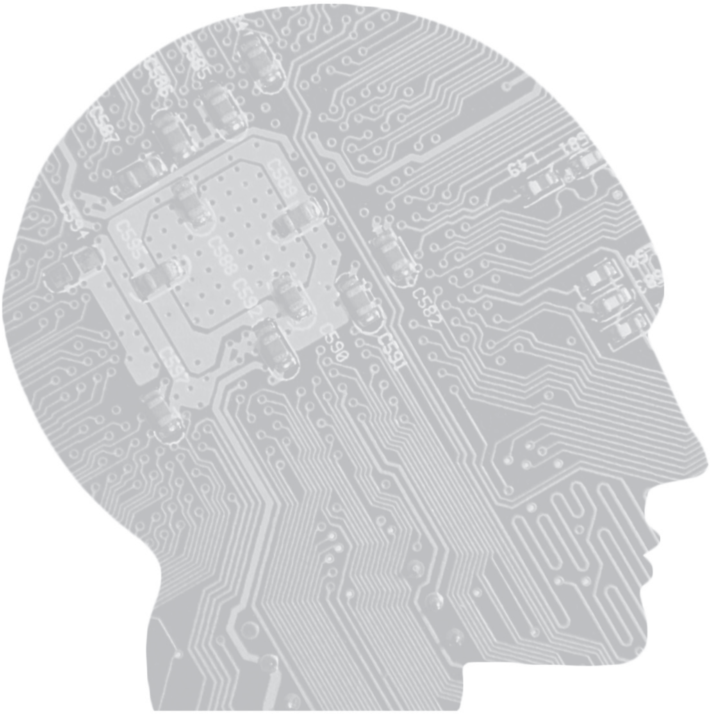
Choose the assertion which fits best your opinion on the game you just played

		Fully disagree	Disagree	Neutral	Agree	Fully agree
1	During play, my attention was very focused on the game cases					
2	I felt actively involved with the game cases					
3	I was able to concentrate well during game play					
4	I liked this way of learning					
5	It was fun to play the game					
6	I could experience for myself what did and did not work in the ABCDE approach					
7	I learned a lot from doing game					
8	During game play, I felt like the emergency doctor in charge					
9	I regularly felt stressed during playing the patient cases					

Do you have any comments to make on the game?

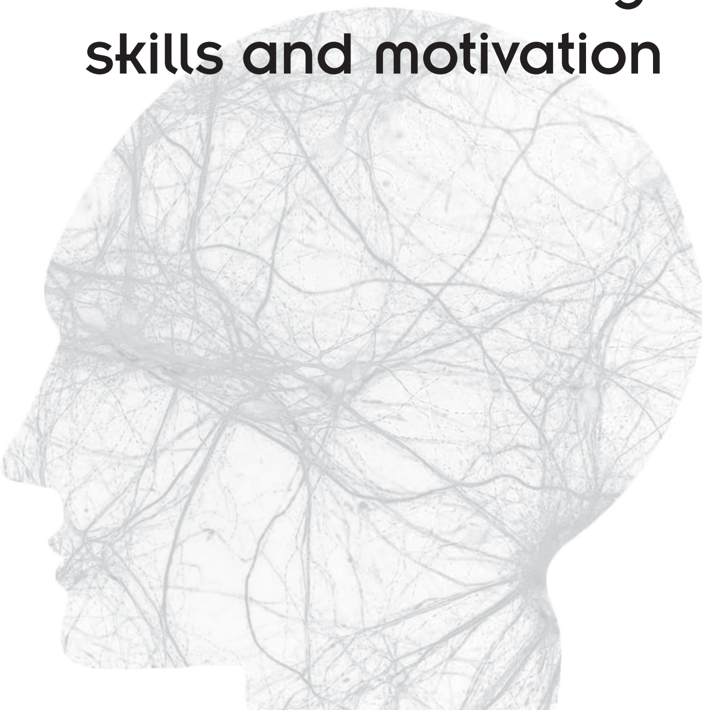
Strong points: ……………………………….

Points of improvement: …………………………….



Chapter 5

An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation



Mary EW Dankbaar
Jelmer Alsma
Els EH Jansen
Jeroen JG van Merrienboer
Jan LCM van Saase
Stephanie CE Schuit

Accepted for publication in Advances in Health Sciences Education

ABSTRACT

Introduction

Simulation games are becoming increasingly popular in education, but more insight in their critical design features is needed. This study investigated the effects of fidelity of open patient cases in adjunct to an instructional e-module on students' cognitive skills and motivation.

Methods

We set up a three-group randomized post-test-only design: a control group working on an e-module; a cases group, combining the e-module with low-fidelity text-based patient cases, and a game group, combining the e-module with a high-fidelity simulation game with the same cases. Participants completed questionnaires on cognitive load and motivation. After a 4-week study period, blinded assessors rated students' cognitive emergency care skills in two mannequin-based scenarios.

Results

In total 61 students participated and were assessed; 16 control group students, 20 cases students and 25 game students. Learning time was 2 hours longer for the cases and game groups than for the control group. Acquired cognitive skills did not differ between groups. The game group experienced higher intrinsic and germane cognitive load than the cases group ($p=0.03$ and 0.01) and felt more engaged ($p<0.001$). Inexperienced students did not profit from working on open cases (in adjunct to an e-module), which nonetheless challenged them to study longer.

Conclusions

The e-module appeared to be very effective, while the high-fidelity game, although engaging, probably distracted students and impeded learning. Medical educators designing motivating and effective skills training for novices should well align complexity and fidelity of cases with students' proficiency level. The relation between case-fidelity, motivation and skills development is an important field for further study.

INTRODUCTION

Although simulation training is becoming widely established within medical education, too little attention is often paid to its effects on motivation and the clinical context[1]. Technology-enhanced simulation training includes computer based simulators, high-fidelity and static mannequins and training with animals or cadavers[2]. It provides learning opportunities for controlled skills practice, without harming the patient. Simulation games can offer learning tasks and instruction in a realistic, engaging computer-based environment, in which trainees directly experience the consequences of their decisions[3]. Well-designed simulation games (or digital games in general) apply game characteristics (players compete in attaining game objectives by following rules and principles) to create a challenging, immersive learning environment which may lead the player into a flow experience[4,5]. Tasks in medical simulation games may contain more or less realistic cases, which can be performed repeatedly without extra costs. Games and simulations are part of the shift from a traditional training model to a learner-centred model, putting the learner into a more active role with challenging and engaging forms of learning[6].
In the next few years, higher education is expected to see an increasing use of games[7,8]. Although several studies have been performed in the last decade, reviews on the effects of simulation or serious games on learning outcomes and motivation have shown mixed and ambiguous results[3,9–13]. Furthermore, there is little consensus on the critical design features that support learning or motivation in games[6].

In contrast, the effectiveness of technology-enhanced simulation training, in comparison with no intervention, has been well established for knowledge and skills[2,14]. In addition, for skills outcomes different design features have proven to be effective in simulations, such as cases with a range of task difficulty, repetitive practice, interactivity and clinical variation[14,15]. Realistic cases in simulation training have often shown to facilitate transfer to practice[16,17]. Similarly, higher-fidelity simulations in emergency care provide greater benefit than lower-fidelity simulation, although definitions of fidelity tend to vary[18]. However, four of five studies comparing high and low-fidelity conditions in training complex skills showed no significant differences in learning outcomes; motivational outcomes were not measured[19]. Recent studies suggest that fidelity is an important factor in simulation-based training, but it is multi-factorial, and the degree of realism required of a simulation is a function of the learning task, the learners and the context[20].

An important consideration in instructional design is cognitive load. For novice learners, high-fidelity cases may be detrimental because they provide too many 'seductive details' and can easily cause cognitive overload. In combination with a complex task, too much (possibly relevant) details in a high-fidelity learning environment may cause overload and impede learning. When confronted with real or high-fidelity tasks, these tasks should initially be relatively simple, decreasing intrinsic load[21]. A simpler presentation of cases (e.g. paper-based) may be just as motivating and also be more profitable for students with little experience. 'Psychological fidelity' pertains to the degree to which a simulated task replicates the skills and psychological factors (stress, involvement) experienced in the real environment. 'Functional fidelity' pertains to the degree to which a simulated task environment reacts to the tasks executed by the learner in a similar way as the real task environment. Depending on the expertise of the learner, the properties of the simulation should be functionally aligned with the criterion task[22]. 'Physical fidelity' pertains to the degree to which the simulated task environment looks, sounds and feels like the real environment[21]. Cognitive load theory assumes that working memory load is affected by intrinsic load (the intrinsic complexity of the learning tasks), extraneous load (the manner in which the tasks are presented) and germane load (the cognitive involvement or learning that actually occurs)[23]. With expertise development, schemes are created which organise knowledge but also reduce working memory load. Proper measurement of the different types of cognitive load can help understand why effectiveness may differ as a function of instructional formats and learner characteristics and how they may interact[24]. Reviews of cognitive load theory underline the need to find load-reducing approaches for knowledge-intensive procedures. An important theme is finding the right balance between stimulating germane processes while offering sufficient structure to avoid cognitive overload[25]. The role of motivation has increasingly been recognized and integrated in cognitive load theory[26].

Training emergency skills is critical for patient safety and substantial resources are involved in training this clinical cognitive skill. It involves systematic evaluation and stabilization of critically ill patients, using the internationally standardized ABCDE approach (Airway, Breathing, Circulation, Disability, Exposure). This method includes several sub skills (procedures) in which the initial resuscitation of critically ill patients is prioritized. We developed a simulation game for residents, where this cognitive skill is trained in a realistic emergency setting, as a preparation for face-to-face training. Our previous study on the effectiveness of this game showed positive effects on residents' clinical cognitive skills[27]. However, this study was inconclusive on answering the question which design features of the simulation game had been responsible for the positive effect.
In the present study, we therefore investigate the effects of adding low-fidelity and high-fidelity patient cases to an instructional e-module on medical students' cognitive skills and motivation to learn with these instructional formats. In this experimental study we used the same simulation game as in our previous study (a virtual emergency care department) as a high-fidelity learning environment. As a low-fidelity environment, we offered the same patient cases as online text-based interactive cases. Both groups started

with an instructional e-module on emergency care. We also added a control group, who exclusively studied this e-module. We chose pre-clerkship (4th-year) medical students as the target population, as they have basic knowledge of emergency care, whereas residents have typically been exposed to these skills before or during their residency program.

We hypothesized that (1) the game group would outperform the cases group on emergency skills, and that the cases group would perform better than the control group on these skills. We expected working on open patient cases would improve performance (compared to instruction only) and we expected the game would stimulate engagement and performance. (2) We expected the high-fidelity game group to be more motivated, study longer, and be more cognitively involved than the low-fidelity cases group.

METHODS

Setting and selection of participants

The study was conducted at the Erasmus MC Medical School, Rotterdam, the Netherlands, with medical students who had finished their Bachelor degree, usually in their 4th-year of study, just before their 2-year clinical clerkships. The basics of emergency medicine are taught in the bachelor curriculum, but this does not include clinical training with simulation patients. Emergency care skills are clinical, primarily cognitive skills: the ability to perform primary assessment of a seriously injured or ill patient, to determine the priority of the necessary actions, and to start initial treatment[28].

The recruitment took place in March 2014 after a 4th-year lecture and through the Medical School's online communication channel for 4th-year medical students. Students were recruited 5 weeks before the assessment to allow time for self-study at home. When they consented to participate they were randomly assigned (using the Excel random function) to the game, cases or control group and received a personal access code. An i-Pad was raffled among the participants who showed up at the assessment. The skill assessments were carried out from 14 to 18 April 2014.

Study design

The control group had access from home to an instructional e-module on emergency care, followed by a knowledge test. The cases group, in addition to the e-module, worked on online text-based cases ('low-fidelity condition'). The game group, in addition to the e-module, worked on a simulation game with the same cases ('high-fidelity condition'). All groups completed questionnaires on motivation and cognitive load after working on the study material. After four weeks of study-time, participants did a mannequin-based

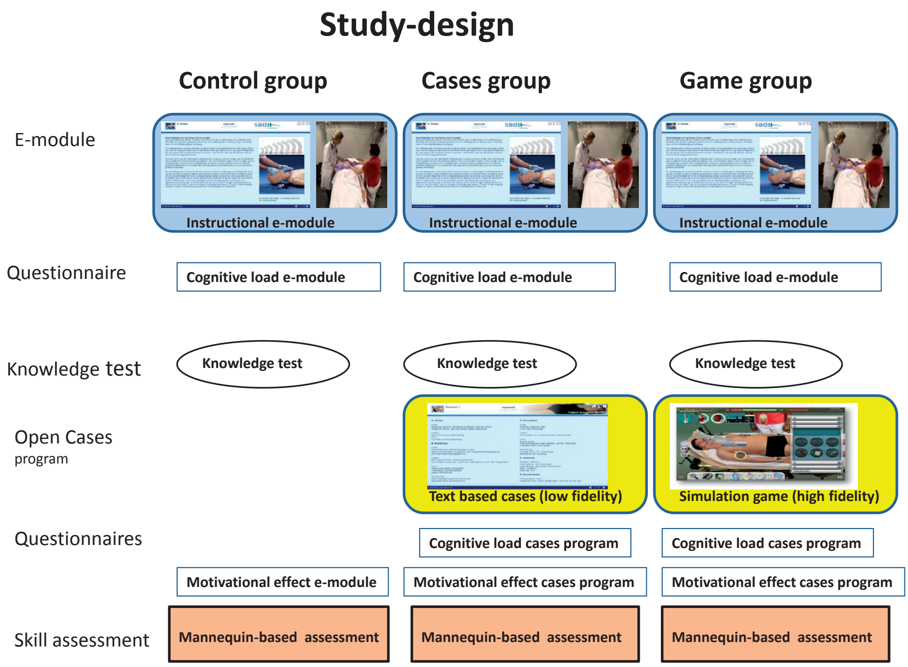


Figure 1: Study-design

scenario assessment. As inter-case variability is known to be an important source of error[29,30] all students did an assessment with 2 cases (figure 1).

Materials

E-module

All groups started their online training with an instructional e-module on emergency care, aimed at developing knowledge on systematically evaluating and stabilizing acutely ill patients, using the internationally standardized ABCDE approach (Airway, Breathing, Circulation, Disability, Exposure). In this approach the Airway, Breathing and Circulation are assessed, using the "look, listen, feel" steps. The information was presented in text with photos, questions with feedback and a 15-minute demonstration video ('worked case') by an experienced doctor and nurse with a standardized patient. The e-module contained no open cases and took about 2 hours to study; students were free to return to the e-module during their work on the cases.

Text-based cases

The text-based cases program was developed for the research study and was designed to train the ABCDE approach. It offered six online, interactive text-based patient cases,

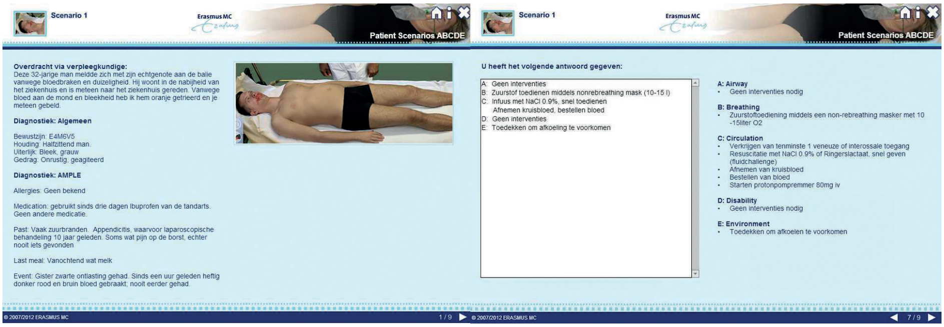


Figure 2: Two screenshots of the text-based cases

each with one patient photo with brief information on the patients' condition and open questions on the initial assessment and treatment of the patient, followed by written feedback. For instance: 'where do you start with this patient and where do you focus on in the physical examination?' was followed by a chronological description of the assessment of the patient (learners can compare their answers). All information on the patients' condition was provided in text; the patient did not react to wrong decisions or interventions. The six patient cases were about acutely ill patients, e.g. gastro-intestinal bleeding leading to circulatory shock and subarachnoid bleeding with seizures. If a case was done for a second time, the same questions and feedback were provided. This *low–fidelity* condition provided lower functional fidelity (feedback on the actions of the learner was text-based, thus less similar to the real task) and lower physical fidelity (the task environment looked and sounded less like a real clinical environment), compared to the game condition. We estimated the cases program took 2-4 hours to study (screenshots of the Cases are in Figure 2).

5

Simulation game ('abcdeSIM')

The computer-based simulation game (abcde*SIM*) was designed for residents to train cognitive emergency care skills (perform the ABCDE approach on different patients). Its design was based on analysing the task demands of this approach and offered an online realistic emergency department with the same six cases, which now took the form of virtual patients. The game started with a storyline in which the patient was presented to the player with brief information on the patients' condition. The player then performed physical examination, aided with tools (e.g. a penlight symbolizes "look", a stethoscope symbolizes "listen" and a hand symbolizes "feel") which had to be used in a correct way to gather the correct information. The stethoscope would only provide breathing sounds if put on the right spot on the thorax, and different locations provided different lung sounds. The player could also order diagnostics (e.g. laboratory testing, ECG); results were presented after a certain timeframe. All sounds, pupil reflexes, lab

Figure 3: Two screenshots of the abcde*SIM* game

results etc. had to be acquired and interpreted by the player and this information should lead to proper treatment of the patient with oxygen, infusion fluids, medication, etc. The (absence of) treatment lead to changes in the patients' vital signs, real time displayed on a medical monitor (with sounds) to increase fidelity and give immediate feedback on the patients' condition. The vital signs were generated by a high-fidelity model of human physiology. Each patient had to be stabilized within 15 minutes; a timer was presented. The storyline ends with an explanation of how the patient faired after the players care, including narrative feedback on the actions and a score. The score depended on the number of correct decisions taken according to the ABCDE approach (e.g. put right oxygen mask on patient: 20 points) and the efficiency of actions (less minutes means more points). Each time a game-case was played, the condition of the patient could vary, depending on the interventions made. Players could compare their scores with peers and with a 'high score' list, to stimulate competition between players. The game did not explicitly train communication skills. This *high-fidelity* condition provided higher functional fidelity (alignment of cognitive demands in the game tasks with the real tasks, including realistic feedback from the patient's reaction on the learner's actions) and higher physical fidelity (the sub tasks, such as using the stethoscope and tools, such as a vital sign monitor looked and sounded like the ones in a real emergency department). The psychological fidelity, related to the cognitive skill at stake including stress, boredom etc. was similar in both conditions. We estimated the game took 2-4 hours to study (screen shots of the Game are in Figure 3).

Assessment instruments and questionnaires

Knowledge test

The self-diagnostic knowledge test consisted of 24, four options Multiple Choice (MC) questions on emergency care. After finishing the test, a fail/pass result was presented (the standard was 75% correct) with the number of correct answers. Students could do the test up to 3 times; the last score was saved.

Skill assessment

Students were assessed at our institution with two 15-minutes mannequin-based (Laerdal Crash Kelly®) scenario assessments on diabetic coma and urosepsis with delirium (in random sequence), taking their low level of experience into account. In both scenario's, students had to work according to the ABCDE-approach, as taught in the E-module. Assessors, who were qualified instructors in emergency courses, introduced and led the scenario. They were familiar with the assessment instrument, blinded for the intervention group and changed scenarios halfway. To get a passing score for the scenario, student had to make a proper diagnosis and start treatment, showing adequacy in the skills that were assessed. The assessment instrument consisted of a *clinical competency scale* (6 items on the ABCDE method), a *communication competency scale* (3 items on communication with the patient and nurse), both rated on a 7-point scale (1=very weak, 7=excellent), and a *global performance scale* (10-point scale to rate 'independent functioning in caring for acutely ill patients in the Emergency Department', 10=perfect). The construct validity and inter-rater reliability of the assessment instrument was validated in a separate study with residents. The clinical competency scale had good construct validity (factor-analysis explained 65% of total variance); the clinical competency and global performance scale had moderate interrater reliability (ICC single measures=0.49/0.50[2]); the communication competency scale had poor inter-rater reliability (ICC=0.27). Internal consistency of scales was high (α=0.91 and 0.86 for clinical and communication competency scales resp.)[31].

Questionnaire on cognitive load

The cognitive load questionnaire has been validated in other work[23]. It consisted of 10 questions; 3 on *intrinsic cognitive load* (e.g. "the content of the e-module/cases/game was very complex"), 3 on *extraneous cognitive load* (e.g. "the explanations and instructions in the e-module/cases/game were very unclear") and 4 on *germane cognitive load* (e.g. "the e-module/ cases/ game really enhanced my understanding of the problems that were discussed"). The questionnaire was scored from 0 (= not all applicable) to 10 (= very much applicable). Sores ≥ 6 for intrinsic load indicate that content is perceived as complex.

Questionnaire on motivation and learning time

A questionnaire on motivation to learn with one of the three formats was developed, based on a previous study on the game effectiveness with residents[27]. The 9-item questionnaire consisted of 2 constructs: engagement ("it was fun to work through the material", 6 items) and feedback ("during learning, I could tell whether I was doing well", 3

---

[2] For Interrater reliability, ICC values (single measures) of 0.21-0.40 are considered 'poor', 0.40 – 0.60 'moderate', 0.60-0.80 'substantial' and 0.80-1.00 'almost perfect'.

5

items); scored on a 5-point Likert scale (1= fully disagree, 5= fully agree). For the control group, the item on the challenging character of cases was left out, leaving 8 items for this group. At the end of this questionnaire, participants also answered some general questions on age, gender, the amount of time they spent on the study material, their clinical experience and whether they had experience with acutely ill patients (questionnaire can be found as Appendix 1). All participants reported learning time separately for the e-module and (if applicable) for the cases or game. Game-time was also computer-recorded.

Statistical analysis

We assessed associations between categorical variables using chi-squared tests. ANOVA analysis and Independent t-tests were used to compare groups on continuous and ordinal variables. Unless the distribution of scores is severely skewed, data from rating scales can be analyzed as if they were interval scales without introducing bias[32]. For statistically significant findings in ANOVA, post hoc analysis was done with S-N-K adjustments. Effect sizes (ES) were calculated using Glass's delta, as homoscedasticity (same variance) is not assumed[33]. Effect sizes of 0.50 were considered medium and ES ≥ 0.80 were considered large[34]. We used Pearson's coefficient to calculate associations. Inter-case reliability in performance assessment was calculated using the Intraclass correlation (ICC). We used regression analysis to determine the factors that were related to clinical skill levels. Analyses were performed using SPSS; we present 95% confidence intervals.

Ethical issues

The study was approved by the Dutch ethical board for research in Medical Education (NVMO, no 210). All participants signed an informed consent form.

RESULTS

Student characteristics

In total 79 students were recruited for the study; they were randomly assigned to the game (n= 30), cases (n= 30) and control group (n= 19). As we were particularly interested in the effects of fidelity on motivation and performance in the cases and game group, we chose to make the control group smaller (25% of the participants). In total 61 students showed up for the assessment 4 weeks later (77 %); 25 from the game, 20 from the cases and 16 from the control group.

There were no significant differences between groups in 'no shows'. The participating students on average were 23 years old, had some clinical experience (in clerkship or with acute patients); a small majority was female. There were no differences between intervention groups in age, clinical experience, or Grade Point Average (GPA) in the Bachelor, but groups did differ in gender: overall 48% of the participants were male, but the control group had more male students (69%) than the game group (28%), (X^2 (2) = 7.2, p = 0.03).

Knowledge test

There were no significant differences in results on the knowledge test between the three intervention groups. The game group had a score of 20.4 (SD = 2.1) the cases group 19.7 (SD = 3.4) and the control group 19.4 (SD = 5.3) out of 24 points. Thus, all groups had the same knowledge after doing the e-module.

Assessment results emergency care skills

There were no significant differences between groups in assessment scores for clinical competencies, communication competencies, and global performance (Table 1). There was no difference in skills between male and female students. The internal consistency (α) of the competency scales for case 1 was 0.84 and 0.86 for the clinical and communication scale respectively; for case 2 it was 0.83 and 0.81.
The inter-case reliability as measured with ICC (single cases) was 0.54 for global performance, 0.41 for the communication competencies and 0.39 for the clinical competencies.

	Total	Control group	Cases group	Game group	F	p-value
n	61	16	20	25		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)		
Clinical competencies (6 items, 7-point scale)	5.4 (0.71)	5.6 (0.64)	5.2 (0.75)	5.4 (0.68)	2.10	0.13
Communication competencies (3 items, 7-point scale)	5.5 (0.52)	5.7 (0.35)	5.4 (0.59)	5.5 (0.52)	2.69	0.08
Global performance (10-point scale)	7.5 (1.0)	7.9 (0.89)	7.2 (1.1)	7.4 (1.0)	2.22	0.12

Table 1: Assessment scores per intervention group (average scores over 2 scenarios)

Learning time and game data

Self-reported total learning time was higher for both the game and cases groups (ca. 4 hours) in comparison with controls (ca. 2 hours), with large variation within groups (F (2, 56) = 5.1, p ≤ 0.05, Table 2). All groups spent the majority of their learning time on the e-module. The game and cases groups did not only spend additional time on the cases or game, but also spent more time on the e-module compared to the controls, although this difference was not statistically significant (F (2,56) = 0.96, p > 0.10). There was no significant difference in total learning time between the game group and cases group. Two students in the game group did not use the game (they were included because there was an 'intention to treat').

	Control group	Cases group	Game group	F	p-value
n	16	20	25		
	Mean (SD)	Mean (SD)	Mean (SD)		
E-module	133 (78)	177 (118)	146 (98)	0.96	0.39
Game/Cases		70 (46)	95 (53)	2.7	0.11
Total self-reported learning time	133 (78)	247 (139)	241 (124)	5.1	0.01*
Logged game time in minutes			90 (49)		

* Significantly different between control and cases/game group

Table 2: Self-reported learning time per intervention group (in minutes)

Analysis of performance of longer playing students in the game group (above the median of 88 min., n= 12) compared to shorter playing students showedslightly higher (not significantly different) scores for clinical competencies (M= 5.6) and global performance (M= 7.8), but they did not outperform the control group. Communication competencies remained the same (M= 5.5).

The mean game score for the game was 3172 points (SD= 1599). There was an association between game time and game score (r= 0.80, p < 0.0001), and between game time and clinical competencies (r= 0.45, p = 0.03). There was no association between self-reported time (on text-based or game-based cases) and clinical competencies (r= -0.12).

Factors influencing skills level

In order to analyse which factors determine the level of clinical competencies, a regression analysis was performed with the variables: intervention group, experience (in clerkship or with acute patients), Grade Point Average (GPA), and total self-reported learning time. GPA was the only significant factor predicting skill level in the total group of students (β = 1.89, p = 0.01). There was no correlation between GPA and game time (r= 0.05, p > 0.1), indicating independence of effects.

Cognitive load

The mean intrinsic cognitive load score of the e-module for all groups was 4.2 (SD = 2.02, n = 54), extraneous cognitive load was 2.6 (SD=1.52) and germane cognitive load was 7.9 (SD = 1.50). Comparing cognitive load of the e-module with cognitive load of the game or cases (for the game/cases group, using a paired t-test) showed that only intrinsic cognitive load of the e-module (M = 4.7, SD = 2.1) was lower than for the game or cases (M = 5.7, SD = 1.6, p = 0.02, ES = 0.45) (CI around the mean difference for intrinsic load is 0.19 ≤ 0.93 ≤ 1.68). The game group experienced significantly higher levels of intrinsic and germane cognitive load than the cases group with large effect sizes, but comparable extraneous cognitive load (Figure 4).

