ORIGINAL SCIENTIFIC REPORT

A Serious Game Can Be a Valid Method to Train Clinical Decision-Making in Surgery

Maurits Graafland · Maarten F. Vollebergh ·
Sjoerd M. Lagarde · M. van Haperen ·
Willem A. Bemelman · Marlies P. Schijven

Abstract

Background A serious game was developed to train surgical residents in clinical decision-making regarding biliary tract disease. Serious or applied gaming is a novel educational approach to postgraduate training, combining training and assessment of clinical decision-making in a fun and challenging way. Although interest for serious games in medicine is rising, evidence on its validity is lacking. This study investigates face, content, and construct validity of this serious game.

Methods Experts structurally validated the game's medical content. Subsequently, 41 participants played the game. Decision scores and decision speed were compared among surgeons, surgical residents, interns, and medical students, determining the game's discriminatory ability between different levels of expertise. After playing, participants completed a questionnaire on the game's perceived realism and teaching ability.

Results Surgeons solved more cases correctly (mean 77 %) than surgical residents (67 %), interns (60 %), master-degree students (50 %), and bachelor-degree students (39 % ($p < 0.01$). Trainees performed significantly better in their second play session than in the first (median 72 vs. 48 %, $p = 0.00$). Questionnaire results showed that educators and surgical trainees found the game both realistic and useful for surgical training. The majority perceived the game as fun (91.2 %), challenging (85.3 %), and would