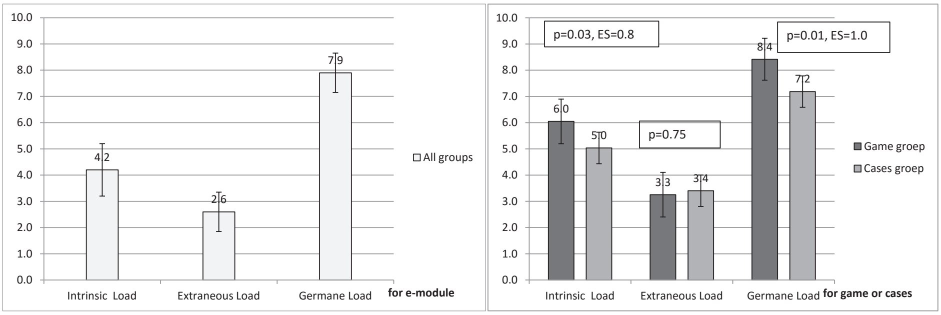


Figure 4: Cognitive load scores for the e-module (all groups) and for the game and cases

In summary, the e-module content was not perceived as very complex, the explanations were clear and students were very actively processing the information from the e-module. The game was perceived as more complex than the cases program and students were more actively processing the tasks. Explanations in the game and cases program were equally clear.

Motivation questionnaire

Mean scores on engagement were different for the three groups (Figure 5). The game group experienced higher engagement than both the cases and control groups (both $p < 0.001$), with large effect sizes. The game group experienced more feedback than the control group ($p < 0.05$, with large effect sizes; CI for mean differences is $0.11 \leq 0.70 \leq 1.28$), but not compared to the cases group ($p = 0.2$). There was no association between the engagement scores and game time ($r = 0.22$, $p > 0.1$).

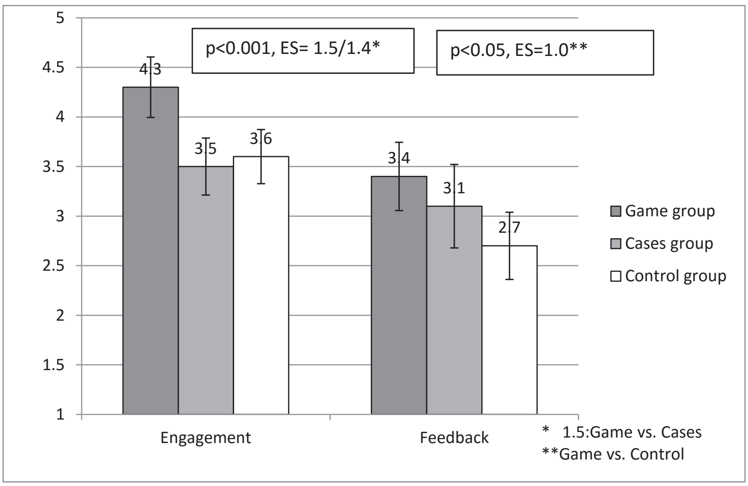


Figure 5: Engagement and feedback scores per intervention group for the e-module, cases and game

5

In open remarks from students from the game group they stated they liked learning with the game, but would have preferred more specific feedback on their actions after doing a case and would have liked additional instructional demonstration videos including the use of the instruments.

DISCUSSION

This study compared the effects of high-fidelity and low-fidelity open patient cases and a control condition (working on an e-module only) on students' cognitive skills and motivation. We showed that complex cognitive emergency care skills of 4th-year medical students were not improved by adding open patient cases as part of a (high-fidelity) simulation game or (low-fidelity) text-based cases to an instructional e-module. Students who worked on the patient cases experienced them as more complex than the instructional e-module, and spent an additional 2 hours studying. Surprisingly, despite this extra study-time, their performance in two scenario-based assessments did not improve. A second finding was that the simulation game, that considerably enhanced students' motivation, was perceived as more complex and stimulated students to put more effort in the cases compared to the text-based cases. Nonetheless, the game group did not study longer and showed the same level of performance as the text-based cases group. The levels of extraneous load were equally low for both cases groups, indicating learning was not hampered by unclear explanations.
Our first hypothesis that the game group would outperform the cases group, and the cases group would outperform the control group on emergency skills was clearly not confirmed. All groups showed the same level of performance in emergency care, despite large differences in learning time. Our second expectation that the game group would be more motivated and show more cognitive involvement than the cases group was confirmed, although, despite their higher engagement, they did not study longer and did not improve their performances.

Why did these 4^{th}-year students not profit from the 2 hours of highly engaged, extra learning time spent on the text- and game-based cases? All groups showed the same knowledge levels on the self-test after studying the e-module. Maybe our study was underpowered and real, existing differences remained unnoticed? Although we cannot exclude this possibility, we believe it is unlikely this would have changed our main conclusions, considering the mean assessment scores of the experimental groups (with limited variation within groups). Apparently the e-module, including a demonstration video on the ABCDE approach, was quite effective. Hence a possible explanation is we observed a ceiling effect and the cases had nothing to add in terms of learning.

Although the assessment scores may seem relatively high, assessors took the low level of students' experience in emergency care into account; as a result these scores cannot be interpreted in absolute terms. The fact that students spent considerable extra time on the cases, and reported high levels of germane load in both conditions makes this assumption unlikely. The opposite argument, the possibility that the assessment setting with a mannequin was too difficult for the students, is in fact more likely. Working with mannequins is not implemented in our medical curriculum and is different from working with simulated cases. Maybe the assessment instrument also was not sufficiently differentiating lower skills levels? The high reliability scores of the competency scales, indicating good variability within groups, and the good inter-case reliability do not provide support for this assumption. The fact that neither the cases nor the game group improved learning indicates that students probably experienced a problem with the open patient cases; we expect they were too complex to solve without guidance. This is confirmed by the high germane load scores in both case conditions and the fact that the intrinsic load in these conditions was perceived as higher than intrinsic load in the e-module. Moreover, the 2 hours extra learning time was only partly spent on the cases; the majority of the 4 hours learning time was spent on the e-module. Students may have returned to the e-module for supporting information to solve the case problems. Apparently for these novices in emergency care, the cognitive gap between the e-module and the six open cases was too large. This assumption is supported by remarks from students that they would have liked more video demonstrations and more specific feedback in the cases. Spending extra time on the game and text-based cases may even have resulted in a shift from being "unconsciously incompetent" to being "consciously incompetent", making them more insecure during assessment (but possibly more eager to learn in the future).

In summary, we believe the e-module (with the video) was quite effective in teaching the principles of the ABCDE approach, the open case problems were too complex for the students and possibly the assessment situation was too demanding as well. Our finding that for initial skills acquisition, solving open cases does not necessarily add anything in terms of learning to (online) instruction with a demonstration video, is in line with research showing that studying worked examples is superior to solving training problems, attributed to the fact that examples allow for building a cognitive scheme[35,36]. In addition, the effectiveness of online instruction for learning skills is consistently confirmed in review studies[3,37,38], as has the effectiveness of instructional videos. For instance, in a critical care situation, students who watched a video demonstration or practiced hands-on in a high-fidelity simulator generated the same responses[39]. Another study with surgical novices showed that video training in adjunct with expert instruction did not improve the development or retention of surgical skills more than video training alone[40].

Contrary to our expectations, the high-fidelity game group spent the same amount of learning time as the low-fidelity cases group although, as expected, the game group was more motivated and was more cognitively involved in learning. The cases group was less engaged to learn with their format, but apparently was sufficiently motivated to learn about the ABCDE approach, resulting in the same time on task and the same skills level. Was the game more distracting than helpful for learning? The association between game time and clinical skills indicates functional alignment between game-tasks and assessed tasks probably did not fail completely, as this association was independent from academic performance or students' motivation. However, since longer playing students (> 90 minutes) still did not outperform the control group, the game is probably distracting for novices, specifically in the beginning. A possible explanation is that the attributes which created higher physical fidelity in the game (the tasks and tools in the game, which looked and sounded like a real emergency room) were very engaging for students, but at the same time confused them, creating overload in combination with the task demands, and deteriorated learning. The attributes which created higher functional fidelity (alignment of cognitive task demands with real tasks, including realistic feedback from the patients' reactions) may be only effective after players spent more time, got used to the sounds and tools, and were able to concentrate on the cognitive tasks, if at all. Further research is needed to confirm this conjecture. Measurement of cognitive load during game play rather than one measurement after playing different game cases would be useful to gain more insight in this process. In addition, the extraneous load items (currently addressing "unclear instructions"), might be extended with items addressing distracting elements in the learning task.

The abcde*SIM* game was originally developed for residents. Compared to 2nd year residents who did show improved clinical cognitive skills after playing the same game for (on average) 130 minutes, these students spent limited game-time (90 minutes) and also had lower game-scores[27]. This is an important finding, indicating an 'expertise reversal effect' where a rich learning environment (with tasks at a specific level of complexity) may benefit experts, but is counter-productive for novice learners[41]. Another important difference between residents from the previous study[27] and students in the current study is that residents were preparing for an emergency department traineeship two weeks later, increasing their dedication to learn this skill. Students volunteering for a research study do not have the same focus during learning.
Is it possible that other game attributes than fidelity stimulated engagement and cognitive effort? The serious game and text-based cases mainly differed in functional and

physical fidelity, but the game also had a scoring system to stimulate competition and longer play. Although we cannot exclude the possibility that engagement was positively influenced by the scoring system, competition did not lead to more time on task or higher skills levels in the game group compared to the cases group. Mayer concluded from several studies there is not yet convincing evidence that competition in games stimulates learning[42].

In summary: the high-fidelity simulation game enhanced motivation and cognitive effort compared to the cases, but appears to be distracting and impeded learning for novices. We expect especially the physical fidelity (tools and sounds in the game) provided too many details and might have created cognitive overload. Our finding that high-fidelity presentation of cases did not improve learning for novices is in line with a number of other studies. Formats with different levels of authenticity (paper-based cases, video cases and simulated patients) did not influence students' skills and performance on the short and longer term[43]. Two studies comparing text-based and video-enhanced virtual patients showed better respectively worse reasoning outcomes for the text-only format, with small effect sizes[44]. Schematic representations in (non-medical) simulation games often were more effective than realistic representations[12]. However, Brydges et al. found that allowing students to progress in their practice on simulators of increasing fidelity improved transfer of a broad range of clinical skills[45]. More research is needed on the interplay between complexity of cases, (functional and physical) fidelity, motivation and performance development for different proficiency levels.

Implications for medical education

This study yields a number of implications for designing engaging and effective online skills training in medical education. For initial cognitive skill acquisition, an instructional e-module with one or more demonstration videos is a powerful instructional format. Additional patient cases stimulate students to study longer and be cognitive involved, but do not necessarily result in learning when these tasks are too complex. Worked cases, part task practice, hints and gradually increasing task complexity in open cases can be used to help students build the cognitive schemes required as a preparation for doing more complex open cases[21]. Although high-fidelity patient cases in a simulation game may enhance motivation and cognitive involvement, they can easily distract novice students and impede learning. In particular physical fidelity can easily create cognitive overload for novices. Learning task fidelity should only gradually increase as learners become more proficient[46]. As the same game, originally developed for residents, resulted in improved clinical competencies with family practice residents[27], our results from the two studies are an illustration of the expertise reversal effect, where a rich learning environment may benefit experts, but is counter-productive for novice learners. Our finding that motivation and cognitive involvement were unrelated to learning outcomes is an important notion for game-based learning and research, as objectively measured learning outcomes are not always part of models in use. Moreover, a simulation game that motivates and stimulates learners has little practical meaning in the absence of real learning effects.

Limitations

This study has a number of limitations. First, the sample size of the control group was limited, as a result comparisons between the control group and the game or cases group may have been underpowered. Considering the assessment scores and limited variation within groups, we don't expect larger samples would have changed our conclusions. However, in a follow-up study we recommend to use larger samples for the control group. A second limitation was the fact that patient cases appeared to be too complex, which made it difficult to investigate the main research question regarding the effect of fidelity in cases. Based on the combined data from the assessment, game play and questionnaires we were able to draw some conclusions, but in a future study it is advisable to test complexity of cases before the start of a research study, e.g. in a pilot. A third limitation was the problem of confounding, as the game differed in several attributes from the text-based condition, such as physical and functional fidelity and a scoring system. We argued it is unlikely the scoring system had a considerable influence on the main results, but it was not possible to separate the effects of physical and functional fidelity. Although in games different attributes often are intertwined, for design-based research it is important to develop and investigate game versions with unique attributes. A last limitation is the fact that this study is focused on cognitive emergency care skills of 4th-year medical students. As competency level of learners is an important factor in the effectiveness of instructional design, findings cannot be generalised to other groups or domains.

Conclusions

This study compared the effects of an emergency care simulation game with a text-based cases program (with the same cases) and a control condition (working on an e-module only) on students' clinical cognitive skills and motivation. For initial skills acquisition, an e-module with a video demonstration is a powerful instructional format. These students, who were inexperienced in emergency care, did not profit from additional work on open cases, which appeared to be too complex, but nonetheless challenged them to study longer. The high-fidelity simulation game increased complexity and did not improve their skills-level, even though students put more effort into it and felt more engaged with it than with the low-fidelity cases. Probably the game distracted students and impeded learning. In the design of motivating and effective cognitive skills training for novices, complexity and fidelity of cases should well be aligned with students' proficiency level. More design-based research is needed on the relation between case-fidelity, motivation and skills development for novices and experts to make informed choices in this increasingly important field.

Contribution
The authors wish to thank Geoff Norman for reviewing the paper.

REFERENCES

1. Kneebone R. Evaluating Clinical Simulations for Learning Procedural Skills: A Theory-Based Approach. Acad Med. 2005 Jun;80(6):549–53.
2. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek J, Wang A, Erwin P, Hamstra S. Technology-Enhanced Simulation for Health Professions Education; A systematic review and meta-analysis. JAMA. 2011;55905(September):978–88.
3. Sitzmann T. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. Pers Psychol. 2011 Jun 27;64(2):489–528.
4. Huang WH, Huang WY, Tschopp J. Sustaining iterative game playing processes in DGBL: The relationship between motivational processing and outcome processing. Comput Educ. Elsevier Ltd; 2010;55(2):789–97.
5. Kiili K. Digital game-based learning: Towards an experiential gaming model. Internet High Educ. 2005;8:13–24.
6. Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simul Gaming. 2002 Dec 1;33(4):441–67.
7. Johnson L, Adams BS, Cummins M, Estrada V, Freeman A LH. Horizon Report 2013 Higher Education Edition. Consortium T new media, editor. Austin, Texas; 2013.
8. Prensky M. The Digital Game-Based Learning Revolution. New York: Mc. Graw Hill; 2001. 1-19 p.
9. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, Schünemann HJ. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach. 2010 Jan;32(1):16–27.
10. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. Elsevier Ltd; 2012 Sep;59(2): 661–86.
11. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012 Oct;99(10):1322–30.
12. Wouters P, van Oostendorp H. A meta-analytic review of the role of instructional support in game-based learning. Comput Educ. Elsevier Ltd; 2013 Jan;60(1):412–25.
13. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, Simeoni Z, Tran M, Yukhymenko M. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. Rev Educ Res. 2012 Feb 1;82(1):61–89.
14. Issenberg S, McGaghie W, Petrusa E, Gordon D, Scallese J. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach. 2005;(27(1):10-28
15. Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hatala R. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Med Teach. 2013 Jan;35(1):e867–98.
16. Bransford J, Brown A, Cocking R. How people learn: Brain, Mind, Experience, and School. Washinton DC: National Academic Press; 2000. 3-127 p.
17. Schuwirth LWT, van der Vleuten CPM. The use of clinical simulations in assessment. Med Educ. 2003 Nov;37 Suppl 1:65–71.

18. Ilgen JS, Sherbino J, Cook DA. Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. Acad Emerg Med. 2013 Mar;20(2):117–27.
19. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. Med Educ. 2012 Jul;46(7):636–47.
20. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. Acad Med. 2014 Mar;89(3):387–92.
21. Van Merrienboer JJ, Kirschner PA. Ten steps to complex learning. A systematic approach to four-component instructional design. 2nd Editio. Routledge; 2012. 49-77 (Chapter 4) p.
22. Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? Adv Health Sci Educ Theory Pract. 2012 Mar;17(1):137–44.
23. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educ Psychol. 2003 Mar;38(1):63–71.
24. Leppink J, Paas F, Van der Vleuten CPM, Van Gog T, Van Merriënboer JJG. Development of an instrument for measuring different types of cognitive load. Behav Res Methods. 2013 Dec;45(4):1058–72.
25. De Jong T. Cognitive load theory, educational research, and instructional design: Some food for thought. Instr Sci. 2010;38(2):105–34.
26. Paas F, Tuovinen JE, Merriënboer JJG, Aubteen Darabi a. A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. Educ Technol Res Dev. 2005;53(3):25–34.
27. Dankbaar M, Bakhuys Roozeboom M, Oprins E, Rutten F, van Saase J, van Merrienboer J, Schuit S. Gaming as a training tool to train cognitive skills in Emergency care: how effective is it? In: Schouten B, Fedtke S, Schijven M, Vosmeer M, Gekker A, editors. Games for health, Proceedings of the 4th conference. Springer; 2014. p. 150.
28. Breedveld. Rapportage werkgroep Kwaliteitsindeling SEH Spoedeisende hulp: Vanuit een stevige basis. 2009;
29. Swanson DB, van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art revisited. Teach Learn Med. 2013 Jan;25 Suppl 1(December):S17–25.
30. Norman G. Research In Medical Education: Three decades of progress. BMJ. 2002;324(7353):1560–2.
31. Dankbaar ME, Stegers-Jager KM, Baarveld F, Van Merrienboer JJ, Norman GR, Rutten FL, Van Saase JL, Schuit SC. Assessing the Assessment in Emergency Care Training. PLoS One. 2014 ;9(12):e114663.
32. Streiner D, Norman G. Health measurement scales. Oxford University Press; 2008.
33. Grissom R, Kim J. Effect sizes for Research - Univariate and multivariate applications. 2nd ed. New York: Routledge; 2012. 66-73 p.
34. Hojat M, Xu G. A Visitor's Guide to Effect Sizes. Adv Health Sci Educ Theory Pract. 2004;9:241–9.
35. Stark R, Kopp V, Fischer MR. Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. Learn Instr. Elsevier Ltd; 2011;21(1):22–33.

5

36. Leppink J, Paas F, van Gog T, van der Vleuten CPM, van Merriënboer JJG. Effects of pairs of problems and examples on task performance and different types of cognitive load. Learn Instr. Elsevier Ltd; 2014 Apr;30:32–42.

37. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008 Sep 10;300(10):1181–96.

38. Sitzmann T, Kraiger K, Stewart D, Wisher R. The comparative effectiveness of web-based and classsroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.

39. Baxter P, Akhtar N, Landeen J, Norman G. Teaching critical management skills to senior nursing students: videotaped or inter-active hands-on instruction? Nurs Educ Perspect. 2012;33(2): 106–10.

40. Nousiainen M, Brydges R, Backstein D, Dubrowski A. Comparison of expert instruction and computer-based video training in teaching fundamental surgical skills to medical students. Surgery. 2008 Apr;143(4):539–44.

41. Kalyuga S, Ayres PL, Chandler P, Sweller J. The expertise reversal effect. Educ Psychol. 2003;38(1): 23–31.

42. Mayer RE. Computer games for learning: an evidence-based approach. E-Pub. Cambridge, MA: MIT press; 2014. 264 p.

43. Durning SJ, LaRochelle J, Pangaro L, Artino AR, Boulet J, van der Vleuten C, Hemmer P, Denton D, Schuwirth L. Does the authenticity of preclinical teaching format affect subsequent clinical clerkship outcomes? A prospective randomized crossover trial. Teach Learn Med. 2012 Jan;24(2): 177–82.

44. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010 Oct;85(10):1589–602.

45. Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximize educational benefit. Acad Med. 2010 May;85(5):806–12.

46. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 2015;119–27.

APPENDIX 1: MOTIVATION QUESTIONNAIRE

Choose the assertion which fits best your opinion on the study material

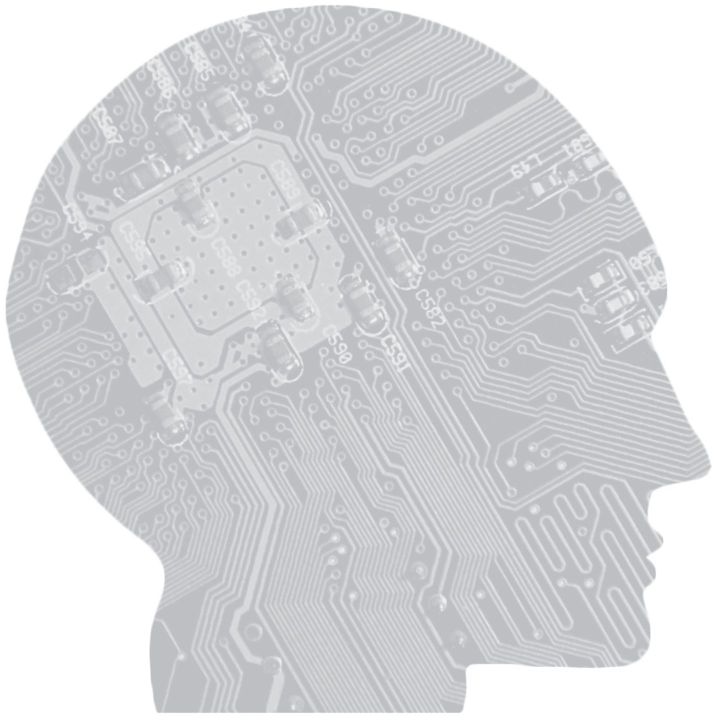
		1 Fully disagree	2 Disagree	3 Neutral	4 Agree	5 Fully agree
E 1	I think the e-module/cases/game can be valuable for me					
E 2	I think this study material is helpful for me to master the ABCDE approach					
E 3	I liked this way of learning					
E 4	It was fun to work through the material					
E 5	After working through the material, I felt encouraged to study it again					
E 6	*It was challenging to perform well on the patient cases* (not for control group)					
F 1	During learning, I could tell whether I was doing well					
F 2	I could experience for myself what did and did not work in the ABCDE approach					
F 3	I received sufficient feedback					

Do you have any remarks on the content of the material (positive or negative)?

Positive: ……………………………….

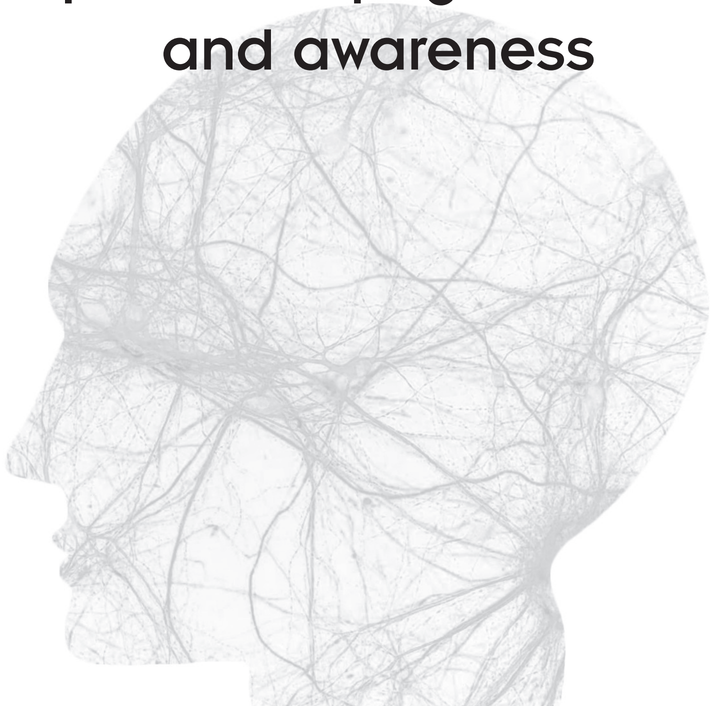
Negative: …………………………….

E: Engagement Questionnaire; F: Feedback Questionnaire



Chapter 6

Comparative effectiveness of a serious game and e-module on patient safety knowledge and awareness



Mary EW Dankbaar
Olivier Richters
Cor J Kalkman
Gerrie Prins
Olle J ten Cate
Jeroen JG van Merrienboer
Stephanie CE Schuit

Submitted for publication

ABSTRACT

Purpose

Serious games are engaging training formats with the potential to teach patient safety awareness at relatively low costs. We investigated whether undergraduate medical students developed better patient safety knowledge and awareness and were more motivated after using a serious game, compared with peers who studied the same topics with an e-module.

Methods

Fourth-year medical students were randomly assigned to either a serious game that included video-lectures, biofeedback exercises and patient missions or an e-module, that included text-based lectures on the same topics. A third group acted as a historical control group without extra education. After the intervention, all students completed a knowledge test, a self-efficacy test and a motivation questionnaire. During the following 10-week clinical rotation they filled out weekly questionnaires on perceived stress and patient-safety awareness and behavior.

Results

In total 103 students participated, 32 in the game, 34 in the e-module and 37 in the control-group. Knowledge was equally improved in the game and e-module-groups compared to the control group. Average learning-time was 3 hours for the game and 1 hour for the e-module-group. During rotations, students in the three groups reported low and similar levels of stress and patient-safety awareness. Students who had treated patients successfully during game missions experienced higher self-efficacy but less stress during their rotation. The game, particularly the patient mission-part, was evaluated as considerably more engaging than the e-module, with large effect sizes.

Conclusions

Video-lectures and text-based lectures are equally effective in developing knowledge on patient safety topics. Although serious games are strongly engaging for students and stimulate them to study longer, they do not necessarily result in better patient safety awareness. More research in this domain is needed.

INTRODUCTION

Medical injuries in hospitals pose a significant threat to patients and incur substantial costs to society[1]. A decade after the publication 'To err is human'[2], Landrigan et al., reported that 'harms remain common, with little evidence of widespread improvement.'[3]. An important determinant of patient safety in hospitals is awareness and technical and non-technical skills of junior doctors and their supervision. They are typically first responders to deteriorating patients during nights and weekends; lack of patient-safety training for clinicians appears to be a major contributor to preventable harm[4]. Moreover, experiencing medical errors can have severe psychological impact on students[5]. Traditionally, medical curricula have focused on three major competencies: medical knowledge, technical skills and clinical decision making. The aviation industry previously recognized that many major incidents were the result of failures in non-technical skills[6] and have now incorporated team training and situational awareness in their training programs, but introduction of such training in healthcare has been slow in most countries[7,8].

The Institute of Medicine has called for reform of health professions education to advance healthcare safety and quality[9]. The WHO's World Alliance for Patient Safety developed guidelines and resources for a patient safety program, covering topics as the systems approach, being a team player and learning from errors[10]. Considering the already packed medical curricula, the scale of training needs, tight training budgets and limited training time in health care organizations, current training formats no longer suffice to meet the needs for patient safety training. Novel, effective and more efficient training formats are needed in adjunction to traditional training.

Online instructional formats such as e-modules[11–14] and computer-based simulation programs[15,16] offer flexible and cost-effective learning opportunities and are known to enhance knowledge and skills. They can be used for self-study or 'blended' with classroom training. Serious games offer new experiential, engaging learning opportunities for complex skills learning[17]. Games in patient safety education have the potential to teach awareness and the basics of teamwork skills that are typically acquired in simulation settings, but at a fraction of the costs[18]. Effectiveness studies on serious games in general however have shown mixed and ambiguous results[19–22]. So far, very few comparative studies on effectiveness of online programs or serious games in patient safety education have been conducted. To move patient safety education one step further, it is important to know the comparative educational value of these instructional formats. If they prove to be engaging as well as effective for learners, their flexibility in use (anytime/place) and scalability (a high-quality online program can teach large groups of trainees at low costs per person) might offer advantages over traditional formats. We compared the effects of a serious game on patient safety with an e-module on the same topic.

In the present study, we investigated whether undergraduate medical students developed better patient safety knowledge and awareness and were more motivated after studying a serious game (experiential learning) in comparison with a simple e-module (instructional learning). The game included video-lectures on patient-safety issues, biofeedback exercises for stress management and patient-missions to stimulate patient-safety awareness. The e-module was text-based, with the same content. We compared both groups with a control-group who received no extra education. Following Kirkpatrick's framework of evaluation[23,24] we assessed students' satisfaction with the game and e-module (level 1), their knowledge and self-efficacy (level 2), and their self-reported stress and patient-safety awareness during practice (between level 2 and 3). Our hypotheses were:

a) knowledge on patient-safety will equally improve among students in the game and e-module-groups, but more than in the control-group, as video and text-based lectures have a positive, but similar effect on knowledge.
b) Perceived stress in subsequent clinical rotations will be lower in the game-group than in the e-module or control-group, as biofeedback exercises provide more adequate feedback than text-based exercises or nothing.
c) Patient-safety awareness and self-efficacy will be higher in the game-group than in both other groups, as there were no real substitutes for the game-based missions in the e-module.
d) Students in the game-group will be more motivated to learn with the game than students from the e-module-group.

METHODS

Setting, design and selection of participants

The research population consisted of fourth-year undergraduate medical students, doing a clinical introduction course immediately before the start of a 10 week Internal Medicine rotation (first rotation). The one-week introduction course consisted of clinical training and patient safety education.

Students who consented to participate from September 2013-February 2014 were randomly assigned to the game or e-module-group, using the Excel random-function. Students who did the introduction course 6 months earlier, from April–September 2013, and consented to participate were used as historical controls. We used this design to prevent the risk of contamination of conditions as students from the e-module or game-group might share access with the control-group (many 4th year-medical students know each other well). Primary outcome measures were (i) motivation to do the e-module or game, (ii) knowledge on patient safety, (iii) self-efficacy and (iv) self-reported stress and patient safety awareness during the clinical rotation.

On the first day of the introduction course, students were asked to participate in the study. When students from the control-group consented, they received no extra instruction. When students from the game/e-module-group consented, they received the game (game-group) or link to the module (e-module-group) on the same day with a brief instruction. On the last day of the course, all groups completed a knowledge test, self-efficacy test and evaluation questionnaire. Students from the game-group returned the game and game-data were extracted. During the following 10-week clinical rotation, they filled out a weekly questionnaire on perceived stress and patient safety issues. Blinding was secured by the fact that one of the authors (OR) performed the randomization and data-collection from online questionnaires, and another author (MD) analyzed the data.

Every 2 weeks, approximately 13 students started the introduction course and the following clerkship Internal Medicine. These fourth-year medical students are approximately 23-24 years old, and 60 to 70% is female. We raffled an i-Pad among participating students and continued recruitment until groups consisted of ± 40 students. Data collection took 12 months.

Materials

Serious game 'Air Medic Sky-1'

A a serious game on patient safety was developed for starting residents and medical students ("Air-Medic-Sky-1"). Learning goals are: to stimulate patient safety awareness (e.g. using 'hear-back', recognizing 'red flags' in clinical situations) and learning to perform effective stress management (e.g. being able to use simple breathing exercises to deal with stress).

The game offers three parts: a) brief video lectures by international experts about patient safety topics such as communication, focus under stress, teamwork; b) biofeedback-breathing exercises to enable learners to consciously influence their own stress levels. The game comes with a biofeedback device with finger sensors to measure heart rate variability and skin conductance. Players gain points by watching video lectures and by doing breathing exercises. After gathering a certain amount of points, the player will be invited to participate in c) patient care missions to disaster areas around the world in a virtual flying hospital, where patients can be diagnosed and treated. During these missions, players learn to combine treating several patients simultaneously, while communicating and collaborating with superiors and nurses. Players gain patient-care

Figure 1: Three screenshots of the Air Medic Sky-1 game and a photo of the finger sensors

points by doing this successfully. From previous beta testing we knew that students need at least 3 to 4 hours to play all main sections of the game (See Figure 1 for game screenshots).

E-module

For the benefit of this study we developed an e-learning module on patient safety, covering the topics from the video lectures (communication, focus under stress, teamwork, etc.) and stress management. The information on patient safety was offered as structured, written text, with limited interaction. Breathing exercises were also provided in text. No replacement for the health care missions was provided, as interactively treating patients was not possible in this format. We estimated that students needed 1-2 hours to study the e-module (Figure 2 shows screenshots).

Table 1 provides an overview of the two experimental conditions and learning goals.

Assessment instruments and questionnaires

Knowledge test

The knowledge test consisted of 126 multiple choice (MC) questions; 100 true/ false questions, 24 three, four- or five-option MC questions and 2 matching questions (max 70 points). The knowledge test was designed by patient-safety and education experts at UMC Utrecht and was derived from the video lectures (Test is available upon request).

Figure 2: Three Screenshots of the e-module

Self-efficacy test

Existing self-efficacy tests are not sufficiently dedicated to patient-safety. To assess self-efficacy regarding patient-safety, we designed a 12-item validated questionnaire on

6

Learning goals	Serious game	E-module	Assessment instrument
Knowledge on patient safety issues	Video lectures	Written text on the same content	Knowledge test
Stress management	Biofeedback exercises	Written text on biofeedback exercises	- Self-efficacy/stress management - (Self-reported) stress management during clinical rotation
- Stimulate patient safety awareness - Learn to perform effective teamwork & communication	Patient missions: diagnosing and treating patients, while communicating and collaborating optimally with nurses, supervisors and the patient's family	-	- Self-efficacy/communication - (Self-reported) patient safety behavior during clinical rotation: awareness of adverse events and actions undertaken

Table 1: Overview of how two intervention programs cover the learning goals

ability in communication issues (e.g. hand over patient information, perform debriefing) and on recognizing patient safety threats (e.g. recognition of sleep deprivation). Students rated their self-efficacy on a 1-100 scale (1= I cannot do this, 100= I can do this perfectly, see Appendix 1). Turner et al. have provided validity support for this approach[25].

Questionnaire on perceived stress and patient safety awareness

In the absence of valid patient safety competencies assessment tools[26], we designed a questionnaire aimed at perceived stress and awareness of patient safety issues during the 10-week clinical rotation. For perceived stress, the Perceived Stress Scale, a widely used validated psychological instrument for assessing the degree to which situations in one's life are appraised as stressful was used as a basis[27]. We used three questions which were relevant in our context: how often students felt stressed, whether they felt they were able to cope with their tasks, and whether they felt to be in control. In addition, three questions on patient-safety awareness were included: whether students had experienced any adverse events, what their own action in response to this event was and to provide an example of events and actions. All 6 questions related to the preceding week and were filled out on a weekly basis. Scales were rated from 1 (never) to 5 (very often). High scores indicated high levels of stress or adverse events, questions stated negatively were reversed before analysis (see Appendix 2 for the weekly questionnaire). Examples were reported separately.

Evaluation questionnaire and interviews

Several questionnaires exist to evaluate (new) online instructional formats. For evaluation of the e-module and game, we used a combination of questionnaires on usefulness, ease-of-use, satisfaction, attitude, motivation[28,29] and engagement[30]. This resulted in a 23 item-questionnaire, with items scored between 1 (strongly disagree) to 5 (strongly agree, indicating positive opinions). Negatively stated questions were reversed before analysis. The questionnaire ended with an open question on positive and negative aspects of the game/e-module (see Appendix 3). In order to explore the strong and weak aspects of the game further, short semi-structured interviews have been conducted with seven game-players, starting with an open question on the value of the game and then going through the three parts. Important additional information from the interviews and open questions was summarized.

Statistical analysis

We assessed associations between categorical variables using chi-squared tests. ANOVA, post-hoc and independent sample t-tests were used to compare groups on continuous variables; paired t-tests were used to analyze differences within groups. Unless the distribution of scores is severely skewed, data from rating scales can be analyzed as if they were interval without introducing bias[31]. Effect sizes (ES) were calculated using Glass's delta[32]; ES ≥ 0.80 were considered large[33]. Associations were calculated using Pearson's coefficient; we report coefficients ≥ 0.50. Confidence intervals of 95% were used to correct multiple correlations and for general analyses. A factor-analysis was performed on the self-efficacy and evaluation questionnaire, determining important constructs. The reliability of scales was assessed using Cronbach's alpha. We treated missing data with pairwise deletion and used SPSS for the statistical analysis.

Ethical issues

The study was approved by the Ethical Review Board of the Netherlands Association for Medical Education (NVMO, no 209). All participants signed an informed consent form.

RESULTS

Participants

Of 52 eligible students 39 consented to participate and were admitted to the control-group (75%). Of 156 eligible students, 90 consented to participate in the e-module/game-group (58%), they were randomly allocated to either the game or e-module-group. Students with empty or incomplete surveys (less than 2 questionnaires, usually because they reportedly could not find time to do the game or e-module) were eliminated from the study, leaving 37 students in the control-group (54% female), 34 in the e-module (68% female) and 32 in the game-group (78% female). In total 103, students participated, with no significant differences in gender between groups (x^2 (1)=4.49, p>0.11).

6

Knowledge test

The reliability of the knowledge test was low: α=0.50. Test scores of the e-module-group (M=35.5, SD=4.1) and game-group (M=36.8, SD=4.1) were significantly higher than in the control-group (M= 32.3,SD=5.5, F(2,90)= 7.7, p=0.001), with medium to large effect sizes (ES=0.6 and ES=0.8 for the e-module and game-group resp.) There were no differences in scores between the game and e-module-group.

Self-efficacy

The reliability of the 12-item self-efficacy scale (1-100) was good: α=0.79. Factor-analysis showed that two constructs could be identified within the scale: *self-efficacy on communication* in *patient safety issues* (e.g. 'perform a debriefing', 6 items, α=0.81) and *self-efficacy on recognition of patient safety threats* (e.g. 'recognize sleep deprivation', 6 items, α=0.79). Mean scores were low (12-item-scale, M=67.1, SD=9.2) and did not differ between groups on either scale. All students showed higher scores on the communication subscale (M=68.8, SD=10.8) than on the recognition of threats subscale (M=65.4, SD=12.7), t(99)=2.30, p<0.05).

Reported stress and patient safety awareness

The reliability of the 3-item, 10-week reported stress scale (1-5) was high: 0.92. The reliability of the 2-item, 10-week reported patient safety behavior scale (1-5) was good: 0.78. Average reported stress in the game-group was lower than in the e-module-group, but overall there were no significant differences between the three groups on stress or patient safety behavior and scores were low (Table 2).

	Control-group Mean (SD)	E-module group Mean (SD)	Game group Mean (SD)	P* (2-tailed)
Reported stress	2.2 (0.5)	2.4 (0.4)	2.1 (0.4)	0.2
Reported patient safety behavior	1.7 (0.5)	1.6 (0.4)	1.6 (0.6)	0.7

5 point scale, 5=high level, 95% Confidence Interval. * Over 3 groups

Table 2: Reported stress and patient safety behavior during a 10-week rotation per group (n=68)

Table 3 presents an overview of examples of adverse events students reported during their rotation and their action(s) undertaken in response. On average, students reported 1.3 adverse events per person. The e-module-group reported fewer incidents (19%) than the control or game-group (±40%). Medication and diagnosis related adverse events were most often reported. Students took action in half of the reported events.

Learning time and game data

Self-reported average learning time for the game-group was 2.9 hours (SD=1.1h) and for the e-module-group 0.9 hours (SD=1.0h, p<0.001). There was no association between learning time and outcome measures.

Due to the fact that the game did not yet have an “auto-save”-function, we only had log-data from 18 students. For this group, average game-time was 3.3 hours (SD=1.6h). Logged game-time correlated with self-reported learning time (r=0.77, p=0.001). Students spent the majority of the game-time doing missions (M=2 h, SD=0.4), followed by breathing exercises (M=0.7 h, SD=0.3) and video-lectures (M=0.7 h, SD=0.4). There was some association between knowledge scores and game-time (r=0.69, p=0.02), in particular with video-lecture points (r=0.84, p<0.001). Game-time was also associated with patient-care mission points (r=0.87, p<0.001), indicating a learning effect within the game. Game-time and patient-care points correlated with self-efficacy (r=0.51/0.50, p=0.04) and negatively with reported stress during rotation (game-time: r=-0.65, p=0.02; patient care points: r=-0.78, p=0.002). This indicates students who had successfully

Type of adverse events reported by students (n = 86)		***Action(s) undertaken as reported by the students (n=38)***	
	n		n
Control group	36 (41%)	Control group	15 (39%)
E-module group	16 (19%)	E-module group	11 (29%)
Game group	34 (40%)	Game group	12 (32%)
Total	86	Total	38
Examples per category			
Medication & Fluid (n = 30) • Incorrect dose of medication (too high/low) • Too much fluid resuscitation in case of cardiac compromised patients • Withdrawal of drugs due to unclear drugs prescription system • Prescription of wrong medication • Ignoring drug interaction		Medication & Fluid (n = 11) • Pointing out a drug interaction to the resident • Corrected the amount of fluids given during an operation (to prevent acute heart failure)	
Diagnosis & diagnostics (n=24) • No precautions in case of contrast allergy • Lab investigations (e.g. blood sample) for wrong patient • Too late request for lab investigations		Diagnosis & diagnostics (n=9) • Telling the resident he had to order the lab for today instead of requesting it for one day later, because there was an emergent indication	
Communication (n=5) • Changes of therapy not communicated to the nurse • No clear handover of therapy changes due to use of verbal instead of written communication resulting in inadvertent closure of a nephrostomy catheter.		Communication (n=5) • Not clear whether action was undertaken	
Remaining events (n = 26) • Lack of adherence to hygiene measures • Already reported mistakes by the official procedure (MIP), unclear what exact adverse event was		Remaining events (n =13) • Speak up to the resident internal medicine about not washing hands after patient contact and touching feces • Making report about the incidence (MIP)	

Table 3: Overview of adverse events reported by students during rotation and related actions undertaken (n=68)

treated virtual patients and played longer were more self-confident in patient safety issues and experienced less stress. There was no association between biofeedback or video-lecture points and reported stress, indicating an independent effect of the patient missions.

Evaluation of the game and the e-module

Factor-analysis showed that three constructs could be identified within the 23-item evaluation scale: the format is *educational* to learn on patient safety (e.g. 'the game/e-module helps me to be more effective at patient safety', 9 items), the format is *engaging* (e.g. 'when playing the game/doing e-module, I felt actively involved', 9 items) and the

format is *easy-to-use* ('I quickly became skillful with the game/e-module', 4 items). Reliability of scales was high to good (α=0.93/0.92/0.74 respectively).
The serious game was evaluated as more engaging than the e-module, with very large effect sizes; the e-module was evaluated as more easy to use, with a medium effect size (Figure 3). Although the game was evaluated as slightly more educational than the e-module, the difference was not statistically significant.
Most students responded to the request to name a few positive and/or negative aspects of the e-module or game. In summary, the game was perceived as 'fun to do', with interesting topics; in particular the patient-missions were appreciated for training in decision making, offering challenge, creating awareness, and learning to deal with stress ("time flies", "the missions generate stress so you become more aware of what is coming").
The video-lectures were evaluated as interesting, but it took too long before the player could go to the fun part (the missions), and after a while they became boring ("hard to keep concentration", "not interactive"). The breathing exercises were evaluated as useful but less realistic, they often took too long and did not always work well ("exercises do not match reality"). The e-module was evaluated as dealing with important topics and educational, but boring and with too little interaction ("too much text on one page", "not interactive", "no examples").

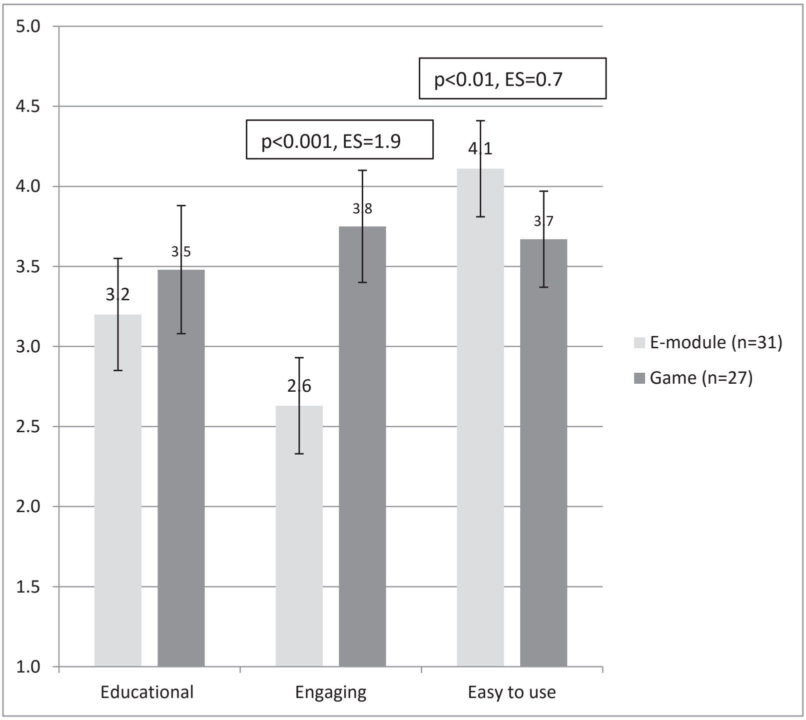


Figure 3: Evaluation of the e-module and game on three dimensions (ES=Effect Size)

The interviews confirmed that the missions were the most motivating part: educational and realistic, with an attractive storyline. One particularly well-appreciated part of the game was the 'red flag game': an interactive exercise with clinical situations depicting 'red flags' in communication and team performance that should be recognized as specific threats to patient safety (with a score and feedback). Several students mentioned that this is a good format and they would remember the messages. The scoring-system in general motivated some students to keep going in the game.

DISCUSSION

We allocated fourth-year medical students randomly to educational patient-safety content either in the form of a serious game or an e-learning module. Compared to students in a control-group who had received no instruction, patient-safety knowledge improved both after studying with the game and the text-based e-module, with large (game) to medium (e-module) effect sizes. Students who spent more time on the video-lectures in the game demonstrated more patient-safety knowledge. Our first hypothesis that patient-safety knowledge would equally improve using video-lectures (in the game) or text-based lectures (in the e-module), compared to no extra education in the control group was thus confirmed. Video-lectures from experts can be inspiring, but text is more flexible to read and can be adjusted to the readers' interests. This is consistent with results from media-research indicating that information can be presented in a variety of media formats with equal learning outcomes, but at very different costs and access[34].

The game-group spent considerably more time learning: three hours versus one hour in the e-module-group. During the rotation, students from all three groups reported low levels of stress and patient-safety awareness, with no differences between groups. Our second hypothesis that perceived stress levels during rotation would be lower in the game-group (due to the biofeedback) than in the other groups had thus to be rejected. An important possible explanation is the low level of experienced stress by all students, who are at a very early career stage and have not yet experienced the burden of clinical responsibility. Students sometimes reported stress during rotations, but this appeared to be more related to how they felt they must present themselves to medical staff[35]. It would be interesting to perform a follow-up study in first-year residents who experience higher levels of responsibility and possibly work-related stress. A comparison could be made between the biofeedback exercises in the game and an audio-based mindfulness training which has shown to be effective in reducing stress levels in medical students[36]. Interestingly, students who had treated patients successfully during game-missions, reported higher levels of self-efficacy and experienced less stress during their rotations.

However, since the game group as a whole did not show better self-efficacy or less stress, this may be a result of more motivated students playing longer and being more self-confident, unrelated to the intervention.

Our third hypothesis, that self-efficacy and patient-safety awareness would be higher in the game-group than in the e-module or control-group, had to be rejected. Both self-efficacy and self-reported awareness on patient-safety was low and did not differ between groups. The low self-efficacy levels of students before their first rotation can be explained by the absence of clinical experience; it is hard for students to estimate how they will be able to deal with stress and communication. Developing patient-safety awareness and changing behavior are complex processes. Although students from the game-group spent most of their game-time treating virtual patients and doing exercises on communication and teamwork, it is a large step from this experience to behavior in actual clinical practice. A systematic review on patient-safety education showed only a few reports of positive outcomes on higher-level skills and attitudes[7,39]. More knowledge on effective design to develop awareness and assertive behavior is needed. Probably a blended learning model, including team scenario training, would be beneficial.

The reports on adverse events and actions by our students underline the need to do so: during their first 10-week-rotation, each student reported on average almost 1,5 adverse events; only half of these students undertook action. The game-group reported somewhat more adverse events than the e-module-group, as expected, but not more than the control-group. This may be explained by the fact that patient-safety education in our institution changed somewhat. During data-collection for the control-group, the focus was more on the human-factors approach. During data-collection for the intervention-groups, education was less recognizable as separately oriented on patient-safety and more on effective and open communication in general. This may have caused a greater awareness of adverse events among control-group students.

The observation that one group of students does not profit from the patient-missions does not mean that the game is not effective for other groups, for example residents. In an earlier study on the effectiveness of a case-based simulation game in emergency care we found that although the game was effective for residents, it was too difficult for students and only residents showed improved skills[40].

Our fourth hypothesis that students in the game-group would be more motivated to learn was confirmed. Students evaluated the serious game as considerably more engaging than the e-module. The e-module was evaluated as more easy-to-use. Particularly the patient-missions in the game, with their realistic scenarios and challenging tasks, and exercises with clinical patient-safety situations, were experienced as very engaging. This observation is in line with other research reporting that a majority of medical students, including many who do not regularly play video games, hold favorable views on its use in medical education to experience different clinical situations[41]. The importance of the notion of motivation for (medical) education has recently been recognized; several studies have shown that motivation influences learning and cognitive performance[42].

To the best of our knowledge, this is the first study on the comparative effectiveness of two modern digital formats on patient-safety learning outcomes and motivation.
Yet, our study has several limitations. This study was done with 4th-year medical students. Results cannot be generalized to other groups such as residents. As we used a historical comparison for the control-group, we did not have control over historic changes in education. Patient-safety education in the medical curriculum is in constant flux. As explained above, we would expect the control-group to be more sensitive to recognizing adverse events during their rotation than the intervention groups, implying this higher 'baseline' may have decreased the difference in patient-safety skills compared with the intervention groups. Results might have differed if the control-group was included in the second phase of the study.

The research study was conducted with students who have limited clinical responsibilities during their rotations. Non-technical skills such as recognizing and dealing with stress, communicating well during difficult circumstances are central learning goals of the game and were designed with the 'overwhelmed' young resident in mind. These issues are less relevant to medical students making their first entry into the clinic. Nonetheless we decided to first compare the effects of the game with another format in medical students, because it is hard to motivate busy residents to spend extra study time on different educational formats and tests without evidence for effectiveness. Moreover, the results of this study may serve as a baseline for future studies with residents.

Practical implications

The study results yield a number of practical implications for patient-safety education. First, patient-safety knowledge can be developed at low costs for large groups of trainees, using a combination of (interactive) video and text-based lectures. In addition, challenging interactive cases on patient-safety issues, offer an engaging and low-cost format for patient-safety education. Thirdly, although serious games are engaging for students and stimulate them to study longer on patient safety issues, more research is needed on how games can stimulate patient safety awareness and behavior.

Conclusion

This study showed that video-lectures from a game and text-based lectures from an e-module are equally effective in developing knowledge on patient-safety topics. Biofeedback exercises may be adequate with learners who perceive higher stress-levels, but future studies should define their optimal format. Although serious games, in which students can actively practice their clinical and communication skills in simulated scenarios, are strongly engaging and stimulate them to study longer, more research is needed into the effects of game-design features on learning outcomes. Online learning is still a young field, with the potential to provide flexible, motivating and scalable learning opportunities for patient safety education.

Funding and Competing Interest statement

This study was financed by the Erasmus University Medical Center Rotterdam, The Netherlands. The serious game 'Air Medic-Sky-1' was developed and is owned by the University Medical Center Utrecht, the Netherlands. The legal owner had no role in the data collection and analysis, decision to publish, or preparation of the manuscript. The game is currently not commercialized, but may be offered for sale or license in any form in the future.

REFERENCES

1. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA. 2003;290(14):1868–74.
2. Kohn L, Corrigan J, Donaldson M. To err is human: building a safer health system. The national academies. 2000. 453-454 p.
3. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann D a, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010;363:2124–34.
4. Aboumatar HJ, Thompson D, Wu A, Dawson P, Colbert J, Marsteller J, Kent P, Lubomski LH, Paine L, Pronovost P. Development and evaluation of a 3-day patient safety curriculum to advance knowledge, self-efficacy and system thinking among medical students. BMJ Qual Saf. 2012 May;21(5): 416–22.
5. Mankaka C, Waeber G, Gachoud D. Female residents experiencing medical errors in general internal medicine: a qualitative study. BMC Med Educ. 2014;14:140.
6. Weiner E, Nagel D. Human factors in aviation. New York: Academic Press; 1988.
7. Nie Y, Li L, Duan Y, Chen P, Barraclough BH, Zhang M, Li J. Patient safety education for undergraduate medical students: a systematic review. BMC Med Educ. BioMed Central Ltd; 2011 Jan;11(1):33.
8. Leotsakos A, Ardolino A, Cheung R, Barraclough B, Walton M. Educating future leaders in patient safety. J Multidiscip Healthc. 2014;Sept.
9. Greiner A, Knebel E. Health Professions Education: A Bridge to Quality (Free Executive Summary). Health Professions Education: A Bridge to Quality (Free Executive Summary). 2003. 192 p.
10. WHO WA for patient. WHO Patient Safety Curriculum guide for medical schools. 2009.
11. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008 Sep 10;300(10):1181–96.
12. Sitzmann T, Kraiger K, Stewart D, Wisher R. The comparative effectiveness of web-based and classsroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.
13. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med. 2006 Mar;81(3):207–12.
14. Dankbaar ME, Storm DJ, Teeuwen IC, Schuit SC. A blended design in acute care training: similar learning results, less training costs compared with a traditional format. Perspect Med Educ. 2014 Sep;3(4):289–99.
15. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek J, Wang A, Erwin P, Hamstra S. Technology-Enhanced Simulation for Health Professions Education; A systematic review and meta-analysis. JAMA. 2011;55905(September):978–88.
16. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ. 2010 Jan;44(1):50–63.
17. Westera W, Nadolski R, Hummel H, Wopereis I. Serious games for higher education: a framework for reducing design complexity. J Comput Assist Learn. 2008;24(5):420–32.
18. Kalkman CJ. Serious play in the virtual world: can we use games to train young doctors? J Grad Med Educ. 2012;4(March):11–3.

19. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. Elsevier Ltd; 2012 Sep;59(2): 661–86.

20. Sitzmann T. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. Pers Psychol. 2011 Jun 27;64(2):489–528.

21. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, Simeoni Z, Tran M, Yukhymenko M. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. Rev Educ Res. 2012 Feb 1;82(1):61–89.

22. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, Schünemann HJ. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach. 2010 Jan;32(1):16–27.

23. Kirkpatrick DI. Evaluating Training Programs: The four levels. 2nd Editio. Berett-Koehler, editor. San Fransisco; 1998.

24. Yardley S, Dornan T. Kirkpatrick's levels and education "evidence". Med Educ. 2012 Jan;46(1): 97–106.

25. Turner NM, Van De Leemput AJ, Draaisma JMT, Oosterveld P, Ten Cate OTJ. Validity of the visual analogue scale as an instrument to measure self-efficacy in resuscitation skills. Med Educ. 2008;42: 503–11.

26. Okuyama a., Martowirono K, Bijnen B. Assessing the patient safety competencies of healthcare professionals: a systematic review. BMJ Qual Saf. 2011;20:991–1000.

27. Cohen S, Kamarck T, Mermelstein R. Perceived Stress Scale. J Health Soc Behav. 1983;24:385–96.

28. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 1989;13(3):319–40.

29. Lund A. Measuring Usability with the USE Questionnaire. STC Usability SIG Newsl. 2001;8:2.

30. Dankbaar M, Bakhuys Roozeboom M, Oprins E, Rutten F, van Merrienboer J, van Saase J, Schuit S. Preparaing residents effectively for emergency care skills training with a serious game. Submitted.

31. Streiner D, Norman G. Health measurement scales. Oxford University Press; 2008.

32. Grissom R, Kim J. Effect sizes for Research - Univariate and multivariate applications. 2nd ed. New York: Routledge; 2012. 66-73 p.

33. Hojat M, Xu G. A Visitor's Guide to Effect Sizes. Adv Health Sci Educ Theory Pract. 2004;9:241–9.

34. Mayer RE. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005. 616 p.

35. Verdonk P, Räntzsch V, de Vries R, Houkes I. Show what you know and deal with stress yourself: a qualitative interview study of medical interns' perceptions of stress and gender. BMC Med Educ. 2014;14:96.

36. Warnecke E, Quinn S, Ogden K, Towle N, Nelson MR. A randomised controlled trial of the effects of mindfulness practice on medical student stress levels. Med Educ. 2011;45:381–8.

37. Allery LA. Educational games and structured experiences. Med Teach. 2004 Sep;26(6):504–5.

38. Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simul Gaming. 2002 Dec 1;33(4):441–67.

39. Gordon M. Non-technical skills training to enhance patient safety. Clin Teach. 2013 Jul;10(3): 170–5.

40. Dankbaar M, Alsma J, Jansen E, van Merrienboer J, van Saase L, Schuit S. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. Adv Health Sci Educ Theory Pract. 2015 DOI: 10.1007/s10459-015-9641-x.

41. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ. 2010 Jan;10:50.

42. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach. 2014 May;36(5):371–84.

APPENDIX 1. QUESTIONNAIRE ON SELF-EFFICACY (12 ITEMS)

Please indicate how you assess your ability to …

1 Perform a debriefing within a team, e.g. before an operation	1-100
2 Hand over patient information	1-100
3 Recognize signals of threats to patient safety during teamwork	1-100
4 Conduct a debriefing after a team task	1-100
5 Raise issues of threat to safety of a patient among the medical staff	1-100
6 Physically reduce high levels of stress in yourself	1-100
7 Recognize signs of depression in yourself	1-100
8 Recognize signs of depression in colleagues	1-100
9 Recognize signs of sleep deprivation in yourself	1-100
10 Recognize signs of sleep deprivation in colleagues	1-100
11 Focus on one important task when dealing with multiple things at a time	1-100
12 Approach a senior staff member personally for his or her negligence	1-100

Item 1-6: self-efficacy on communication; item 7-12: self-efficacy on recognition of patient safety threats

APPENDIX 2. QUESTIONNAIRE ON PERCEIVED STRESS AND PATIENT SAFETY BEHAVIOUR

Perceived Stress Scale questionnaire

1 = Never 2 = Almost Never 3 = Sometimes 4 = Fairly Often 5 = Very Often

	1	2	3	4	5
1. During the last week, how often have you felt nervous and “stressed” in your clerkship?					
2. During the last week, how often have you found that you could not cope with all the things that you had to do during clerkship?					
3. During the last week, have you felt you were in control during your clerkship?					

Patient Safety behaviour

The next three questions are related to 'adverse events'. An ‘adverse event’ is a (possibly) negative outcome resulting from a medical intervention, which is not due to the underlying condition of the patient. Examples are violations of rules, near misses and mistakes.

1 = Never 2 = Almost Never 3 = Sometimes 4 = Fairly Often 5 = Very Often

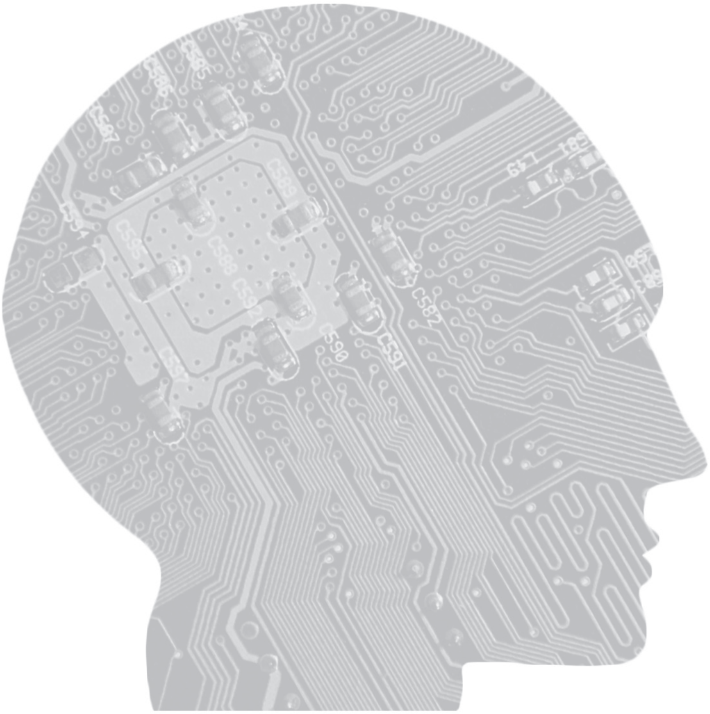
	1	2	3	4	5
4. Have you seen/experienced any adverse events during the last week?					
5. Have you taken any action in response to this event/ these events?					
6. (if question 5 was answered affirmatively): Can you give an example of such action?					

APPENDIX 3. EVALUATION QUESTIONNAIRE ON GAME AND E-MODULE

Original Q*	**Educational (to learn on Patient Safety, 9 items)**	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
Use 1	The game/e-module helps me to be more effective at patient safety					
Use 2	The game/e-module is useful for patient safety education					
Use 3	The game/e-module gives me control on patient safety activities during my clerkship					
Use 4	The game/e-module meets my needs					
Engage 4	The content of the game/e-module is educational					
Satis 1	I am satisfied with the game/e-module					
Satis 2	I would recommend the game/e-module to a fellow students					
Attitude 1	It is a good idea to use the game/e-module for my study					
Moti 4	I try to learn more about patient safety to be a better doctor in the future					
	Engaging (9 items)					
Engage 1	When playing/doing the game/e-module, my attention was entirely on the game					
Engage 2	When playing/doing the game/e-module, I felt actively involved					
Engage 3	I could concentrate fully throughout the game/ e-module					
Satis 3	The game/e-module is fun to use					
Satis 4	The game/e-module is pleasant to use					
Attitude 2	I like the idea of using a game/e-module to learn about patient safety topics					
Attitude 3	Overall, I enjoyed using the game/e-module					
Attitude 4	I would have liked to continue playing the game/e-module					
Moti 2	I like what we are learning in this game/e-module					
	Ease of use (4 items)					
Ease 1	The game/e-module is easy to use					
Ease 2	Learning to operate the game/e-module initially is full of problems					
Ease 3	I easily remember how to use the game/e-module					
Ease 4	I quickly became skilful with the game/e-module					

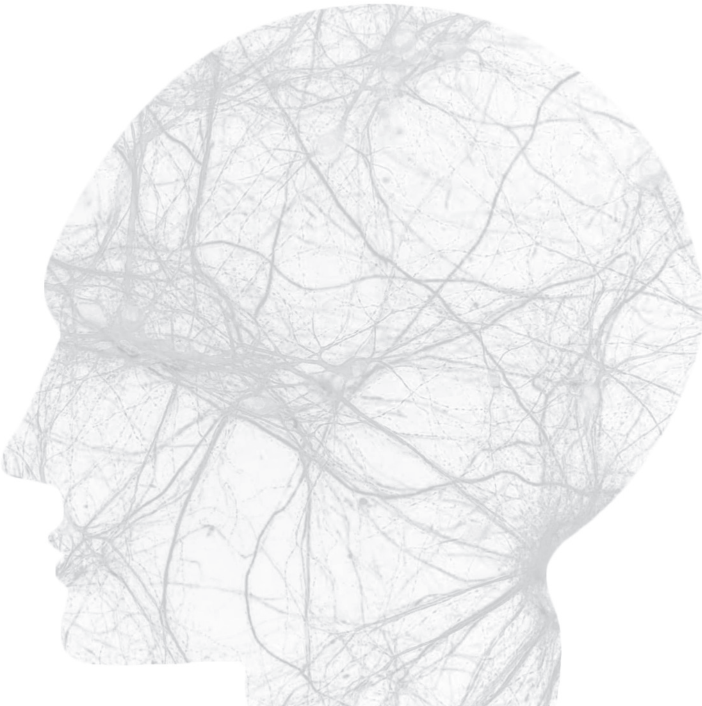
What do you feel as the most positive aspect(s) of the game/ e-module?
What do you feel as the most negative aspect(s) of the game/ e-module?

*Codes in left column indicate the questionnaire from which the item is originating (questionnaire on usefulness, satisfaction, ease of use, engagement, attitude towards using or motivation



Chapter 7

How to systematically assess serious games applied to health care



Maurits Graafland
Mary EW Dankbaar
Agali Mert
Joep Lagro
Laura D de Wit-Zuurendonk
Stephanie CE Schuit
Alma Schaafstal
Marlies P Schijven

JMIR Serious games 2014; 2(2):e1

ABSTRACT

Background

Serious or *applied games* are digital games with the purpose to improve individual's knowledge, skills, or attitudes in the "real" world. Serious games applied to medical or health-related purposes are growing rapidly in numbers and in types of applications. Serious games have been shown to be at least as effective as conventional tests in improving cognitive abilities in the elderly[1] and even more effective than conventional neuropsychological interventions when it comes to improving neuropsychological abilities of alcoholic patients[2]. Serious game-based interventions have been used to support rehabilitation in disabled patients, showing equal effectiveness compared to conventional training programs[3]. Games have been applied to promote healthy behavior in children[4] and educate patients[5,6]. Serious game-based patient education has also been shown to increase the treatment adherence among adolescents with leukemia[7]. A third application for serious games is training medical personnel[8]. Serious games have been shown to add to Advanced Life Support training[9] and improve understanding of geriatrics principles among medical students compared to conventional training methods[10]. Patients[11], students, and professionals[12] generally view game-based interventions as fun and challenging.

Although results for serious games in terms of effectiveness for such purposes are promising, their implementation as "serious" modalities for prevention, treatment, or training in health care is hindered by lack of understanding of the underlying concepts among health care professionals, or even distrust. Before doctors and patients consider using serious games as a useful solution for a health care-related problem, it is important that they understand what problem is being addressed by the game and that a proposed claim on effectiveness is indeed trustworthy. Many clinicians are currently undereducated in judging a serious game's safety or effectiveness. Information on individual games is often hard to find in disorganized app stores and websites[13]. Studies on serious games' validity and effectiveness remain scarce[8,14]. The idea of applying a video game in health care may even be resentful to certain clinicians or patients. In addition, threats to data safety fuel distrust towards electronic applications in health care altogether[15]. Such issues menace the practical application of serious games throughout health care, subsequently limiting investments in smart solutions that may actually prove beneficial in the end.

This article discusses the first tool for the systematic assessment of serious games applied to medical use, for educators and clinicians. The information collected and organized accordingly, will aid health care practitioners to understand and appraise the risks and benefits of specific serious games in health care in a uniform manner.

ASSESSMENT FRAMEWORK

To our knowledge, there is currently no systematic framework for the assessment of serious games in health care described in literature. Therefore, the Dutch Society for Simulation in Healthcare (DSSH)[16] has developed a consensus-based framework, categorizing important items that assess a serious game's safety and validity. Eight individuals (see Acknowledgements section for details) from six different institutions experienced in designing, applying, or researching serious games for health care-related purposes participated. The reporting standards for *non*-game mobile health apps for medical purposes (mHealth), published by Lewis[17] and Albrecht[18], was used as a basis. This system is applied by the peer-reviewed mHealth app assessment initiative of the *Journal of Medical Internet Research*[19]. Due to inherent differences in the functionality of *games* compared to purely informational *mHealth* applications, this framework required re-evaluation.

The panel reviewed the items from these reporting standards during two meetings. All items in the Albrecht framework[18] were systematically evaluated. For each of the 5 categories, items irrelevant to serious games were removed and if necessary, extra items were added. During the second panel meeting, the framework was re-evaluated and all members approved the final version.

The framework described provides 62 items in 5 main themes (Table 1), aimed at assessing a serious game's rationale, functionality, validity, and data safety. It specifically does *not* aim to assess its effectiveness in terms of success or user attractiveness. The panel defined serious games (other than a regular medical application) as digital applications instigating a specific behavioral change to its user, in the form of skills, knowledge, or attitudes useful to reality[20]. The framework does therefore *not* apply to (mobile or Web-based) digital health apps with a purely informational purpose, for which the *mHealth* app assessment framework is designed[18].

ASSESSING MEDICAL SERIOUS GAMES

Game Description

When evaluating a specific serious game, it should be thoroughly described and registered (including information about the manufacturer or owner to whom the game should be attributed and the version). Equally to mobile applications, a special interest is taken into the owner's policy concerning revenues from sponsoring and advertisements, both during development as well as its use. Sources of revenue and affiliations

	Category	Item	Question
Game description	Meta-data	Operating system	Operating systems of the game
		Version	Version
		Web-link	Web-link
		Project type	Commercial, non-commercial, other
		Access	Public / restricted / other
		Adjunct devices	Is an adjunct device needed?
	Development	Funding	How was development funded? Eg, funding agencies, investors
	Sponsoring / Advertising	Advertisement policy	Is the game free of commercial pop-ups? If not, what is advertised?
		Sources of income	Are there sources of income within the game?
		Sources of income outside game	What are the sources of income of the owner/distributor?
	Potential conflicts of interest	Affiliations	What affiliations do the publishers have that could influence content or user group?
		Conflicts of interest	What interests do the publishers have that could influence the game's content or user group?
		Disclosure	Are conflicts of interest disclosed?
Rationale	Purpose	Goal or purpose	What is (are) the purpose(s) of the game?
		Disclosure	Is (are) the purpose(s) disclosed to users?
	Medical device	Medical device	Is the serious game a medical device, or not?
		Class	If yes, which class?
		Approval by legal bodies	If yes, does it comply to the necessary requirements (FDA-ap- proval, CE-mark?).
	User group	Specific user groups	For each user group: disease/condition, or health care profession.
		Description	Please specify gender, age (range), and other relevant descriptive items.
		Limits	Are there age limits, or other limits?
		Disclosure	Is the intended user group disclosed?
	Setting	Patient care	Is the game used in patient care?
		Training courses	Is the game used in training courses or -curricula?
		SCORM compliancy	If used in training courses or curricula, is the serious game SCORM-compliant
Functionality	Purposes / didactic features	For every purpose of the game:	
		Learning or behavioral goals	What content will the player learn?
		Relation learning and gameplay	How does the learning content relate to the gameplay?

	Category	Item	Question
		Instruction	What intervention leads to the learning transition (eg, tutorial, in- structions (in-game))
		Assessment (progress) in game	Through which parameters is progress in the game measured?
		Assessment parameters	Which parameters are to designers' opinion indicative for measuring learning effects?
	Content Management	Content Management system	Is the Content Management System restricted to specified persons or institutions?
		User uploaded content	If no, are users allowed to upload their own content?
		Content monitoring	How is uploaded content checked?
		Restrictions and limits of the serious game	Please describe restrictions and limits of the serious game. What content on the learning goals is not covered?
	Potentially undesirable effects	Potentially undesirable effects	What potential undesirable effects could the game have?
		Disclosure	Are such potential undesirable effects disclosed to the user?
		Measures taken	What measures are taken to prevent potential undesirable effects?
Validity	Design process	Medical expert complicity	Were medical experts (content experts) involved in the design process from the start?
		User group complicity	Were representatives from the user group involved in the design process from the start?
		Educationalist complicity	Were educationalists involved in the design process from the start?
	User testing	User testing	Did user testing take place? What were the results, and how were these incorporated in the design?
	Stability	Platform stability	Does the game produce the same results on different platforms?
	Validity (effectiveness)	Face validity	Do educators and trainees view it as a valid way of instruction?
		Content validity	How is its content validated to be complete, correct, and nothing but the intended medical construct?
		Construct validity	Is the game able to measure differences in skills it intends to measure?
		Concurrent validity	How does learning outcome compare to other methods assessing the same medical construct?
		Predictive validity	Does playing the game predict skills improvement in real life?

	Category	Item	Question
Data protection	Data protection and privacy	Data processing	How is data collected in the serious game?
		Patient privacy	Are patient-specific data stored in the game? If yes, are patient informed consent criteria met according to relevant national standards?
		Data ownership	Who owns and stores the data resulting from play?
		Data storage period	During what period are data stored?
		Data removal	Can the user delete data temporarily and/or permanently?
		Data storage security	Is the data storage secured in conformity with laws of the countries stated above?
		Data transmission security	Is the data transmission secured in conformity with laws of the countries stated above?
		Disclosure	Are all items on "data protection" disclosed to the user?

Table 1. Items relevant for the assessment of a serious game used for health care-related purposes.

(eg, pharmaceutical industry) may bias or threaten a serious game's validity for obvious reasons. These should be fully disclosed to the game's users. Sources of income within a game can be equally relevant to the costs required for the initial purchase.

Rationale

This clarifies the game's purpose outside the game. This external purpose (eg, improving eye-hand coordination in laparoscopic surgery) may differ from the actual goal *in* the game (eg, completing a quest in an underground world[21] or playing a tennis game[22]. This clearly differs from the Albrecht framework, because most mHealth apps have a single obvious purpose (internal goal = external goal). A game's purpose relates to the intended user group and the setting in which it is used, similar to mHealth apps.

Additionally, serious games might fall within the scope of the medical devices, requiring specific guidelines to be implemented, set by the US Food & Drug Administration (FDA), European Committee (Conformité Européenne, CE), or national equivalents. This specifically applies to games with a distinct diagnostic or therapeutic purpose. Moreover, integration of serious games into electronic learning environments may demand certain technical requirements. The industry has set standards to improve the interoperability of e-learning content (the Sharable Content Object Reference Model; SCORM)[23]. Its implementation will improve the integration of educative serious games in learning management software.

Functionality

Functionality of a serious game clearly differs from that of an mHealth app. These usually contain "dry" content (e.g., medical information) or an obvious functionality (e.g., communicating or registering information), whereas a game requires the user to operate or interact with the content, with the ultimate goal to change ones behavior in real life (i.e., learning). To understanding this process, information is required on the game's content, how the instruction is delivered, how performance is assessed and how these aspects are integrated in the gameplay[24,25].

Consequently, it is important to register information on the game's content management. For instance, users may be able to add content themselves, making content validation an important issue. This directly influences the game's content's validity.

Finally, undesired results or negative transfer of learning could occur in the interaction with a serious game, which is not the same concept as "gaming the game" (ie, cheating), an effect that may very well enhance learning[24]. If validation research is not present, at least a logical connection between game play and behavioral or learning goals should be present and disclosed by the developer.

Validity

Validity determines whether an instructional instrument (such as a serious game) adequately resembles the construct it aims to educate or measure. More formally, "the degree to which evidence and theory supports the interpretations of [game] scores entailed by the proposed use of [the game]"[26]. The American Psychological Association has set a series of standards to measure validity[26]. Whereas many validity types have been described, validity research in medical education usually contains several consequential phases[27,28]. First, experts should scrutinize the game's content to determine its legitimacy (*content* validity). Second, experts and novices judge the instrument's apparent similarity to the construct it attempts to represent (*face* validity). *Construct* validity reflects the ability of the instrument to actually measure what it intends to measure (ie, the difference in performance between groups of users with different levels of experience in reality). *Concurrent* validity reflects the correlation between performance on the serious game and their performance on an instrument believed to measure the same construct (eg, a simulator or course). The ultimate goal is to prove a game's *predictive* validity: does performance in the game lead to better outcomes in reality? Most validation research currently published in the medical domain uses these concepts[29]. For individual cases, relevance of specific validity types may differ. When considering mHealth apps in general, content validity may be the sole source of validity.

Validity research is frequently a long and costly enterprise. Many newly developed serious games have therefore not yet undergone validity research[8]. The framework therefore determines a number of steps to pre-assess a serious game's potential as a valid instrument, with regard to its design and initial testing phases. This encompasses the involvement of user groups, content experts, or educationalists in the design (if relevant to the game's purpose). Next, if a game has undergone user testing and stability testing, the game is more likely to have higher face- and content validity.

Data Protection

Threats to user privacy are imminent in electronic and mobile health apps, especially when patient-specific data are measured or entered in the game[15]. This considers data "at rest" on devices or servers, as well as data "in transit". It must be clear whether data is collected by the game, who owns the data and whether users can request to remove their data. Storage and analysis of personal data should be disclosed to users and must be in conformity with the laws applied in countries the serious game is distributed in. Special care must be taken if patient information is collected. These items are in general conformity with the requirements for mHealth apps described in Albrecht's framework[18].

DISCUSSION

When using serious games in health care, end users (clinicians, patients, or educators) must decide whether games are safe and effective enough to be used for their intended purposes. In order to do so, they need consistent, transparent, and reliable assessments. Are applied games really stating their claim in this field? In the framework described in this article, both developers and end users are supported in assessing relevance, validity, and data safety of an applied game. In order to become a "qualified game", developers should disclose comprehensive information on their products and claims. They must provide transparency to meet the standards. The *Journal of Medical Internet Research* and the Dutch Society for Simulation in Healthcare[16] have launched an international peer-reviewing initiative for serious games in health care.

The safe application of technology-enhanced solution remains the responsibility of the health care provider. Choosing if a serious game answers to the user's needs, can be based on information concerning 5 main areas described in this article. The majority of the items cannot be assessed using objective parameters. For instance, claiming a specific serious game's predictive validity should be supported by solid evidence. A comprehensive evaluation by a panel of experts in the form of a quality label could form a more practical solution.

Guidelines have been recently published reporting standards to support clinicians and patients in distinguishing high quality mhealth apps[17,18] and medical websites[30]. These standards form the basis for the framework described in this article. These standards have two important shortcomings when it comes to games. First, explicit information on a serious game's content and didactic features is required, as the external purpose of a serious game is frequently less obvious to the user than in the case of mHealth apps. Second, serious games require additional validation steps (e.g., construct and predictive validity), compared to non-interactive information platforms. Game play is dynamic and learning goals in game play are often not disclosed to the user. In fact, the user learns by playing the game, whereas discovery in itself may be part of the game play. Disclosing learning goals would thus be counterproductive.

There are several limitations to the framework described in this study. It considers validity of the serious game's content and its didactic functionality. Validity does not predict a game's success nor its attractiveness to the user, which also depend on its entertainment capability and distribution method[31]. It does not wish to objectify which game is most fun, but merely which game is most valid. A second consideration is that in the scientific field of validity research in medicine, validity concepts other than the one used in this framework have been proposed[32]. The "classical" validity concepts (content-, face-, construct-, concurrent-, and predictive validity) have been most frequently used in validity research in medicine and therefore the most logical to encompass in the framework presented in this article[27,28].

In summary, this consensus-based tool provides the end users the support required when assessing the effectiveness and relevance of serious games in health care. An FDA-approval or CE-mark is simply insufficient for this purpose. In order to prevent wrongful application and data theft of unsuspecting patients or medical students, this information on medical serious games should become publically available to all end users. This will aid the prescription of safe and effective games to patients and the implementation of games into educational programs.

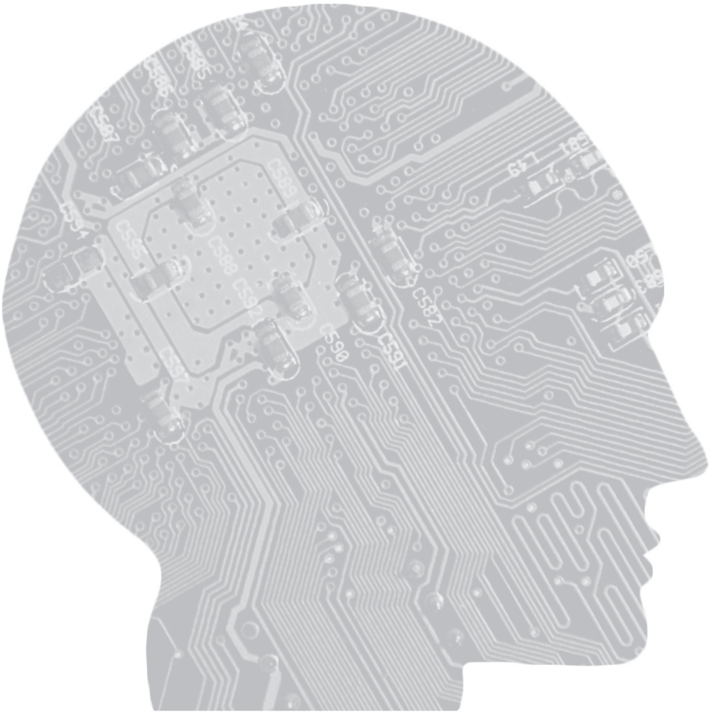
Acknowledgments

The authors received funding from the Patient safety project (grant ref PID 101060), of the *Pieken in de Delta*-program by the Dutch Ministry of Economic Affairs, Agriculture and Innovation, the city of Utrecht and the province of Utrecht, the Netherlands. The funding agency had no role in the preparation of the consensus tool, nor the preparation of the manuscript.

REFERENCES

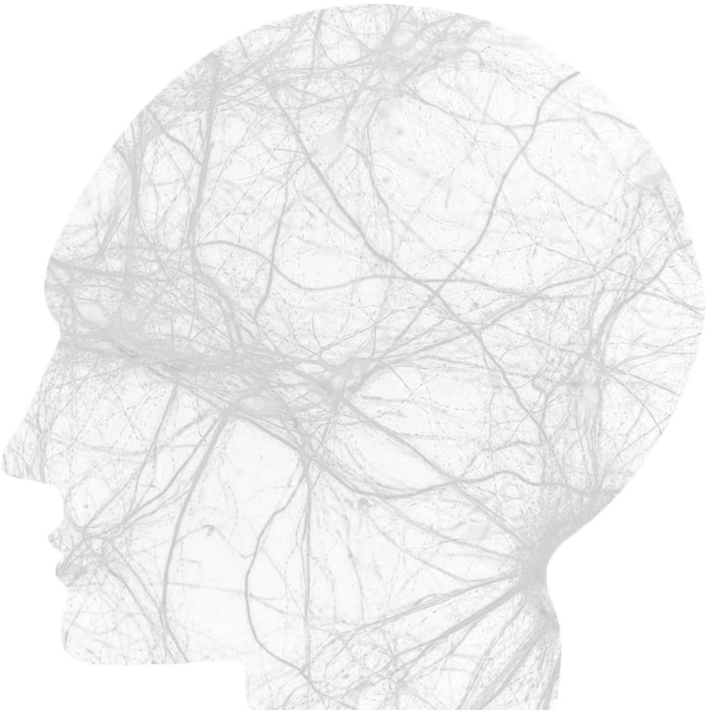
1. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 2. Gamito P, Oliveira J, Lopes P, Brito R, Morais D, Silva D, et al. Executive functioning in alcoholics following an mHealth cognitive stimulation program: randomized controlled trial. J Med Internet Res 2014;16(4):e102
3. Prange GB, Kottink AI, Buurke JH, Eckhardt MM, van Keulen-Rouweler BJ, Ribbers GM, et al. The Effect of Arm Support Combined With Rehabilitation Games on Upper-Extremity Function in Subacute Stroke: A Randomized Controlled Trial. Neurorehabil Neural Repair 2014 May 29.
4. Majumdar D, Koch PA, Lee H, Contento IR, Islas-Ramos AD, Fu D. "Creature-101": A Serious Game to Promote Energy Balance-Related Behaviors Among Middle School Adolescents. Games Health J 2013 Oct;2(5):280-290
5. Cooper H, Cooper J, Milton B. Technology-based approaches to patient education for young people living with diabetes: a systematic literature review. Pediatr Diabetes 2009 Nov;10(7): 474-483.
6. Read JL, Shortell SM. Interactive games to promote behavior change in prevention and treatment. JAMA 2011 Apr 27;305(16):1704-1705.
7. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008 Aug;122(2):e305-e317
8. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330.
9. Buttussi F, Pellis T, Cabas Vidani A, Pausler D, Carchietti E, Chittaro L. Evaluation of a 3D serious game for advanced life support retraining. Int J Med Inform 2013 Sep;82(9):798-809.
10. Lagro J, van de Pol MH, Laan A, Huijbregts-Verheyden FJ, Fluit LC, Olde Rikkert MG. A Randomized Controlled Trial on Teaching Geriatric Medical Decision Making and Cost Consciousness With the Serious Game GeriatriX. J Am Med Dir Assoc 2014 Jun 6.
11. Forsberg A, Nilsagård Y, Boström K. Perceptions of using videogames in rehabilitation: a dual perspective of people with multiple sclerosis and physiotherapists. Disabil Rehabil 2014 May 16: 1-7.
12. Graafland M, Vollebergh MF, Lagarde SM, van Haperen M, Bemelman WA, Schijven MP. A Serious Game Can Be a Valid Method to Train Clinical Decision-Making in Surgery. World J Surg 2014 Aug 27
13. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013;13:23
14. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 2012 Sep;59(2): 661-686.
15. Kotz D. A threat taxonomy for mHealth privacy. In: Crowcroft J, Manjunath D, Misra A. editors Third International Conference on Communication Systems and Networks (COMSNETS 2011).: IEEE; 2011 Presented at: Third International Conference on Communication Systems and Networks; 2011; New York, NY p. 1.
16. Dutch Society for Simulation in Healthcare. Official Website Internet. DSSH official website 2014: 2014.

17. Lewis TL. A systematic self-certification model for mobile medical apps. J Med Internet Res 2013;15(4):e89
18. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013;15(12):e277
19. Journal of Medical Internet Research (JMIR). 2014. Apps Peer-Review launched Internet URL: http://mhealth.jmir.org/announcement/view/67 [accessed 2014-11-04] [WebCite Cache ID 6TpSDiKIJ]
20. Michael DR, Chen S. Serious Games: Games That Educate, Train, and Inform. Boston, MA: Course Technology PTR;2006.
21. Jalink MB, Goris J, Heineman E, Pierie JP, ten Cate Hoedemaker HO. Construct and concurrent validity of a Nintendo Wii video game made for training basic laparoscopic skills. Surg Endosc 2014 Feb;28(2):537-542.
22. Badurdeen S, Abdul-Samad O, Story G, Wilson C, Down S, Harris A. Nintendo Wii video-gaming ability predicts laparoscopic skill. Surg Endosc 2010 Aug;24(8):1824-1828.
23. SCORM. SCORM Explained Internet URL: http://scorm.com/scorm-explained/ [accessed 2014-11-04]
24. Anderson LW, Krathwohl DR, Bloom Bs. A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. New York, NY: Longman; 2001.
25. Van Staalduinen JP, de Freitas S. A Game-Based Learning Framework: Linking Game DesignLearning Outcomes. In: Khine MS, editor. Learning to Play. New York, NY: Peter Lang Publishing, Inc; 2011:29.
26. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.
27. Schijven MP, Jakimowicz JJ. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005;14(4):244-246
28. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003 Oct;17(10):1525-1529
29. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract 2014 May;19(2):233-250
30. Health on the Net Foundation Code of Conduct (HONcode). Operational definition of the HONcode principles Internet URL: http://www.hon.ch/HONcode/Webmasters/Guidelines/guidelines.html [accessed 2014-11-04]
31. Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of Serious Games: An Overview. Advances inHuman-Computer Interaction 2013;2013:1-11
32. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006 Feb;119(2):166.e7-166.16.


C582
C591

Chapter 8

Discussion



In health care and medical education, with its exponential growth in knowledge and increasing demands on the needed competencies of doctors, there is a need for new, engaging and cost-effective training models[1]. Blended learning offers the opportunity to combine the advantages of online learning (flexibility and efficiency[2]) and classroom learning (social interaction and feedback). Different online formats can be combined with classroom learning. The use of blended learning is increasing quickly and a precipitous growth is predicted[3]. Some studies show blended learning is more effective than face-to-face learning[4,5], while others show no difference[6]. The **first aim** of this thesis was to investigate whether a blended training design is more effective or equally effective compared to classroom training for the acquisition of knowledge. We also compared costs of both designs.

Online learning, in a blended course or in a completely online course, can be used to improve the access, efficiency and effectiveness of education and training[1,2]. One of the challenges of designing online learning is to engage students during self-study. Serious games are potentially powerful new educational formats for training complex skills, while motivating learners through direct experiences[7]. Since little is yet known on the effectiveness and critical design features of serious games[8–10], the **second aim** of this thesis was to investigate the effectiveness and critical design features of serious games for performance, specifically training complex cognitive skills and attitudes, while enhancing motivation.

The blended training design we studied was a post-graduate course on emergency care for nurses. The two serious games that were studied were both originally developed for medical residents: one on emergency care (abcde*SIM*) and one on patient safety (Air Medic Sky-1).

In this chapter, we will first summarize the main conclusions from the six studies reported in this thesis with the implications for designing blended learning and serious games. Subsequently, the strengths and limitations of this thesis and the recommendations for future research are stated. We will conclude with the key points from this thesis.

CONCLUSIONS

We started with a study comparing the effectiveness of a blended design in postgraduate training for nurses with a conventional classroom training (Q1). The blended course is composed of one-third online and two-third classroom training. We evaluated training costs of both designs a well. Participants from the blended and conventional groups showed equal knowledge test results and were equally satisfied with the training, the blended group being more self-confident than the conventional group. Training costs of the blended design were two-third of the conventional design. We concluded that blended learning, combining online self-study with classroom training, is equally effective and attractive as classroom training. In post-graduate training, a blended design can lead to a significant reduction in costs.

It is essential to use validated assessment instruments to evaluate the efficacy of novel educational interventions. Therefore, our second study was a validation study on commonly used formats in the assessment of cognitive emergency care skills (Q2). Checklists have poor validity and reliability in assessing complex emergency care skills. In contrast to the intuitive attractiveness of detailed checklists, more global assessment instruments (such as competency scales or global performance scales) have a better reliability and validity. The limited number of assessment scenarios and raters, as typically used in emergency care assessment around the world, results in moderate validity and reliability.
In our third study we investigated the effectiveness of the abcde*SIM* game in training complex emergency care skills (Q3), using the more global assessment instruments. We investigated whether residents who use a serious game as an additional preparation for classroom training, show better performance on skills before training than a control group that only used a course manual for preparation. We found that residents who used the game spent more time on task (+2.5 hours), felt engaged with the game and showed better clinical cognitive skills than the control group. Both groups were equally motivated for the classroom training. After two weeks of training, no differences between the groups existed. A high-fidelity serious game can be used as an effective preparation for instructor-led emergency care courses, to train complex cognitive skills for residents. As this study was inconclusive on the question of which game-features had been responsible for the learning effects, in the next study we investigated the effect of fidelity in cases on the acquisition of skills.

Our fourth study consisted of a three-group randomized design on the effects of adding high-fidelity patient cases (the abcde*SIM* game) versus adding low-fidelity (text-based) cases to an instructional e-module on cognitive emergency skills, motivation and cognitive load of 4th-year medical students (Q4). The control group only studied the e-module, which includes information on the ABCDE approach and a demonstration video. Learning time was two hours longer for both the game group and the cases group compared to the control group, however acquired cognitive skills did not differ between groups. Students who spent more time on the game showed better skills than students who spent less time on the game, indicating the game was probably distracting for novices in the beginning and, if at all, only facilitated learning after longer play. We especially expect that the physical fidelity (tools and sounds in the game) provided too many details, creating cognitive overload. Only after playing longer students got used to the physical fidelity and were able to profit from the functional fidelity in the game. The game group experienced higher intrinsic and germane cognitive load and felt more engaged with the format than the cases group. For initial skills acquisition, an e-module with a demonstration video is a powerful instructional format. Additional open cases stimulate students to study longer, but do not result in learning when these tasks are too complex. Although high-fidelity patient cases in games enhance motivation and cognitive effort, they also increase complexity, can easily distract novice students and impede learning.

Our next study investigated whether students develop better patient safety knowledge and awareness and are more motivated after using a highly interactive serious game, in comparison with a limited interactive e-module on the same topic (Q5). A control group received no extra instruction. The serious game (Air Medic Sky-1) includes video lectures, interactive patient cases and biofeedback exercises. The e-module offers text-based information on the same topics, but with limited interactivity. Average learning time was three hours for the game group and one hour for the e-module group. Video lectures (in the game) and text-based lectures were equally effective in developing knowledge on patient safety topics. The game was evaluated as considerably more engaging than the e-module. During clinical rotations, students in all three groups reported low and similar levels of patient-safety awareness and stress. Patient safety knowledge can be gained through either video-lectures or text-based lectures. Although serious games on patient safety are strongly engaging for students and stimulate them to study longer, they do not necessarily result in better patient safety awareness.

The last contribution on serious games described a general framework for systematically evaluating serious games applied to health care and medical training (Q6). The framework provides a set of standardized criteria to evaluate the rationale behind a serious game's (e.g., goals and user groups), its content and didactic functionality (e.g., the relation between learning, game play and game assessment parameters), effectiveness (e.g., predictive validity of game outcomes on real-life performance) and data safety (e.g., data ownership). Using this framework, developers support end users in evaluating the usability and expected effectiveness of a specific serious game. We expect this will stimulate a thoughtful and safe implementation of serious games in medical education, based on its added value in a specific educational situation.

IMPLICATIONS FOR INSTRUCTIONAL DESIGN

In this section the implications for designing blended learning and serious games that followed from our studies will be discussed, in the context of educational theory and existing research. We will start with the question of effectiveness and design of blended learning (related to the first aim of this thesis), thereafter the effectiveness and design of serious games will be discussed, specifically for performance and motivation (related to the second aim of this thesis).

Blended and online learning

Blended learning can be summarized as being at least as effective as classroom learning[4–6,11–14]. The cost reduction we reported is mainly due to savings in indirect training costs (less absence costs). In blended learning in undergraduate education, classroom time is usually not reduced, but dedicated to other learning goals (e.g., more interaction). Therefore, reduction of training costs primarily exists in post-graduate education, where part of the training time is replaced with self-study. A risk of shifting class time to self-study time is making the course more demanding for participants, which may reduce their motivation[12]. This is an important consideration in the design. Simply reducing class time and adding online material does not result in optimal learning outcomes; an integrated (re)design of both components is needed[15]. Although the initial investment in developing a blended course is considerable, for repeated trainings or large groups of learners it is usually worth the investment. A variety of combinations of classroom learning and online learning are possible, depending on the learning goals, context and learner characteristics.

Different online formats which we used in our studies were effective for gaining knowledge and can be used in a blended design: e-modules containing text, exercises with feedback and demonstration video's; video-lectures and clinical cases with questions and feedback. This is in line with research stating that information can be presented in a variety of formats with equal outcomes but different costs[1,2,16–18]. On the one hand, our study (Q5) showed that video-based or text-based information, without interactivity, is experienced by students as boring; on the other hand, knowledge test results showed both formats enhanced learning. Although learner preferences are typically uncorrelated or even negatively correlated to learning outcomes[19], feedback is a key didactic feature in both online and classroom learning, as it enables students to identify their own strengths and weaknesses in mastering knowledge or skills[12,16]. In addition, self-assessment questions with feedback have been shown to stimulate learning[20], hence the knowledge test results may have been better still if the information was offered in an interactive way. An e-module including a demonstration video of a complex skill (providing visualization), also appeared to be very effective and was highly appreciated by novice students (Q4). This is in agreement with research findings showing that for initial skill acquisition, studying worked examples is superior to actively solving training problems or studying abstract principles. Effectiveness of worked examples can be attributed to the fact that this imposes less cognitive load on the novice learner and enables schema construction[21].

The development of a serious game requires considerable resources; for knowledge-oriented learning goals there is in general little added value of this format. Interactive text or video lectures can well be used to explain concepts or procedures at lower costs.

A critical element in a blended design is finding a good strategy to secure online preparation by learners, so that classroom time may benefit from this preparation. One strategy is imposing a minimal test score as a condition for participating in the classroom session; another strategy is to have students prepare, individually or as a group, assignments that are presented in class. Collaborative learning can also be built into a blended design and can strengthen motivation and self-regulation[11]. Collaborative learning theory provides a good basis for further research and development in group work and blended learning[22]. Finally, workload and support for teachers to develop these new blended courses has to be considered. This requires embedding the use of technology within a wider institutional strategy for learning and instruction, providing training and support by instructional designers[2].

Blended learning designs can also be directed to the development of skills, e.g. using a simulation program or serious game as a preparation for instructor-led skills training. As we showed, residents' working on a serious game resulted in improved complex cognitive skills before training (Q3). We expect training time can be reduced, which would make the blended design more cost-effective. A study in Advanced Life Support training, comparing the effectiveness of a blended design (one day training supplemented with online interactive cases) with a two-days course, showed the same skill level for both delivery formats, with reduced costs in the blended design[23]. Further research is needed to confirm our assumption. Training time may also be maintained, dedicating extra time to enhance the skills level at the end of training or for training other (e.g. communication) skills.

In conclusion a blended learning design is equally effective as classroom training. Advantages of blended designs are: adaptability to the learner's knowledge level and pace, flexibility of access in time and place and scalability of learning. In post-graduate education, blended learning can also save training costs. A variety of effective didactic features can be used, depending on the learning goals: presentation of information with exercises and feedback, interactive cases, demonstration videos. An important consideration in the design is the amount of online self-study. Securing online preparation is a critical success factor in blended learning. Theories of multimedia learning and collaborative learning are useful as a basis for further research.

Serious games and performance

In our validation study on commonly used formats in the assessment of complex cognitive skills – as targeted by the abcde*SIM* game - we concluded that more global assessment instruments have a better reliability and validity than detailed checklists. After preparing training in emergency care skills with an additional serious game, residents showed improved clinical skills with limited self-study-time, compared to residents who studied course material only. This game contained patient cases with feedback in a high-fidelity environment. Game-based learning can be associated with *task-centered models of learning* complex skills, such as the cognitive apprenticeship model or the four-component instructional design model (4C/ID). Research based on these theories has shown that in order to learn with tasks effectively and facilitating transfer to practice, a balance should be found in offering worked cases (examples, demonstrations) and open cases in a variety of problems. Support and guidance of learners is another important element[24–26]. In contrast, medical students did not profit from additional work on this game, compared to what a similar group learned from an e-module alone. Both residents and students were motivated to work with the game.

Our combined results from the two studies show that the same game can effectively improve complex cognitive skills for residents, but not for inexperienced students. These findings indicate an '*expertise reversal effect*', where open or worked cases interact with expertise level in effectiveness[27]. Novices in general profit more from example-based, tutorial conditions, whereas for more experienced learners in a certain domain, the superiority of worked cases disappears (and may even inhibit learning) and additional learning is facilitated by problem solving in a rich learning environment[28]. Discovery learning, with little or no guidance, may only be beneficial for more experienced students; for novices a more guided form of discovery learning is preferable, with instructional support and modelling, as found in research on simulations[29,30] and games[31]. Although we offered students an instructional e-module as a preparation for the serious game, the cognitive gap between this instruction and the open cases apparently was too large. Despite the extra learning time and cognitive involvement, skills do not improve when the learning tasks are too complex. High-fidelity presentation of cases in games may increase complexity even more, but also enhance motivation and cognitive effort. Simulation games, in particular the high physical fidelity it contains, can easily distract novice students, creating cognitive overload and impeding learning. Learning task fidelity should only gradually increase as learners become more proficient[32].

Students who worked with the game on patient safety did not show better patient awareness than students who worked on text-based information or a control group (Q5). Developing patient-safety awareness and team skills is a complex process; it's a large step from treating virtual patients and doing communication exercises in a game, to behavior change regarding adverse events in clinical practice. So far, only a few reports on successful training of patient safety and team skills exist[33]; often scenario-based team training is recommended[34]. A blended design, offering this type of game in adjunct with scenario-based team training may be needed.

Serious games and motivation

Motivation is recognized as an important component of learning, which influences cognitive performance[27], although it is often not considered in current approaches to medical education[35,36]. Our studies showed that two types of motivation should be distinguished: motivation to learn a task and motivation to work with a particular instructional format. Residents who used the game felt engaged with the format, but were just as motivated for the emergency care training (the task) as residents who did not use the game. Obviously, both groups of residents were highly motivated to develop their skills, as the training was just prior to their 6 months-internship in an Emergency Department. Medical students working on this game felt more engaged and invested more mental effort than students who worked on the same cases in a text-based format; however both groups spent the same amount of learning time – hence they seemed to be equally motivated for the task. The game on patient safety was also evaluated as considerably more engaging than the e-module, especially the patient cases. Which game features enhance motivation for trainees?

A huge number of studies on serious games discuss the proposed features that engage learners, but relatively few studies describe empirical research. As a result there is little consensus on the main elements which influence motivation and learning[31,37–41]. Features that are mentioned as relevant for engaging learners in games are: contextual information (a story or theme, rules, goals), challenges, feedback and instructions, control and interaction (with a program or peers)[39,42,43]. Learner processes which are associated with motivation include enjoyment, depth of involvement, engagement and task persistence; often described as 'flow': the state in which learners are so involved in an activity that nothing else seems to matter[44]. Flow derives from activities that are optimally challenging, in which the goals and feedback are clear, concentration is focused with a high level of cognitive effort, a high degree of control, and users are absorbed to the extent that they lose a sense of time and self[37]. The challenges a player faces in the game should be closely matched to his or her skill level in order to facilitate the flow experience[7,44]. This can be associated with Vygotsky's zone of proximal development, indicating that if the challenge is significantly lower than the learner would be able to handle (given his or her skill level), the learner might feel bored, while with too much challenge the learner might feel anxiety[45]. The 'flow channel' can be extended by providing some guidance or help from others. In order to keep learners in a flow state, challenges should increase with the progress of learners' skill level[7]. In contrast with views that game-based learning is to be considered as playful learning with low mental effort[46], our findings are in line with the theory of flow and cognitive load that games can stimulate both cognitive effort and engagement. Research findings on simulations point in the same direction. High-fidelity simulations stimulate motivation, learning and transfer because of their similarity with the performance environment (the 'yang'); the increased cognitive load however may reduce initial learning for novice learners (the 'yin')[47]. Challenging tasks and fidelity in games and simulations are motivating for learners, but too many motivational and cognitive stimuli may impede learning[48]. In order to maintain the benefits of high-fidelity, scaffolding and guidance can be used[49]. A study in which students were allowed to progress in simulated tasks of increasing fidelity led to superior transfer in a broad range of skills[50].

Self-determination theory provides an additionally relevant perspective for understanding motivation and learning. This theory posits that individuals' natural motivation to learn is driven by fulfilling three innate psychological needs: a sense of autonomy, a sense of relatedness and a sense of competence[51]. Self-determination theory identifies intrinsic motivation (doing something because it is inherently interesting) and integrated external motivation (doing something because it is relevant or leads to a desirable outcome)[51]. Intrinsic motivation results in high-quality learning and creativity, but integrated external motivation also comes with greater persistence and higher engagement[51,52]. *Choice* and the opportunity for *self-direction* appear to enhance intrinsic motivation, as they afford a greater sense of autonomy[51]. In the perspective of this theory, opportunities for self-directed learning in games create intrinsic motivation, in combination with interest in the game's subject. Game-based research also indicates learner control (as opposed to program control) has a positive effect on user's motivation[37]. When residents are aware that they will need certain skills in clinical practice or students know that they will be assessed, integrated external motivation usually is an additional important motive to study. As self-determination theory suggests that extrinsic rewards can undermine intrinsic motivation[51], it would be interesting to investigate the effect of a scoring system and competition in games on intrinsic motivation and performance.

In conclusion, our studies showed that serious games with realistic tasks can facilitate learning of complex cognitive skills, provided that the complexity of the tasks is matching the competency-level of the learner. Games often enhance motivation, but in the absence of improved performance there is little practical value in more engaged learning. High-fidelity tasks in games enhance motivation and cognitive effort even more, but may create cognitive overload and be detrimental for novice learners. In game design, complexity of tasks for the learner can be managed by offering guidance (worked cases, hints) and gradually increasing task difficulty and task fidelity with a variety of problems. Games appear to stimulate motivation by offering a challenging environment with meaningful tasks and learner control, creating a sense of autonomy for learners. Challenging tasks which match the skills level of the learner likely create a 'flow' experience and optimal learning. More design-based research in games is needed on the effects of case complexity and design features on performance and motivation, for novices and experts.

STRENGTHS AND LIMITATIONS OF THIS THESIS

The strengths of this thesis are its relevance for medical education, the design-based, empirical approach and the grounding of findings in theoretical perspectives of learning. Health care and medical education need novel, more flexible models for training and learning, thus making the topic of this thesis relevant. In addition, as fewer and sicker patients are admitted to hospitals, for shorter time periods, learning opportunities shrink, while the number of admitted medical students is increasing. More online, self-directed learning, with a variety of clinical cases and in adjunct to clinical experiences, is needed for expertise development[53]. In this thesis we investigated the effects of different choices in blended and online learning on performance and motivation. The second strength, the design-based research approach, is reflected in the fact that we performed several empirical studies in which we compared the effects of instructional formats, didactic features and presentation forms in serious games. The vast majority of papers on serious games in education speculate about the potential benefits of games for learning, or describe game design in general terms[39]. By formulating research questions based on a multi-level design framework, we aimed to prevent confounding, arising when multiple factors simultaneously influence the dependent variable, and outcomes can be interpreted in more than one way[54]. Furthermore, we did this for two separate serious games, in different domains. A third strength is grounding our findings in theory and evidence-based principles of learning. This thesis explained empirical findings from the studies in the context of theoretical perspectives on complex skills learning and motivation. Understanding the role of technology-based learning from the perspective of theories on learning, instructional design and motivation is important to bring this young field forward.

8

The limitations of the research studies lie in the generalizability of the results and in the potential problem of confounding. Despite the fact that we investigated two games with different learning goals (related to emergency care and patient safety) and with different user groups, the generalizability of results is compromised as results cannot be generalized to other user groups or learning goals. We did one study where a game on emergency care was used by residents; this study showed the game improved their cognitive skills. The two other studies we performed on games were done with 4th-year medical students, although the primary target groups of both games were residents. We decided to first investigate efficacy of the game(s) in medical students. We chose this approach because the medical expertise level of students in general is more homogenous than the expertise level of residents. In addition, it is hard to motivate busy residents to spend extra study time on educational formats without evidence for their effectiveness. Moreover, emergency care is a high-stake skill with limited space for testing new educational formats. Yet, as the competency level of learners is an important factor in the efficacy of instructional design, results from undergraduate students cannot be generalized to residents. Potentially confounding factors originate from the fact that in two studies we compared the effects of different presentation forms (Q4) and formats (Q5), keeping the content the same. Games are formats with often unique and intertwined features, in which presentation forms are hard to isolate (e.g. physical and functional fidelity) or for which an equivalent is not easy to develop in another format. This leaves the risk that more than one didactic feature may have differed and thereby influenced the outcomes. We tried to control this potential problem, which is common in design-based research, by minimizing the differences between conditions, carefully describing the didactic features in each intervention and subsequently critically analyzing the learning outcomes and its most probable relation to the design features.

FUTURE RESEARCH

Research findings provide some answers, but also raise new questions. Design-based research, grounded in theory and in evidence-based principles on learning, is useful in advancing our knowledge on the effects of serious games and simulation programs on performance and motivation. These online programs can either be used independently, or blended with classroom learning. Research on game-based learning and online learning in general, is still a young field. Based on this thesis, and personal experience in developing and evaluating online programs, we propose the following further studies:

Design-based research on the tasks and features that improve performance and motivation

More research is needed on the effects of task difficulty and guidance for complex skills learning with different user groups, focusing on questions such as: how can task complexity in games (and simulation programs) be balanced with competency levels of learners to facilitate optimal and engaged learning ('flow')? A second area in which further research is needed, can be summarized as the value-added approach[55], investigating the design features that improve a game's effectiveness for performance improvement and motivation (scoring systems, levels, task fidelity, forms of narrative feedback). Clear definitions of design features are an important part of this type of research. A useful four-dimensional framework for game research includes the learner (role, challenges, progress), pedagogic approaches (adaptation, feedback, hints), representation of the game (control, interaction, multimedia) and the context of learning (goals and tasks, setting)[43]. Cognitive oriented theories of learning and the motivation theory of self-determination provide a useful basis for further research in this dynamic field. A third interesting area for further research is to investigate the optimal blend of a serious game or simulation program with classroom training for complex skills learning. If a serious game is used in preparation for classroom training, can training time be proportionally reduced while retaining performance?

Design-based research on single and multi-player games that improve team skills

Developing communication and team skills is in general considered an important part of the medical curriculum[56]. These skills include interpersonal communication, leadership and decision making. Serious games can offer a story line where team skills are imbedded in a clinical context. Relevant research questions are: What is the added value of game-features for training team skills, compared to simpler simulation programs? What do multi-player games add to single-player games? Offering time-critical tasks, modeling and debriefing are interesting features to investigate in this context. A second area for further research is the optimal blend of a serious game and classroom training for developing team skills.

KEY POINTS OF THIS THESIS

The two aims of this thesis were to investigate the effectiveness of a blended training design compared to classroom training and the effectiveness and critical design features of serious games for performance and motivation. In this final section, key points of the thesis in relation to these aims are presented:

1. Blended learning, consisting of a mix of online and classroom learning, is at least as effective and attractive for learners as classroom learning for knowledge acquisition. In post-graduate training, a significant reduction in training costs can be realized without compromising the learning outcomes.

2. Blended learning facilitates adaptation to the learner's knowledge level, flexibility of access in time and place and scalability of learning. Didactic features that improve online learning are: information with exercises and feedback, demonstration videos, interactive and simulated cases. Workload for students in preparing classroom time is an important factor to consider in course-design.

3. Checklists have poor validity and reliability in the assessment of emergency care skills compared to more global competency scales or rating scales. Multiple raters and cases may further improve the validity of assessment.

4. A serious game on emergency care, containing patient cases with feedback in a high-fidelity environment, improves complex cognitive skills among residents, with limited self-study time. Although the same game is motivating for inexperienced medical students and stimulates them to study longer, it does not improve their emergency skills, compared to what they learn from an instructional e-module. These combined findings indicate an 'expertise reversal effect' between prior knowledge of learners (residents vs. students) and instructional format (game vs. e-module).

5. A serious game on patient safety, containing video-lectures, patient cases and bio-feedback exercises, is motivating for medical students and stimulates them to study longer; however it does not improve their patient safety awareness. Video-lectures from the game and text-based lectures from an e-module are equally effective in developing patient safety knowledge.

6. The development of a serious game requires considerable resources. For knowledge-oriented learning goals there is in general little added value of a game; other educational formats serve this purpose just as well, at lower costs.

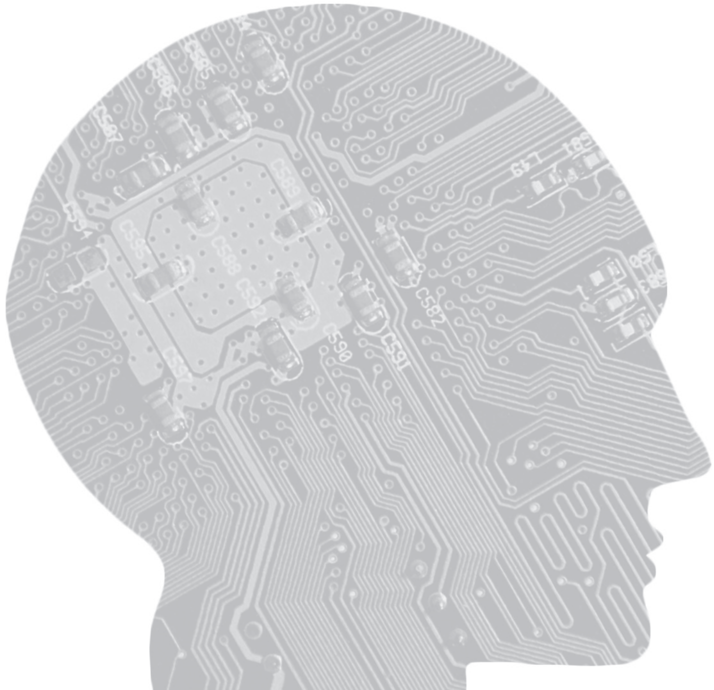
7. Although serious games with realistic tasks often enhance motivation, they are as such not sufficient for improving performance. Games may support skills learning, provided the complexity of the tasks matches the competency-level of the learner. Complexity for the learner can be managed by offering tasks with guidance (worked examples, hints). High-fidelity presentation of tasks stimulates motivation and investment of mental effort, but can nevertheless impede learning by novices.

REFERENCES

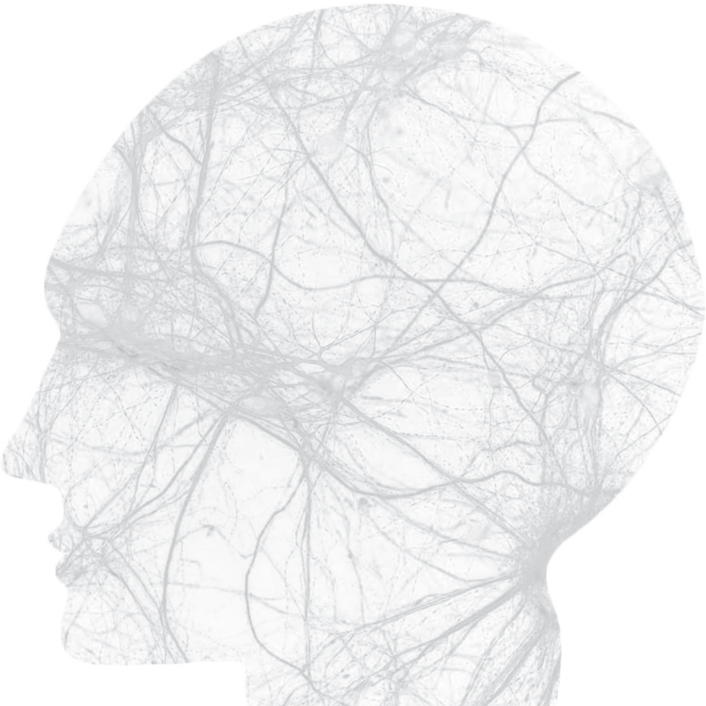
1. Ruiz JG, Mintzer MJ, Leipzig RM. The impact of E-learning in medical education. Acad Med. 2006 Mar;81(3):207–12.
2. Bates T. Managing technological change. San Fransisco: Jossey-Bass; 2000.
3. Moskal P, Dziuban C, Hartman J. Blended learning: A dangerous idea? Internet High Educ. Elsevier Inc.; 2013;18:15–23.
4. Abdelhai R, Yassin S, Ahmad MF, Fors UGH. An e-learning reproductive health module to support improved student learning and interaction: a prospective interventional study at a medical school in Egypt. BMC Med Educ. BioMed Central Ltd; 2012 Jan;12(1):11.
5. Taradi SK, Taradi M, Radic K, Pokrajac N. Blending problem-based learning with Web technology positively impacts student learning outcomes in acid-base physiology. Adv Physiol Educ. American Physiological Society; 2005 Mar 1;29(1):35–9.
6. Beale EG, Tarwater PM, Lee VH. A retrospective look at replacing face-to-face embryology instruction with online lectures in a human anatomy course. Anat Sci Educ. 2013 Aug 19;
7. Kiili K. Digital game-based learning: Towards an experiential gaming model. Internet High Educ. 2005;8:13–24.
8. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, Schünemann HJ. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach. 2010 Jan;32(1):16–27.
9. Sitzmann T. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. Pers Psychol. 2011 Jun 27;64(2):489–528.
10. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012 Oct;99(10):1322–30.
11. Bernard RM, Borokhovski E, Schmid RF, Tamim RM, Abrami PC. A meta-analysis of blended learning and technology use in higher education: from the general to the applied. J Comput High Educ. 2014;26:87–122.
12. Sitzmann T, Kraiger K, Stewart D, Wisher R. The comparative effectiveness of web-based and classsroom instruction: a meta-analyses. Pers Psychol. 2006;59(3):623.
13. Stewart A, Inglis G, Jardine L, Koorts P, Davies MW. A randomised controlled trial of blended learning to improve the newborn examination skills of medical students. Arch Dis Child Fetal Neonatal Ed. 2013 Mar;98(2):F141–4.
14. Sherman H, Comer L, Putnam L, Freeman H. Blended versus lecture learning: outcomes for staff development. J Nurses Staff Dev. 2012 Jul;28(4):186–90.
15. Bates TW. Teaching in a Digital Age. Bates, editor. www.tonybates.ca; 2015.
16. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA. 2008 Sep 10;300(10):1181–96.
17. Clark RC, Mayer RE. E-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning. Elearning. San Fransisco: Pfeiffer; 2008. 476 p.
18. Issa N, Schuller M, Santacaterina S, Shapiro M, Wang E, Mayer RE, DaRosa DA. Applying multimedia design principles enhances learning in medical education. Med Educ. Blackwell Publishing Ltd; 2011 Aug 1;45(8):818–26.

19. Kirschner PA, van Merriënboer JJG. Do Learners Really Know Best? Urban Legends in Education. Educ Psychol. 2013 Jul;48(3):169–83.
20. Cook DA, Thompson WG, Thomas KG. Test-Enhanced Web-Based Learning. Acad Med. 2014;89(1): 169–75.
21. Stark R, Kopp V, Fischer MR. Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. Learn Instr. Elsevier Ltd; 2011;21(1): 22–33.
22. Johnson DW, Johnson RT. An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. Educ Res. 2009;38(5):365–79.
23. Perkins GD, Fullerton JN, Davis-Gomez N, Davies RP, Baldock C, Stevens H, Bullock I, Lockey AS. The effect of pre-course e-learning prior to advanced life support training: a randomised controlled trial. Resuscitation. 2010 Jul;81(7):877–81.
24. Francom GM, Gardner J. What is task-centered learning? TechTrends. 2014;58(5).
25. Van Merriënboer JJ, Clark RE, Croock MB. Blueprints for complex learning: The 4C/ID-model. Educ Technol Res Dev. 2002 Jun;50(2):39–61.
26. Van Merrienboer JJ, Kirschner PA. Ten steps to complex learning. A systematic approach to four-component instructional design. 2nd Editio. Routledge; 2012. 49-77 (Chapter 4) p.
27. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach. 2014 May;36(5):371–84.
28. Kalyuga S, Ayres PL, Chandler P, Sweller J. The expertise reversal effect. Educ Psychol. 2003;38(1): 23–31.
29. Mayer RE. Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. Am Psychol. 2004;59(1):14–9.
30. Baxter P, Akhtar N, Landeen J, Norman G. Teaching critical management skills to senior nursing students: videotaped or inter-active hands-on instruction? Nurs Educ Perspect. 2012;33(2):106–10.
31. Wouters P, van Oostendorp H. A meta-analytic review of the role of instructional support in game-based learning. Comput Educ. Elsevier Ltd; 2013 Jan;60(1):412–25.
32. Leppink J, van den Heuvel A. The evolution of cognitive load theory and its application to medical education. Perspect Med Educ. 2015;119–27.
33. Nie Y, Li L, Duan Y, Chen P, Barraclough BH, Zhang M, Li J. Patient safety education for undergraduate medical students: a systematic review. BMC Med Educ. BioMed Central Ltd; 2011 Jan;11(1):33.
34. Gordon M. Non-technical skills training to enhance patient safety. Clin Teach. 2013 Jul;10(3): 170–5.
35. Schumacher DJ, Englander R, Carraccio C. Developing the master learner: applying learning theory to the learner, the teacher, and the learning environment. Acad Med. 2013;88(11):1635–45.
36. Kneebone R. Evaluating Clinical Simulations for Learning Procedural Skills: A Theory-Based Approach. Acad Med. 2005 Jun;80(6):549–53.
37. Garris R, Ahlers R, Driskell JE. Games, Motivation, and Learning: A Research and Practice Model. Simul Gaming. 2002 Dec 1;33(4):441–67.
38. Huang WH. Evaluating learners' motivational and cognitive processing in an online game-based learning environment. Comput Human Behav. Elsevier Ltd; 2011;27(2):694–704.

39. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. Elsevier Ltd; 2012 Sep;59(2): 661–86.
40. Bedwell WL, Pavlas D, Heyne K, Lazzara EH, Salas E. Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study. Simul Gaming. 2012 May 14;43(6):729–60.
41. Young MF, Slota S, Cutter AB, Jalette G, Mullin G, Lai B, Simeoni Z, Tran M, Yukhymenko M. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. Rev Educ Res. 2012 Feb 1;82(1):61–89.
42. Huang, W.D & Johnson T. Handbook of Research on Effective Electronic Gaming. In: Ferdig RE, editor. IGI Global; 2008.
43. Van Staalduinen JP, de Freitas S. A First Step towards integrating educational theory and game design. In: Felicia P, editor. Research on improving learning and motivation through educational games: multidisciplinairy approaches. Hershey, PA: IGI Global; 2010. p. 28.
44. Csikszentmihalyi M. Flow: the psychology of optimal performance. New York: Cambridge University Press; 1990.
45. Vygotsky L. Mind in society: The development of higher psychological processes. Oxford, UK: Harvard University Press; 1978.
46. Cheng MT, Su T, Huang WY, Chen JH. An educational game for learning human immunology: What do students learn and how do they perceive? Br J Educ Technol. 2013;45(5):820–33.
47. Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? Adv Health Sci Educ Theory Pract. 2012 Mar;17(1):137–44.
48. Huang WD, Johnson TE, Han SHC. Impact of online instructional game features on college students' perceived motivational support and cognitive investment: A structural equation modeling study. Internet High Educ. Elsevier Inc.; 2013;17(1):58–68.
49. Hamstra SJ, Brydges R, Hatala R, Cook D. In Reply to Rubio et al. Acad Med. 2014;89(10):2014.
50. Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximize educational benefit. Acad Med. 2010 May;85(5):806–12.
51. Ryan R, Deci E. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol. 2000;25:54–67.
52. Ryan RM, Deci EL. Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? J Pers. 2006;74(December):1557–85.
53. Norman G. Medical education: past, present and future. Perspect Med Educ. 2012 Mar;1(1):6–14.
54. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med. 2005 Jun;80(6):541–8.
55. Mayer RE. Computer games for learning: an evidence-based approach. E-Pub. Cambridge, MA: MIT press; 2014. 264 p.
56. Aboumatar HJ, Thompson D, Wu A, Dawson P, Colbert J, Marsteller J, Kent P, Lubomski LH, Paine L, Pronovost P. Development and evaluation of a 3-day patient safety curriculum to advance knowledge, self-efficacy and system thinking among medical students. BMJ Qual Saf. 2012 May;21(5): 416–22.



Summary



In medical education and health care, with its exponential growth in knowledge and increasing demands on the needed competencies of doctors, there is a need for new and more cost-effective training models. Blended and online learning are powerful learning concepts, which can be used in different formats (e-modules, games) to gain knowledge, skills and attitudes.

'Blended' (or hybrid) learning combines online learning and classroom learning. Potential benefits are more flexibility in online learning and high-quality, interactive classroom sessions. Evidence from research is not conclusive on whether blended learning is equally effective or more effective than solely classroom learning. The **first aim** of this thesis was to compare the effectiveness of a blended training design with a conventional classroom training for the acquisition of knowledge.

One of the promising new online formats to train complex skills in a motivating and efficient way is provided by serious games. Little is yet known about the effectiveness and optimal design of serious games in health care training. The **second aim** of this thesis was to investigate the effectiveness and critical design features of serious games for performance and motivation.

In **Chapter 1**, the concepts of online learning and blended learning are introduced. For blended learning, different mixes of online and classroom learning can be chosen. For designing online learning, different formats, didactic features and presentation types can be used. A multi-level framework for structuring design decisions and enabling systematic comparisons of choices is presented. Available research on the different levels of this framework is summarized, leading to the six central research questions of this thesis.

Despite the growing popularity of blended learning, little is known on whether blended learning is equally or more effective as classroom-based learning. Our first research question was: What is the effectiveness of a blended training design compared to a classroom-based training design for knowledge acquisition in a postgraduate training program? **(Q1)**

One of the two serious games we used for our studies was designed to train cognitive emergency care skills for residents (physicians who are in training for a medical specialty). Little evidence exists on the validity and reliability of (internationally used) assessment instruments in emergency care. Hence our second research question was: What is the validity and reliability of commonly used formats in the assessment of emergency care skills for residents? **(Q2)** Serious games are new, promising online training formats in health care and education. Research findings are not clear on the effectiveness of this format for motivation and skills development. Therefore our third research question was: Do residents, who use a serious game as a preparation for classroom training, show better complex cognitive skills before training than residents who do not

play the game? **(Q3)** High-fidelity cases are often used in simulation games, although its contribution to learning and motivation is still unclear. Our fourth research question was: What are the effects of adding high-fidelity patient cases (the game on emergency care) compared to low-fidelity (text-based) cases on the acquisition of cognitive skills and motivation of medical students? **(Q4).** The second serious game we investigated was focused on patient-safety knowledge and skills. Therefore our fifth research question was: Do students develop better patient safety knowledge and skills and are they more motivated after using a highly interactive serious game than after using a limited interactive e-module on the same topics? **(Q5)** Serious games are available in a variety of genres. Currently, there is no systematic framework to evaluate games. Therefore, our sixth research question was: How can serious games applied in health care and medical education be systematically evaluated? **(Q6)**

Chapter 2 describes a study comparing the effectiveness of a conventional classroom design in a postgraduate course (11 day course) with a blended course design, consisting of two-third classroom training (7 days) and one-third online self-study (Q1). Participants from the blended group and the classroom-based group did not differ in learning results on knowledge tests. Evaluation results of both groups were equally positive, with the blended group being more self-confident about their knowledge. Evaluation showed they appreciated practicing exercises at their own pace and level. Training costs of the blended design were reduced with one-third compared to the conventional design. Thus, blended learning is equally effective and attractive as classroom training. In post-graduate training, a blended design can lead to a significant reduction in costs, without compromising performance.

In **Chapter 3** we describe the validity and reliability of (internationally) commonly used formats in the assessment of complex cognitive emergency care skills for residents (Q2), as a basis for the effectiveness study in Chapter 4. The scenario assessments were performed by trained raters with a standardized simulation patient; the assessment instrument included a checklist, a competency scale and a global performance scale. Psychometric analyses were done with assessment results of 144 residents and 12 trained raters; 22 residents were videotaped and re-assessed by eight experienced raters. Results showed that checklists have poor validity and reliability in assessing emergency care skills. The clinical competency scale had good validity and a moderate reliability; the global performance scale had a moderate reliability. We concluded that more global assessment instruments have a better reliability and validity than detailed checklists for assessing emergency skills. The limited number of assessment scenarios and raters, as typically used in emergency care assessment around the world, results in a moderate validity and reliability.

In the study presented in **Chapter 4**, we investigated whether residents who used the abcde*SIM* game in addition to course material as a preparation for classroom training ('reading & game' group), show better emergency care skills at the start of the training than a control group who only used the course material ('reading' group) (Q3). The assessment instrument we used to assess the complex cognitive emergency care skills was based on the findings from the validation study described in Chapter 3. Before classroom training, the reading & game group studied 2.5 hours longer and showed higher scores on objectively measured and self-assessed clinical competencies (with a medium to large effect size) than the reading group. The reading& game group felt engaged with the game. Groups were equally motivated for the training. After two weeks of training, we no longer found differences in competencies between groups. We concluded serious games can be used as an effective, motivating preparation for instructor-led emergency care courses to train residents' clinical competencies. Future research will have to show how this effect can be sustained and whether training time can be reduced in combination with this type of online preparation, maintaining learning outcomes. As this study was inconclusive on which design choices had been responsible for the learning effects, in our next study we focused on the effect of fidelity in games.

In the study described in **Chapter 5**, we investigated the effect of high versus low fidelity patient cases, in adjunct to an instructional e-module, on medical students' acquisition of cognitive emergency care skills and motivation (Q4). In a three-group randomized design, a control group worked on an instructional e-module explaining the emergency care approach (including a demonstration video); a 'cases group' combined the e-module with a text-based patient cases program (low-fidelity condition), and a 'game group' combined the e-module with the abcde*SIM* game (high-fidelity condition), using the same patient cases. After a study period, blinded assessors rated students' cognitive emergency care skills in two mannequin-based scenarios. A double assessment scenario was chosen to improve validity of assessment, building on the findings from the validity study described in Chapter 2. We also evaluated different types of perceived cognitive load. Although learning time was two hours longer for both the cases group and the game group compared to the control group, acquired cognitive skills did not differ between groups. However, students who spent more game-time showed better skills than students who spent less game-time. The game and cases group experienced higher intrinsic cognitive load from the cases compared to the e-module. The game group experienced higher intrinsic and germane load than the cases group and felt more engaged with the format. We concluded that for initial skill acquisition, an e-module with a demonstration video is a powerful instructional format and provides sufficient opportunity for learning. Inexperienced students did not benefit from additional work on open cases, which appeared to be too complex, but nonetheless challenged students to study longer. The high-fidelity game increased complexity and did not improve learners' skill level, even though students put more effort into it and felt more engaged than with the low-fidelity cases. Fidelity in cases enhances motivation and stimulates cognitive effort, but may increase complexity and impede learning for novices.

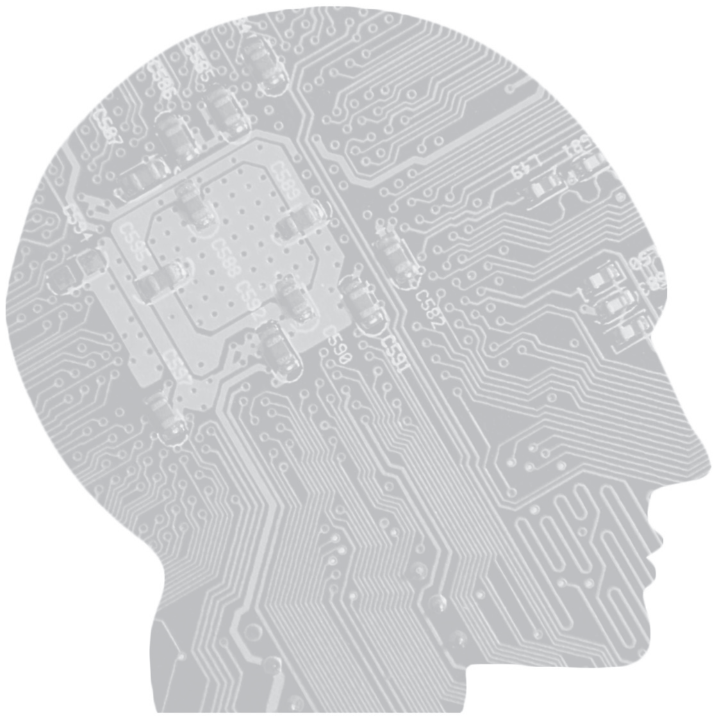
Chapter 6 reports on the question: Do undergraduate medical students develop better patient safety knowledge and awareness, and are they more motivated after using a highly interactive serious game (Air Medic Sky-1) than after using a limited interactive e-module? (Q5). A third group acted as a historical control group without extra education. The game includes video lectures, biofeedback exercises (using finger sensors that keep track of heart rate and skin conductance) and patient missions. The e-module includes text-based lectures on the same patient safety topics. Average learning time was three hours for the game and one hour for the e-module group. The game (particularly the part with the patient missions), was evaluated as considerably more engaging than the e-module, with large effect sizes. We found that patient safety knowledge was equally improved in the game group and the e-module group, compared to the control group. During rotations, students in the three groups reported similar low levels of stress and patient-safety awareness. Interestingly, students who had played the game longer (and treated more 'patients' successfully), experienced higher self-efficacy and reported less stress during their rotation than students who spent less game-time. We concluded that video-lectures and text-based lectures, however, are equally effective in developing knowledge on patient safety topics. Although games with interactive patient cases, where students can practice patient safety situations, are strongly engaging and stimulate longer study, they do not result in improved patient safety awareness. The value of biofeedback exercises for stress-management should be investigated with learners experiencing higher stress levels.

Chapter 7 presents a general framework to systematically evaluate serious games applied to health care and medical training (Q6). The framework provides a set of standardized criteria to evaluate the rationale behind a serious game (e.g., goals and user groups) and assess its content and didactic functionality (e.g., the relation between learning and game play and the game assessment parameters), effectiveness (e.g., predictive validity of game outcomes on outcomes in real life), and data safety (e.g., data ownership). Using this framework, developers can support end-users in evaluating the usability and potential benefits of a specific serious game. We expect this will stimulate a thoughtful and safe implementation of serious games in medical education.

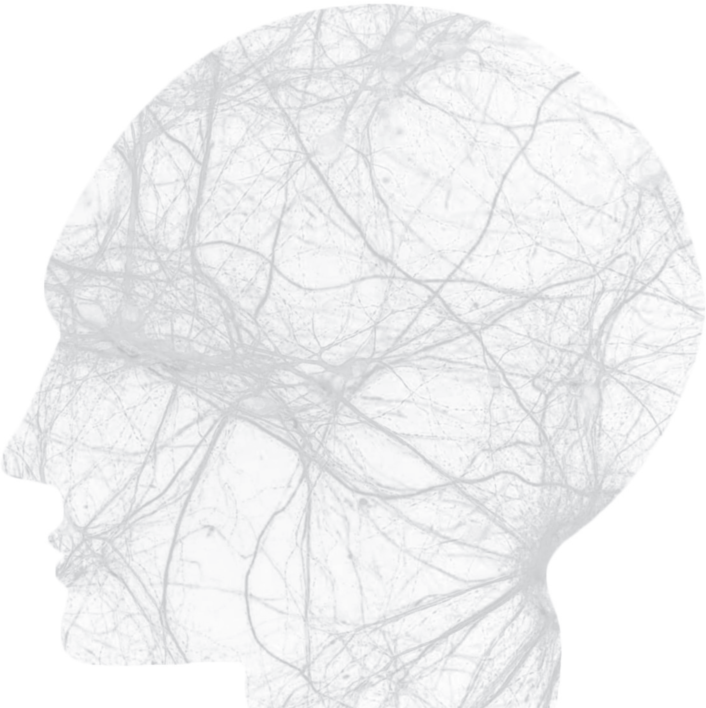
In **Chapter 8** a general discussion of the findings of this thesis is provided, including the main conclusions from the studies, the implications for instructional design and the strengths and limitations of this thesis. Next we present recommendations for future research and state the key points of the thesis. The first aim of this thesis was to investigate the effectiveness of a blended training design compared to classroom training. For knowledge acquisition, blended learning is equally effective as classroom learning. Blended learning facilitates adaptation to the learner's knowledge level, flexibility in time and scalability of learning. In post-graduate training, a significant reduction in training costs can also be realized. Effective didactic methods in the design of the online component are: information in conjunction with exercises and feedback, demonstration videos, interactive and simulated cases. An important consideration in the design is also the amount of online self-study. Securing online preparation is a critical success factor in blended learning. Theories of multimedia learning are useful as a basis for further research.

The second aim was to investigate the effectiveness and critical design features of serious games for performance and motivation. For the assessment of complex cognitive emergency care skills, checklists have poor validity and reliability compared to more global competency scales or rating scales. A serious game with realistic, interactive cases improved complex cognitive skills for residents, despite limited self-study time. Although the same game was motivating for inexperienced students and stimulated them to study longer, it did not improve their cognitive skills, compared to what they learned from an instructional e-module. These combined results indicate an 'expertise reversal effect' between prior knowledge of learners (residents vs. students) and instructional format (game vs. e-module). Serious games provide affordances for flexible practice with simulated tasks; they often enhance motivation, but do not necessarily result in better performance. High-fidelity tasks in games enhance motivation and cognitive effort even more, but may create cognitive overload and be detrimental for novice learners. In designing games, complexity of tasks for the learner can be managed by offering guidance (worked cases, hints) and gradually increasing task difficulty and task fidelity with a variety of problems. Games appear to stimulate motivation by offering challenging, meaningful tasks and learner control, creating a sense of autonomy for learners. Challenging tasks which match the skills level of the learner are likely to effectuate a 'flow' experience and optimal learning. The development of a serious game requires considerable resources. For knowledge-oriented learning goals there is in general little added value of a game; other educational formats serve this purpose just as well and at lower costs.

Although serious games with realistic tasks often enhance motivation, they are as such not sufficient for improving performance. Games may support skills learning, provided task complexity matches the learner's competency-level. More design-based research is needed on the effects of task complexity and design features on performance improvement and motivation, for novices and experts.



Samenvatting



In het medisch onderwijs en de gezondheidszorg, met zijn exponentiële groei in kennis en toenemende eisen aan de vaardigheden van artsen, bestaat behoefte aan nieuwe, meer kosteneffectieve trainingsmodellen. Blended en online leren zijn krachtige leerconcepten, waarbij verschillende formats (serious games, e-modules, simulaties) kunnen worden ingezet om kennis, vaardigheden en attitudes te ontwikkelen.

'Blended' (of hybride) leren combineert online leren en contactonderwijs. Hiermee kan meer flexibiliteit in het online deel worden gerealiseerd, gecombineerd met hoog interactieve klassikale sessies. Over de vraag of blended leren net zo effectief is als alleen klassikaal leren, of effectiever, zijn de onderzoeksresultaten niet eenduidig. Het **eerste doel** van dit proefschrift is de effectiviteit te vergelijken van een blended onderwijsontwerp met klassikaal onderwijs voor kennisontwikkeling.

Serious games zijn relatief nieuwe, veelbelovende formats om complexe vaardigheden op een motiverende en efficiënte wijze te trainen. Er is nog weinig bekend over de effectiviteit en optimale ontwerpkeuzes van serious games voor (medisch) onderwijs en training. Het **tweede doel** van dit proefschrift is de effectiviteit van en kritische ontwerpkeuzes bij serious games te onderzoeken voor de bevordering van leerprestaties en motivatie.

In **Hoofdstuk 1** worden de concepten online leren en blended leren geïntroduceerd. Bij blended leren kunnen verschillende verhoudingen van klassikaal en online leren worden ingezet. Voor het ontwerpen van online leren kunnen diverse formats, didactische methoden en presentatie typen worden gebruikt. Met een conceptueel kader dat de verschillende stappen in het ontwerpproces beschrijft kan dit proces worden gestructureerd en is een systematische vergelijking van keuzes mogelijk. Nadat dit raamwerk is beschreven, wordt beschikbaar onderzoek met betrekking tot effectiviteit van ontwerpkeuzes bij de diverse ontwerpstappen samengevat. Dit is de basis voor de formulering van de zes centrale onderzoeksvragen van dit proefschrift.

Ondanks de toenemende populariteit van blended leren, is nog weinig bekend over de vraag of blended leren net zo effectief is, of effectiever dan klassikaal leren. Daarom was onze eerste onderzoeksvraag: Wat is de effectiviteit van een blended trainingsontwerp voor een kennisgerichte nascholing in vergelijking met een klassikaal ontwerp? **(Q1)** Een van de twee serious games die het onderzoeksobject vormen voor een aantal studies uit dit proefschrift, traint artsen in opleiding tot specialist (aios) in cognitieve acute zorgvaardigheden. Er is nog weinig bekend over de validiteit en betrouwbaarheid van (internationaal) gebruikte beoordelingsinstrumenten in trainingen acute zorg. Daarom was onze tweede onderzoeksvraag: Wat is de validiteit en betrouwbaarheid van veel gebruikte instrumenten in de beoordeling van acute zorgvaardigheden bij aios? **(Q2)** Serious games zijn vrij nieuwe trainingsformats in het medisch onderwijs. Onderzoeksresultaten zijn niet eenduidig over de effectiviteit van serious games voor de leerprestaties en motivatie van cursisten. Daarom luidde onze derde onderzoeksvraag: Hebben AIOS, die een serious game acute zorg als voorbereiding van een klassikale training gebruiken, betere cognitieve vaardigheden aan het begin van de training dan aios die de game niet spelen? **(Q3)** Bij veel simulatiegames wordt casuïstiek met een hoge natuurgetrouwheid ingezet, alhoewel de bijdrage van natuurgetrouwheid aan leren en motivatie nog onduidelijk is, met name voor onervaren studenten. Onze vierde onderzoeksvraag luidde daarom: Wat zijn de effecten van het toevoegen van patiëntcasuïstiek met een hoge natuurgetrouwheid (de game acute zorg) in vergelijking met patiëntcasuïstiek met een lage natuurgetrouwheid (tekstgebaseerde cases) op de verwerving van cognitieve vaardigheden en motivatie door medische studenten? **(Q4)**. De tweede serious game die we hebben onderzocht was een interactieve game voor de ontwikkeling van kennis over en bewustzijn van patiëntveiligheid. Onze vijfde onderzoeksvraag was: Hebben studenten een beter niveau van kennis over en bewustzijn van patiëntveiligheid en zijn ze meer gemotiveerd na het leren met een interactieve serious game dan met een beperkt interactieve e-module over dezelfde onderwerpen? **(Q5)** Er zijn verschillende typen serious games op de markt beschikbaar. Er is echter momenteel geen raamwerk waarmee games kunnen worden geëvalueerd door potentiële gebruikers. Onze laatste onderzoeksvraag was dan ook: Hoe kunnen serious games voor de gezondheidszorg en medisch onderwijs systematisch worden geëvalueerd? **(Q6)**.

In **Hoofdstuk 2** beschrijven we een retrospectieve studie waarin de effectiviteit van een (11-daagse) klassikale nascholingscursus wordt vergeleken met een blended cursusopzet, bestaande uit 2/3 deel klassikaal onderwijs (7 dagen) en 1/3 deel online zelfstudie (Q1). Cursisten uit de klassikale groep en de blended groep verschilden na afloop niet in leerresultaten op de kennistoetsen. Ook waren de evaluatieresultaten met betrekking tot de cursus van beide groepen positief, waarbij de blended groep meer vertrouwen toonde in eigen kennis. Evaluaties gaven verder aan dat deze groep met name het oefenen in eigen tempo en op hun eigen niveau waardeerden. De trainingskosten van de blended opzet bedroegen eenderde minder dan die van de klassikale opzet. We concludeerden dat blended onderwijs even effectief en aantrekkelijk is als volledig klassikaal onderwijs. Bij postinitieel onderwijs kan het leiden tot een significante reductie van kosten, met behoud van leerprestaties.

In het onderzoek van **Hoofdstuk 3** beschrijven we de validiteit en betrouwbaarheid van (internationaal) veel gebruikte instrumenten in de beoordeling van complexe cognitieve vaardigheden acute zorg voor aios (Q2), als basis voor de effectiviteitstudie van Hoofdstuk 4. De scenariogebaseerde beoordeling werd gedaan door getrainde beoordelaars met (gestandaardiseerde) simulatiepatiënten. Het beoordelingsinstrument omvatte een checklist, competentielijsten en een globale beoordelingslijst. Psychometrische analyses zijn uitgevoerd met de beoordelingsresultaten van 144 aios en 12 getrainde beoordelaars. Van 22 aios werd de beoordeling op video opgenomen en herbeoordeeld door acht ervaren beoordelaars. De resultaten lieten zien dat checklists een lage validiteit en betrouwbaarheid hebben voor de beoordeling van vaardigheden acute zorg. De klinische competentielijst en de globale beoordelingslijst had een goede validiteit en een matige betrouwbaarheid. De conclusie was dat meer globale beoordelingsinstrumenten betrouwbaarder en meer valide zijn dan gedetailleerde checklists bij de beoordeling van complexe cognitieve vaardigheden acute zorg. Het beperkte aantal scenario's en beoordelaars, zoals gebruikelijk bij de beoordeling van trainingen acute zorg wereldwijd, resulteert in een matige betrouwbaarheid en validiteit van de beoordelingen.

In het onderzoek dat in **Hoofdstuk 4** wordt beschreven, hebben we onderzocht of aios die een serious game (de abcde*SIM*) gebruiken in aanvulling op een cursusboek (de 'cursusboek & game groep') ter voorbereiding op een klassikale training acute zorg, voor deze training betere vaardigheden hebben dan aios die alleen het cursusboek hebben gebruikt (de 'leesgroep') (Q3). Het gebruikte beoordelingsinstrument was gebaseerd op de resultaten van de validatie studie uit Hoofdstuk 3 (competentielijsten en een globale beoordeling). Voor de klassikale training studeerde de cursusboek & game groep 2.5 uur langer, had een hoger gemeten niveau van klinische vaardigheden (met een medium tot grote effectgrootte), en beoordeelde zelf de klinische vaardigheden ook beter dan de leesgroep. De cursusboek & game groep was gemotiveerd om de game te spelen. Beide groepen waren even gemotiveerd voor de inhoud van de klassikale training. Na 2 weken training vonden we echter geen verschillen meer in competenties tussen de groepen. We concludeerden dat serious games kunnen worden ingezet als een effectieve, motiverende voorbereiding op een klassikale cursus om klinische vaardigheden van aios te trainen. Nader onderzoek zal moeten uitwijzen hoe dit effect kan worden behouden en of trainingstijd in combinatie met deze voorbereiding kan worden ingekort, met behoud van leerprestaties. Aangezien deze studie geen uitsluitsel gaf over de vraag welke ontwerpkeuzes bepalend zijn voor het leereffect, hebben we ons in de volgende studie gericht op het effect van natuurgetrouwheid van casuïstiek in simulatiegames.

In de studie die we in **Hoofdstuk 5** rapporteren, hebben we het effect onderzocht van hoge en lage natuurgetrouwheid van casuïstiek op de motivatie en verwerving van cognitieve vaardigheden acute zorg van medische studenten (Q4). Deze casuïstiek werd aangeboden in aanvulling op een instructieve e-module. In een gerandomiseerde opzet met 3 groepen, kreeg de controlegroep alleen een instructieve e-module waarin de basisprincipes van acute zorg werden toegelicht (inclusief een demonstratievideo); de 'tekst-cases groep' kreeg naast de e-module tekstgebaseerde patiëntcasuïstiek (de lage natuurgetrouwheid conditie), en de 'game groep' combineerde de e-module met de abcde*SIM* game (de hoge natuurgetrouwheid conditie), met daarin dezelfde patiën-tcasuïstiek. Na een studieperiode werden de klinische (cognitieve) vaardigheden acute zorg van de studenten beoordeeld door (voor de conditie) geblindeerde beoordelaars in twee scenario's met simulatiepoppen. Daarnaast werd de ervaren cognitieve belasting bij studenten gemeten. Ondanks het feit dat de leertijd bij zowel de tekst-cases groep als de game groep twee uur langer was dan bij de controlegroep, verschilden de gemeten cognitieve vaardigheden van groepen niet. Studenten die meer tijd aan de game besteedden, hadden echter wel een hoger vaardigheidsniveau dan korter spelende studenten. Zowel de tekst-cases als de game groep ervaarde een hogere cognitieve belasting bij de patiëntcasuïstiek dan bij de e-module. De game groep ervaarde een hogere cognitieve belasting, ervaarde meer mentale inspanning en voelde zich meer gemotiveerd door het game format dan de tekst-cases groep. We concludeerden dat in de beginfase van vaardigheidsontwikkeling, een e-module met een demonstratievideo veel instructieve waarde heeft. Onervaren studenten profiteren niet van additioneel oefenen met open casuïstiek, indien deze voor hen te complex is, ook al stimuleert deze hen wel tot langer studeren. Hoge natuurgetrouwheid in casuïstiek van een serious game werkt motiverend en stimuleert het leveren van mentale inspanning, maar kan bij onervaren studenten de complexiteit verhogen en het leren belemmeren.

Hoofdstuk 6 beschrijft een studie naar de vraag of medische studenten die een interactieve serious game over patiëntveiligheid (Air Medic Sky-1) gebruiken, meer kennis van en bewustzijn over patiëntveiligheid ontwikkelen en gemotiveerder zijn dan studenten die een beperkt interactieve e-module gebruiken (Q5). Een derde groep fungeerde als een historische controlegroep en kreeg geen extra instructie over patiëntveiligheid. De game bevat video-lessen, biofeedback oefeningen (met vingersensoren die de hartslag en huidgeleiding meten) en interactieve patiëntscenario's. De e-module bevat tekstlessen over dezelfde onderwerpen, maar met beperkte interactiviteit. De gemiddelde leertijd was drie uur voor de game groep en een uur voor de e-module groep. De game, met name het deel met de patiëntscenario's, werd ervaren als meer motiverend (met een grote effect-grootte) dan de e-module. Kennis over patiëntveiligheid was zowel in de game- als in de e-module groep toegenomen ten opzichte van de controle groep. Tijdens het daaropvolgend co-assistentschap rapporteerden studenten van alle drie groepen echter lage, vergelijkbare niveaus van bewustzijn over patiëntveiligheid en stress. Studenten die de game langer speelden (en meer 'patiënten' succesvol hadden behandeld), ervaarden wel een hogere self-efficacy (geloof in eigen kunnen) en minder stress tijdens het co-assistentschap dan studenten die de game korter speelden.

We concludeerden dat video-lessen en tekst-lessen even effectief zijn om kennis over patiëntveiligheid te ontwikkelen. Alhoewel games, met interactieve casuïstiek over patiëntveiligheid, motiverend zijn voor studenten en aanzetten aan tot extra studie, leiden ze niet tot meer bewustzijn over patiëntveiligheid. De waarde van biofeedback oefeningen voor stressreductie dient in vervolgonderzoek te worden onderzocht met cursisten die hogere stressniveaus ervaren.

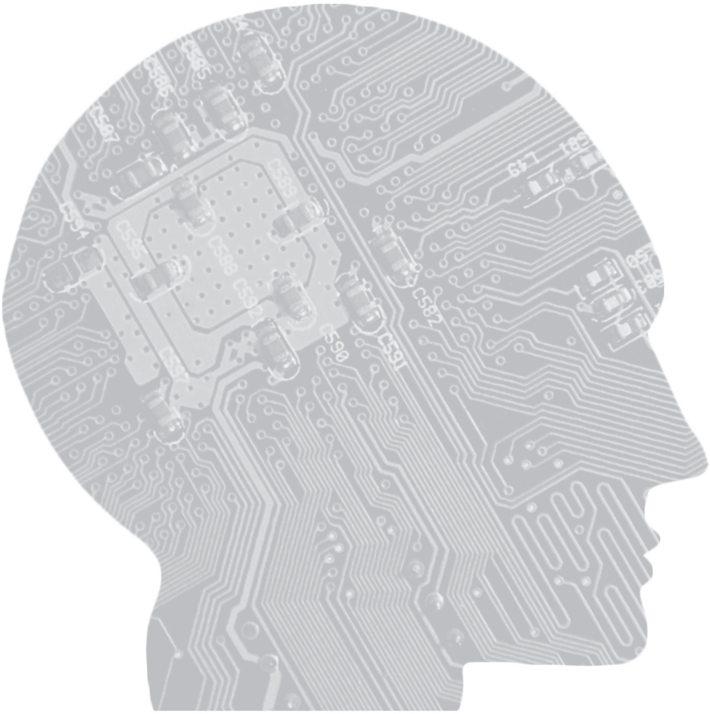
Hoofdstuk 7 presenteert een raamwerk om serious games voor de gezondheidszorg en medisch onderwijs systematisch te evalueren (Q6). Het raamwerk biedt een set van gestandaardiseerde criteria om de rationele van een game (o.a. doelen en doelgroepen), de inhoud en didactische functionaliteit (o.a. de relatie tussen de leerdoelen, gamedoelen en scoringsparameters), de effectiviteit (o.a. de voorspellende validiteit van game resultaten naar resultaten buiten de game), en dataveiligheid (data-eigenaarschap) te evalueren. Met behulp van dit raamwerk kunnen game-ontwikkelaars gebruikers ondersteunen in het beoordelen van de bruikbaarheid en de potentiële voordelen van een specifieke serious game. We verwachten dat dit raamwerk een weloverwogen en veilige inzet van serious games in medisch onderwijs zal ondersteunen.

Hoofdstuk 8 bevat de discussie over de bevindingen van dit proefschrift. Als eerste worden de conclusies met betrekking tot de studies samengevat, de implicaties voor het ontwerpen van blended onderwijs en serious games geschetst en de sterke kanten en beperkingen van de studies benoemd. Tot slot worden aanbevelingen gedaan voor verder onderzoek en de kernpunten van dit proefschrift geformuleerd.

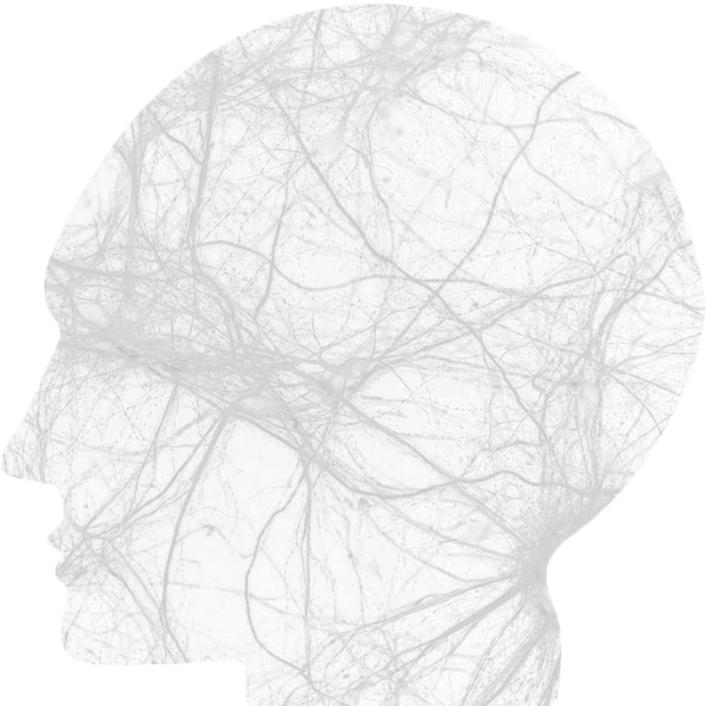
Het eerste doel van dit proefschrift was te onderzoeken of een blended onderwijsopzet net zo effectief is als klassikaal onderwijs voor kennisontwikkeling. Voor kennisverwerving is een blended opzet net zo effectief als een klassikale opzet. Een blended opzet biedt daarnaast meer flexibiliteit in tijd en plaats, betere mogelijkheden voor aansluiting aan het kennisniveau van de cursisten en maakt een groter bereik mogelijk. In postinitieel onderwijs kan hiermee tevens een aanzienlijke kostenreductie worden gerealiseerd. Voor de online component kan een variëteit aan effectieve didactische methoden worden gebruikt, zoals: informatie met oefeningen en feedback, demonstratie video's, interactieve casuïstiek. Bij het ontwerpen van blended leren is de studiebelasting van cursisten een belangrijke overweging. Daarnaast is een goede borging van de voorbereiding door cursisten een kritische succesfactor. Theorieën over multimedia leren vormen een goede basis voor verder onderzoek.

Het tweede doel van dit proefschrift was de effectiviteit van en kritische ontwerpkenmerken bij serious games te onderzoeken voor de ontwikkeling van leerprestaties en motivatie. Voor de beoordeling van complexe vaardigheden acute zorg zijn checklists minder valide dan meer globale beoordelingsinstrumenten, zoals competentielijsten. Een serious game met realistische, interactieve patiëntcasuïstiek leidt bij aios tot betere complexe cognitieve vaardigheden, ondanks een beperkte zelfstudietijd. Alhoewel deze game voor onervaren studenten motiverend is en hen aanzet tot langer studeren, verbeteren hun cognitieve vaardigheden niet, in vergelijking met wat zij leren van een instructieve e-module. Deze gecombineerde resultaten duiden op een 'expertise reversal effect' tussen kennisniveau van de cursist (aios vs. student) en het instructieformat (game vs. e-module). Serious games, met hun mogelijkheid flexibel te oefenen met gesimuleerde taken, zijn vaak motiverend maar leiden niet altijd tot verbeterde leerprestaties. Een hoge natuurgetrouwheid van taken stimuleert motivatie (en mentale inspanning) nog extra, maar kan cognitieve overbelasting veroorzaken en bij onervaren studenten het leren belemmeren. Bij het ontwerpen van games kan de complexiteit voor de cursist worden beperkt door het aanbieden van begeleiding (uitgewerkte voorbeelden, tips) en het geleidelijk opvoeren van de moeilijkheidsgraad en natuurgetrouwheid van leertaken, met een variëteit aan problemen. Het motiverende karakter van games lijkt vooral samen te hangen met een uitdagende leeromgeving, betekenisvolle taken en controle door de cursist, hetgeen een gevoel van autonomie creëert. Uitdagende taken die passen bij het competentieniveau van de cursist kunnen tijdens het leren leiden tot een 'flow'-ervaring en optimaal leren. De ontwikkeling van een serious game vereist echter veel middelen. Voor de ontwikkeling van kennis is er weinig meerwaarde bij inzet van een game; andere educatieve formats zijn net zo goed in staat deze leerdoelen te realiseren tegen minder kosten.

Het spelen van serious games met realistische taken is vaak sterk motiverend voor cursisten, maar bevordert de leerprestaties niet noodzakelijkerwijs. Games kunnen het leren ondersteunen indien de complexiteit van de taken aansluit op het competentieniveau van de cursist. Meer ontwerpgericht onderzoek is nodig naar de effecten van taakcomplexiteit en didactische keuzes op de bevordering van leerprestaties en motivatie, voor minder en meer ervaren cursisten.



Dankwoord



De oorsprong van dit proefschrift ligt in een gesprek tijdens een autorit naar het gamebedrijf IJsfontein, rond februari 2012, waarin Stephanie vroeg: "Mary, ben jij eigenlijk gepromoveerd?" "Nee?" "Zou je willen promoveren, onderzoek doen naar het effect van de game die we hebben ontwikkeld?" "ehh… Ik zou wel willen onderzoeken of het werkt, maar promoveren? Een jaar of zeven alle vrije tijd besteden aan onderzoek doen?" We spraken af dat ik erover na zou denken. Een nieuw, uitdagend project sprak me wel aan, ik was ook erg nieuwsgierig naar de effectiviteit van de abcde*SIM* game die we hadden ontwikkeld met 'de jongens van IJsfontein' en ik had geen kinderen meer thuis … dus waarom niet? Toen Erik-Jan mij aanbood om 1 á 2 dagen per week aan dit promotieonderzoek te kunnen besteden, was de knoop doorgehakt. Het voelde een beetje als een avontuur, waarvan ik niet precies wist waar het zou eindigen. Nu, 3,5 jaar later, kijk ik terug op een periode van weliswaar hard werken, maar vooral de kans om veel nieuwe dingen te leren; soms met vallen en opstaan, soms door een cursus of met behulp van adviezen van collega's. Een nieuwe professionele impuls. In tegenstelling tot mijn verwachting ben ik na een paar jaar intensief met onderzoek naar serious games en blended leren bezig te zijn geweest, er nog lang niet op uit gekeken.

Dr.S.C.E. Schuit, beste Stephanie: veel dank voor je initiatief, voor de inspirerende sessies die we hadden, je onuitputtelijke energie en vindingrijkheid in het vinden van oplossingen. Je enthousiasme werkte aanstekelijk. Ook nadat je hoofd van de SEH en opleider Interne was geworden, vond je tijd om mijn stukken te lezen en van feedback te voorzien. Ik ben er trots op dat onze intensieve samenwerking - soms bij nacht en ontij -, die startte met het abcde*SIM* project, nu heeft geresulteerd in dit proefschrift.

Prof.dr.J.L.C.M. van Saase, beste Jan: veel dank voor de prettige begeleiding, het vertrouwen dat je me hebt gegeven en de vrijheid die ik kreeg in het onderzoek. Door jouw netwerk kreeg ik extra kansen om mijn werk te presenteren op medische congressen. Ik kijk met plezier terug op onze gesprekken over de vorderingen van het onderzoek en jouw ervaringen als opleider en internist.

Prof.dr.J.J.G. van Merrienboer, beste Jeroen: jouw grote expertise en ervaring in onderwijskundig onderzoek en instructie ontwerp was erg waardevol in alle fasen van onze studies. Veel dank voor de prettige samenwerking! Ondanks de vele promovendi die je begeleidt, was je steeds ongelooflijk snel en adequaat in je kritische en leerzame feedback.

Prof.dr.ir.A.P.N. Themmen, beste Axel: het Institute of Medical Education Research Rotterdam (iMERR), dat jij hebt helpen opzetten, was een goede voedingsbodem voor mijn onderzoek. Dank ook voor het beoordelen van het manuscript en je bereidheid plaats te nemen in de grote commissie.

Prof.dr.M. de Hoog, beste Mathijs: vanaf het begin van mijn promotietraject heb je interesse gehad in mijn onderzoek en meegedacht. Hopelijk gaan we aan het onderwerp online leren een vervolg geven voor de medische vervolgopleidingen. Dank voor het beoordelen van het manuscript en je bereidheid plaats te nemen in de grote commissie.

Prof.dr.M.P. Schijven, beste Marlies: we delen al enkele jaren de interesse in onderwijs ict, hebben nauw samengewerkt in commissies van de DSSH en aan artikelen. Ik vind de wijze waarop je technologie integreert met zowel je medische vak als met onderwijs inspirerend. Dank voor het beoordelen van het manuscript en je bereidheid plaats te nemen in de grote commissie.

Prof.dr.W.W. van den Broek, beste Walter: de onderzoeksgroep van iMERR, waar jij de eindverantwoordelijkheid voor hebt, was een goede voedingsbodem voor mijn onderzoek. Dank voor de samenwerking bij diverse onderwijs ict projecten en je deelname in de grote commissie.

Prof.dr.T.J. ten Cate, beste Olle: ruim vier jaar geleden hebben we samengewerkt in het strategieplan online leren voor het UMC Utrecht. Dank voor je waardevolle bijdrage aan het artikel over de Air Medic Sky-1 game over patiënt veiligheid en je bereidheid plaats te nemen in de grote commissie.

Prof.dr.T.van Gog, beste Tamara: dank voor je interesse in mijn onderzoek en je bereidheid deel te nemen aan de grote commissie. Wie weet komt het nog tot een samenwerking in de toekomst.

Prof.dr.H.A.P. Pols, beste Huib: je hebt steeds een grote betrokkenheid getoond bij de ontwikkeling van onderwijs ict. Eerst bij het Erasmus MC, en nu in je rol als rector bij de Erasmus Universiteit. Dat heb ik als zeer stimulerend ervaren, dank daarvoor. Ik voel me vereerd dat je vandaag de voorzitter van de promotiecommissie bent.

Prof.dr.G.R. Norman, dear Geoff: thank you for your help and expertise in the data-analysis for the assessment study and in reviewing the articles on assessment and simulation fidelity in games, it was very valuable for me and has greatly contributed to their improvement.

Prof.C.J. Kalkman, beste Cor: het was een genoegen met jou samen te werken aan het onderzoek naar de Air Medic Sky-1 game. Veel dank voor het ontsluiten van de game-data en je waardevolle bijdrage aan het artikel.

Erik-Jan: zonder de mogelijkheid een deel van mijn werktijd aan het promotieonderzoek te kunnen besteden, was dit boekje er niet geweest. Veel dank voor het vertrouwen dat je in me hebt gesteld en de ruimte die je medewerkers biedt.

Medewerkers en docenten van de SBOH/Schola Medica: dank voor het mogelijk maken van het onderzoek onder aios huisartsgeneeskunde en de hulp bij de uitvoering ervan. Kees Esser: hartelijk dank voor het faciliteren van het onderzoek. Frank Baarveld: dank voor het helpen motiveren van aios en hun opleiders om mee te doen aan het onderzoek, voor je voor je betrokkenheid en je bijdrage aan het assessment artikel. Frans Rutten: dank voor het meedenken met en meewerken aan het onderzoek, je bijdrage aan de ontwikkeling van het nieuwe beoordelingsformulier en aan zowel het artikel over de assessment als over de effectiviteit van de game voor aios. Rosé Vink: dank voor je efficiënte en prettige ondersteuning bij de organisatie van het onderzoek; jouw hulp was onontbeerlijk. Myrddin Stoffels, Stephanie Verwijmeren en Sarien Koppelman: dank voor jullie professionele ondersteuning bij het regelen van de ontelbare praktische zaken. Carlos Kok: dank voor het organiseren van de video-opnames. Docenten/specialisten Walter Henny, Els Jansen, Marty Luiting, Iscander Maissan, Paul Pahlplatz, Maria Peters, Lard Renalda, Dirk-Jan Versluis: veel dank voor de vele uren die jullie hebben besteed aan de videobeoordelingen; dit was een belangrijke basis voor het assessment onderzoek. Huisartsopleiding Nederland: Herman Duesman, dank voor het aanleveren van de toets gegevens en je interesse in het onderzoek.

Ik wil ook graag de aios huisartsgeneeskunde bedanken voor hun medewerking aan het onderzoek en aan de extra beoordelingen. Ook alle master studenten geneeskunde van het Erasmus MC ben ik erkentelijk voor hun medewerking aan het onderzoek en de beoordelingen en de tijd die ze hebben besteed aan een game en/of online leermateriaal. Zonder deze respondenten waren de studies niet mogelijk geweest.

Collega's van TNO: Esther Oprins, Maartje Bakhuys Roozeboom, Gillian Visschedijk en Gerard Veldhuis: dank voor de samenwerking in het onderzoek naar de effectiviteit van de game onder aios, en het mij op gang helpen met de data-analyse. Ik heb veel van jullie geleerd. Prof.dr.J.M.C.Schraagen, beste Jan Maarten: dank voor het inspirerende gesprek over onderzoek naar serious games in de beginfase van mijn traject en het kritisch meedenken met het beoordelingsformulier. Math Reintjens: dank voor de efficiënte data-invoer.

Ronald Nanninga en Lennard Goedknegt van abcdeSIM BV: heel fijn dat jullie steeds snel en goed de inlogcodes voor groepen studenten beschikbaar stelden, de vragenlijsten online plaatsten en hielpen bij het extraheren van de game-data. Veel dank!

Docenten/specialisten Erasmus MC: Jelmer Alsma en Els Jansen: veel dank voor de dagenlange vaardigheidsbeoordeling van geneeskunde studenten en voor de bijdrage aan het artikel. Gerrie Prins: bedankt voor het actief meedenken over de onderzoeksopzet van de Air Medic Sky-1 studie en je bijdrage aan het artikel.

Olivier Richters: dank voor de fijne samenwerking bij het onderzoek naar de Air Medic Sky-1 game; jouw master project. Het was leuk om te zien hoe jij als ervaren gamer kijkt naar de educatieve games. Zonder jouw grote wervingskracht hadden we niet zoveel studenten gehad bij de studie, veel dank voor je inzet en volharding bij het onderzoek.

Maurits Graafland, Agali Mert, Joep lagro, Laura de Wit-Zuurendonk en Alma Schaafstal: dank voor de prettige samenwerking aan het artikel over het systematisch evalueren van serious games.

Collega's van de Desiderius School: dank voor jullie belangstelling, steun en gezelligheid. Mathijs: bedankt voor het ontwerpen van de mooie omslag van het proefschrift en voor het reviewen. Karen: dank voor je goede adviezen, je bijdrage aan het assessment artikel en het reviewen. Susanne, Lokke, Wendy en Laura: dank voor de discussies, de suggesties en jullie interesse in mijn onderzoeksvorderingen. Willeke: dank voor je hulp bij de afronding van het promotietraject.

Collega's van de Zorg-Academie; Liesbeth van Ewijk-van den Bosch: dank voor je interesse in het onderzoek en de prettige samenwerking. Diane Storm en Irene Teeuwen: het artikel dat wij samen schreven over blended leren was een mooie start van het promotietraject. Bedankt voor de prettige samenwerking.

Collega's DCO; Ellen Thuis-de Graaf en Saskia van den Kieboom: dank voor de efficiënte aanlevering van studentgegevens.

Marjolijn: toen je 27 jaar geleden geboren werd, had ik niet verwacht dat ik nog eens een proefschrift zou schrijven, laat staan dat jij me zou helpen en goede feedback zou geven op sommige hoofdstukken. Ik vind het heel bijzonder dat je mijn paranimf bent. Jacinta: onze vriendschap begon ruim 35 jaar geleden, toen we beiden startten met de master onderwijskunde. Sinds die tijd hebben we samen veel meegemaakt. Onze gezamenlijke wandelingen en de concerten die we bezochten waren een zeer welkome afleiding van mijn onderzoekswerk. Dank voor je goede feedback en hulp bij de laatste loodjes. Ik ben heel blij dat je mijn paranimf wilt zijn.

Familie en vrienden: bedankt voor alle gezelligheid, belangstelling voor mijn vorderingen en steun. Ik hoop nu weer wat meer tijd te hebben om samen leuke dingen te ondernemen! Tennisvriendinnen/vrienden: de vele gezellige wedstrijden waren voor mij enorm ontspannend, mede daardoor kon ik het opbrengen om vervolgens weer achter mijn PC te kruipen.

Lieve mama: jij hebt me, samen met papa, altijd alle vrijheid gegeven te doen wat ik graag wilde, maar me ook gestimuleerd er wel hard voor te werken. Daar ben ik je nog steeds dankbaar voor.

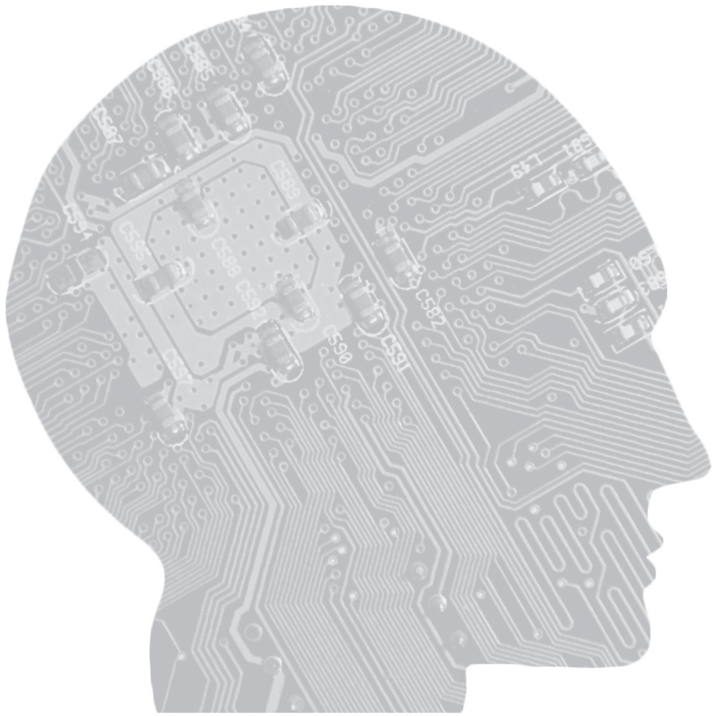
Lieve Ino, love of my life. Zonder jouw hulp bij veel onderdelen van het proefschrift en jouw onvoorwaardelijke steun was dit eindresultaat er niet geweest. Bedankt voor je humor, liefde en betrokkenheid bij alles wat ik doe.

Mary

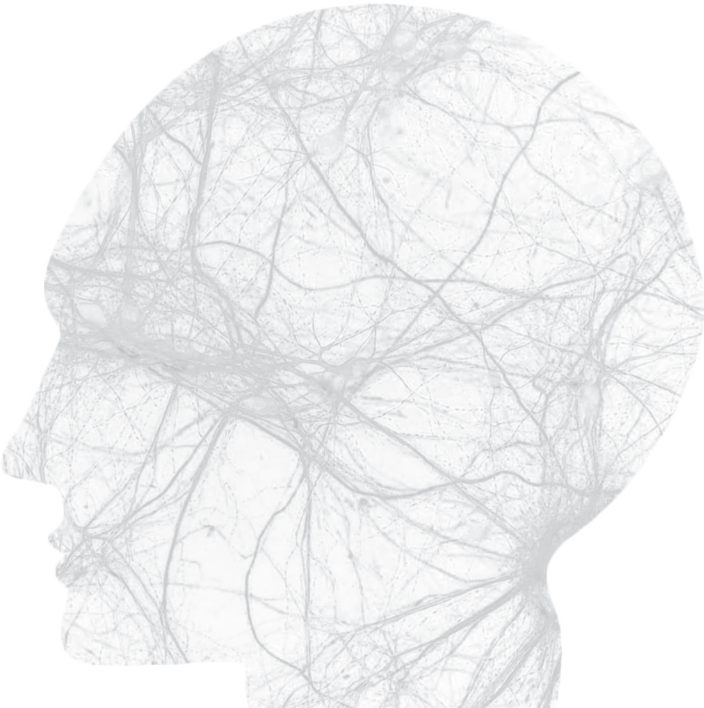
D

The research studies or printing of the thesis were supported by:

Dutch Society for
Simulation in Healthcare



Publications and presentations



PUBLICATIONS

Dankbaar MEW, Richters O, Kalkman CJ, Prins G, ten Cate TJ, van Merrienboer JJG, Schuit SCE. Comparative effectiveness of a serious game and e-module on patient-safety knowledge and awareness. *Submitted*

Oprins E, Visschedijk G, Bakhuys M, Dankbaar M, Trooster W, Schuit S. The Game-based learning Evaluation Model (GEM): measuring the effectiveness of serious games using a standardized method. *Accepted for Publication by the International Journal of Technology Enhanced Learning, 2015*.

Dankbaar MEW, Alsma J, Jansen EH, van Merrienboer JJG, van Saase JLCM, Schuit SCE. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Accepted for publication in Advances in Health Sciences Education, 2015 DOI: 10.1007/s10459-015-9641-x*

DankbaarMEW, Bakhuys Roozeboom M, Oprins EAPB, Rutten FL, van Merrienboer JJG, van Saase JLCM, Schuit SCE. Preparing residents effectively for emergency skills training with a serious game. *Submitted*

Dankbaar MEW, Stegers-Jager K, Baarveld F, van Merrienboer JJG, Norman GR, Rutten FL, van Saase JLCM, Schuit SCE. Assessing the assessment in emergency care training. *Plos-One* 2014; 18.

Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, Schaafstal A, Schijven M. How to systematically assess serious games applied to health care. *JMIR serious games* 2014; volume 2, issue 2.

Dankbaar MEW, Bakhuys Roozeboom M, Oprins EAPB, Rutten FL, van Merrienboer JJG, van Saase JLCM, Schuit SCE. *Games for Health 2014, Conf. Proceed*; Chapter 3 Gaming as a training tool to train cognitive skills in emergency care: how effective is it? 2014; Okt.

Dankbaar MEW, Storm DJ, Teeuwen IC, Schuit SCE. A blended design in acute care training: similar learning results, less training costs compared with a traditional format. *Perspectives on Medical Education* 2014; vol. 2, issue 2.

Dankbaar MEW, de Jong PGM. Technology for learning: how it has changed education. *Perspectives on Medical Education* 2014; vol. 2, issue 2.

Turk RFM, Dankbaar MEW, van Beeck EF. Computerondersteund samenwerkend leren in het coschap Sociale Geneeskunde, *Tijdschrift voor Medisch Onderwijs* 2009; Dec.

Dankbaar MEW. De effectiviteit van e-learning en de implementatie in het medisch onderwijs, *Tijdschrift voor Medisch Onderwijs 2009; Okt.*

Dankbaar M. Communities of practice: kenmerken en kennisontwikkeling. *Handboek Effectief Opleiden* 2006; juni.

Dankbaar M. Methoden voor het ontwikkelen, delen en vastleggen van kennis, *Handboek Effectief Opleiden 2003*; mei.

Dankbaar M, Oprins E. Visies op kennismanagement. *Handboek Effectief Opleiden 2002*; Okt.

Dankbaar M, Oprins E (red). *Kennismanagement; leerprocessen organiseren in werkomgevingen*. CINOP Expertisecentrum 2002; April.

ORAL PRESENTATIONS

Dankbaar MEW, Alsma J, Jansen EH, van Merrienboer JJG, van Saase JLCM, Schuit SCE. *What does a simulation game or text-based cases program add to an instructional e-module? Applying emergency skills on patient cases*. Presentation at the AMEE 2015 Conference in Glasgow, Great-Britain.

Dankbaar MEW, Alsma J, Jansen EH, van Merrienboer JJG, van Saase JLCM, Schuit SCE. *An experimental study on the cognitive and motivational effects of a simulation game*. Scientific presentation at the DSSH (Dutch Society for Simulation in Health Care) Conference March 2015, Alkmaar, the Netherlands. 2nd Price Best Scientific Presentation.

Dankbaar MEW, Stegers-Jager K, Baarveld F, van Merrienboer JJG, Norman GR, Rutten FL, van Saase JLCM, Schuit SCE. *De beoordeling beoordeeld: nadelen van checklists in de beoordeling vaardigheden* Scientific Presentation at NVMO 2014 Conference, Egmond aan Zee, the Netherlands.

Dankbaar MEW, Bakhuys M, Oprins E, Rutten F, van Saase JLCM, van Merrienboer JJG, Schuit SCE. *It's all in the game: Effectiviteit van serious gaming voor de ontwikkeling van* *vaardigheden acute zorg bij AIOS*. Scientific Presentation at NVMO 2014 Conference, Egmond aan Zee, the Netherlands.

Doets M, Zietse R. Dankbaar MEW. *Korte weblectures voor het medisch onderwijs: een aansprekend onderwijsmiddel*. Paper at the 2014 NVMO Conference, Egmond aan Zee, The Netherlands.

Dankbaar MEW, Bakhuys Roozeboom M, Oprins E, Rutten F, van Saase JLCM, van Merrienboer JJG, Schuit SCE. *Games as a training tool to prepare residents in emergency care skills: how effective is it*. Presentation at Games for Health Conference 2014, Utrecht, the Netherlands.

Dankbaar MEW, Bakhuys Roozeboom M, van Saase JLCM, van Merrienboer JJG, Schuit SCE. *It's all in the game: Effectiveness of a serious game for residents' emergency skills training*. Presentation at AMEE 2014 Conference Milan, Italy.

Dankbaar MEW, Bakhuys Roozeboom M, van Saase JLCM, van Merrienboer JJG, Schuit SCE. *Gaming as a tool to train cognitive skills in Emergency Care: how effective is it?* Presentation at Ottawa 2014 Conference Ottawa, Canada.

Dankbaar MEW, *Impact of online learning on health education: formats, design principles and effectiveness*. Presentation for the Institute for Medical Education Research (iMERR), April 2014, Rotterdam, the Netherlands.

Dankbaar MEW, Stegers-Jager K, Baarveld F, van Merrienboer JJG, Norman GR, Rutten FL, van Saase JLCM, Schuit SCE. *Validity and reliability of commonly used formats in assessing emergency skills*. Paper presented at EUSEM (European Society for Emergency Medicine) Conference 2014 Amsterdam, the Netherlands.

Pols HAP, Schuit SCE, Dankbaar MEW. *Revolutions in Medical Education: Transforming Education to Strengthen Healthcare*. Games for Health Conference 2013, Amsterdam, the Netherlands.

Hols-Elders WPM, Dankbaar MEW, de Jong PGM, Ballieux M, Verstegen D, Eggermont S, van der Woert N. *Technology for learning; how it has changed education*. Round table session from the special interest group E-Learning NVMO at the NVMO 2013 Conference, Egmond aan Zee, the Netherlands.

Vos JA, Dankbaar MEW, Doets M, Spierenburg E, Witkowska-Stabel M, de Bruycker N, Timmer C, Vos J, Willemsen A, Blom J, Meinema J, de Vries J. *Digitaal toetsen van klinisch redeneren*. NVMO 2013 Conference Egmond aan Zee, the Netherlands.

Dankbaar MEW. *Serious Gaming for Acute Medicine: Make Training more Efficient and Attractive*. Presentation at IAMSE 2013 Conference St.Andrews, Great Britain.

Dankbaar MEW, *Serious gaming in Medical Education: What's the use?* Invited Lecture for the Master Business Informatics, June 2013, University of Utrecht, the Netherlands.

Dankbaar MEW *The Flipped Classroom en ander Online Onderwijs*. Invited presentation on the Internal Medicine Conference April 2013, Maastricht, the Netherlands.

Dankbaar MEW, Schuit SCE. *The ABCDE game: how to prepare your residents at the ER playfully*. Paper presented at the DSSH Conference 2013, Amsterdam, the Netherlands.

WORKSHOPS

Kamphuis C, Dankbaar MEW, Lagro J. *Serious gaming in medical education. Let's play!* Workshop serious gaming at the AMEE 2015 E-Learning Pre-Conference Glasgow, Great Britain.

Verstegen D, Cunningham AM, Dankbaar MEW, de Jong PGM. What's up? Whatsapp? Technology in Medical Education. Pre-Conference Workshop at the AMEE 2014 Conference, Milan, Italy.

Lagro J, Kamphuis C, Schuit S, Dankbaar M, Graafland M, Schijven M. Serious gaming in het medisch onderwijs. Do's and don'ts. Workshop at the 2013 NVMO Conference, Egmond aan Zee, the Netherlands.

Dankbaar MEW, Verstegen D. Hoe de technologie het medisch onderwijs verandert. Invited Pre-Conference workshop for student representatives at the NVMO 2013 Conference, Egmond aan Zee, the Netherlands.

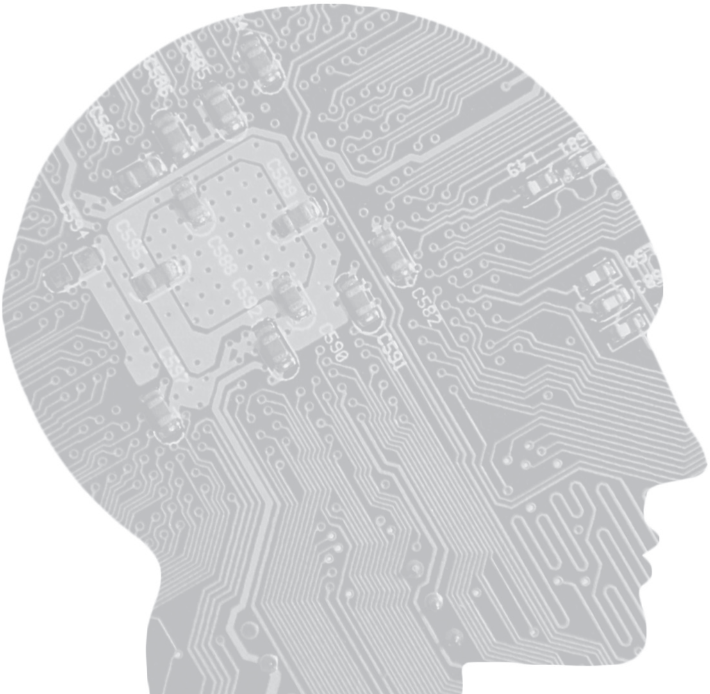
Lagro J, Schijven M, Kamphuis C, Schuit S, Dankbaar M. *Serious gaming in medical education: do's and don'ts*. Workshop at IAMSE Conference St.Andrews, Schotland. June 2013

Dankbaar M, van Bruggen L, Manrique M,Witlowska-Stabel M, de Bruycker N, Willemsen A. *Computer Based Assessment of Clinical Reasoning: how to develop effective questions? Pre-Conference Workshop at* Ottawa Conference 2012, Kuala Lumpur, Malaysia.

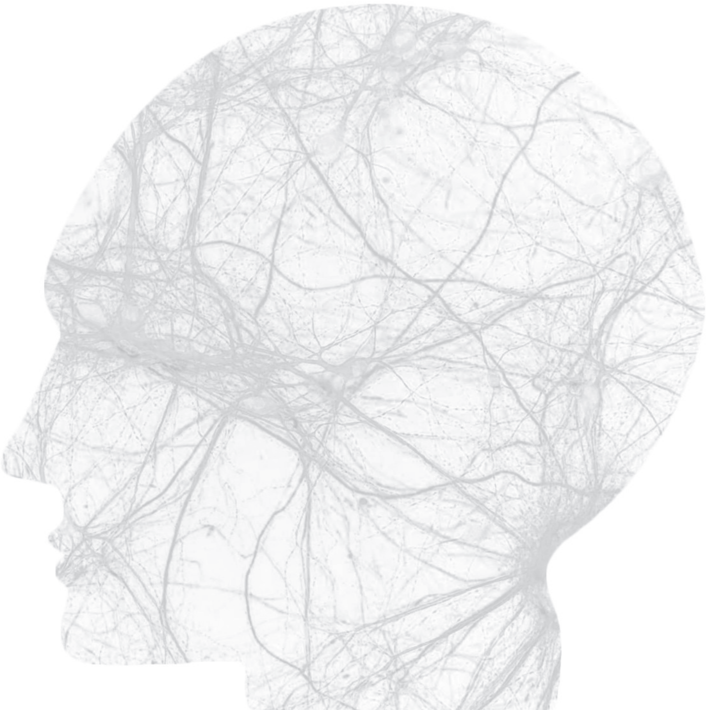
POSTER PRESENTATIONS

Dankbaar MEW, Bakhuys Roozeboom M, van Saase JLCM, van Merrienboer JJG, Schuit SCE. *Serious Games: An effective preparatory tool to train emergency care skills*. Poster presented at the Science Days of the Department of Internal Medicine Erasmus MC, 2014 January, Antwerp, Belgium.

Dankbaar MEW, Doets M, van den Broek WW, *Twitter tijdens colleges: een hip medium om de interactiviteit te verhogen?* Poster presentation at the 2013 NVMO Conference, Egmond aan Zee, the Netherlands.



Curriculum Vitae

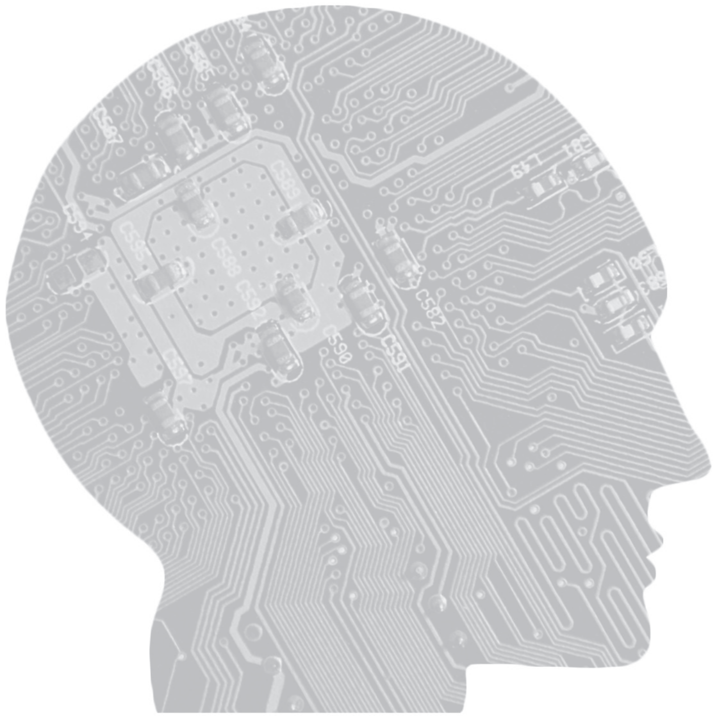


CURRICULUM VITAE

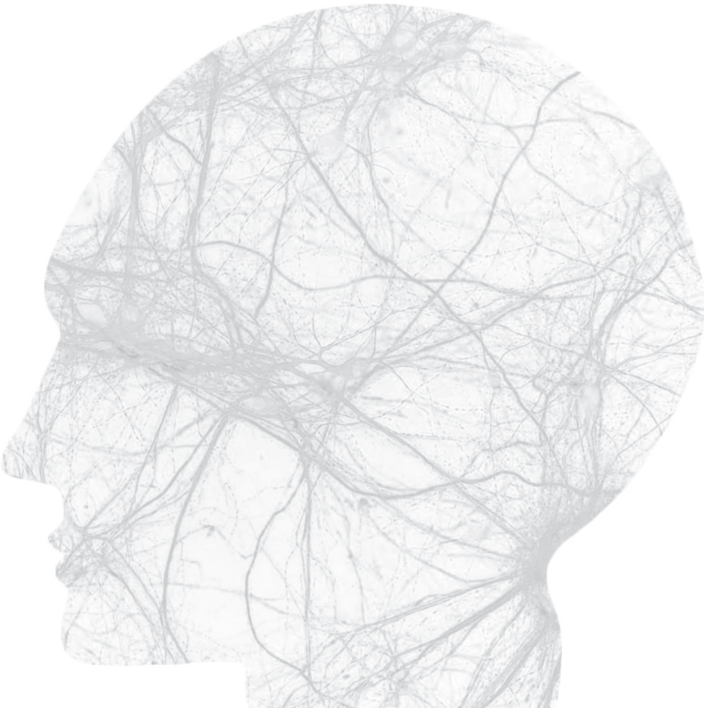
Mary Dankbaar was born in Lichtenvoorde, the Netherlands on September 9 in 1957. She received a Master in Educational Psychology at the Rijks Universiteit Leiden in January 1982. After her graduation, she worked at Teleac as an educational consultant. From August 1986, she spent a (sabbatical) year at AT&T Bell Labs and was involved in research on learning with multimedia. At Teleac, she subsequently worked as a projectmanager, developing multimedia courses for open net distance learning, and as a coordinator internet and new media. From 2000 until 2006, she took a job as a senior consultant at CINOP in Den Bosch (training & learning consultancy). She carried out projects on e-learning strategy, design and implementation for different companies, governance and educational organizations, and was a project leader for several European projects.

In 2006, she was appointed as a program manager e-learning at the Erasmus MC University Medical Center in Rotterdam. In collaboration with teachers, she has designed, implemented and evaluated a variety of blended and online programs for undergraduate and post-graduate medical curricula. She also has worked on the implementation of online assessment in different medical disciplines and has frequently trained teachers in designing online programs. She was the project manager for the development of the abcde*SIM* game, currently used by all starting residents at Erasmus Medical Center and in several other hospitals in the Netherlands and abroad. In 2013, she received, with her team members, the 'smarter learning' E-Learning award for this game. In September 2015, her team was nominated for the 'best educational innovation in medical education' prize by the board of the NVMO (Dutch association for Medical Education), for the blended ABCDE training design for residents.

Next to her work for Erasmus MC, she was project manager 'e-learning strategy' for the University Medical Center Utrecht (October 2010 - July 2011), designing an institution-wide strategy for online and blended learning (programma 'Onbegrensd leren'). From November 2014 until May 2015, she worked parttime for the Erasmus University as project manager 'virtual learning environment tender strategy' and developed, in collaboration with all faculties, an advice to acquire one shared virtual learning environment for the Erasmus University.
From 2008 - 2013, she was chair of the e-learning working group of the NVMO. Since 2012 she is an active member of the Committee on Scientific Research and the Committee on Serious Gaming of the DSSH (Dutch Society for Simulation in Healthcare). She published several articles in peer-reviewed journals on e-learning and a book on knowledge management, and is a frequent speaker at national and international conferences. In April 2012 she started her PhD study described in this thesis at the Institute for Medical Education Research Rotterdam (iMERR) at Erasmus MC.



PhD Portfolio



PhD PORTFOLIO

Summary of PhD training and teaching activities

Name PhD student: Mary E.W. Dankbaar	PhD period:	April 2012- August 2015
Erasmus University Medical Center	Promotor(s):	Prof. dr.J.L.C.M. van Saase
Institute for Medical Education Research Rotterdam		Prof. dr.J.J.G. van Merrienboer
	Co-promotor	Dr.S.C.E. Schuit

	Year	ECTS
1. PhD training		
General courses	2012-2015	6
SPSS, Tridata		
Biomedical English Writing and Communication		
Literature research		
Research design, methodology & instruments		
MOOC Statistics		
Strategies for writing and publishing		
Integrity in Research		
Specific courses		
Research in Medical Education (Onderzoek van onderwijs), Faculty of Health, Medicine and Life Sciences, Maastricht University	2012	1.5
MOOC Teaching online	2014	1.5
(Inter)national Conferences		
Oral presentations (online design, simulation/games, assessment; 15x)	2012-2015	12
Workshops on serious gaming and digital assessment (5x)	2012-2015	5.0
Attending conferences	2012-2015	8.5
2. Teaching activities		
Faculty training on instructional design for online learning	2012-2015	5.5
Workshops on the Flipped Classroom		1.0
Teacher's observation and feedback	2012-2014	1.5
Supervising Master's thesis		3.0
Invited lecture on Serious Games for Master Business Informatics (UU)		0.5
3. other activities		
Chair of the national working group on e-learning of the Dutch Association for Medical Education (NVMO)	2008-2012	
Member of the Committee on Scientific Research of the Dutch Society for Simulation in Healthcare (DSSH)	from 2012	
Member of the Committee on Serious Gaming of the Dutch Society for Simulation in Healthcare (DSSH)	from 2013	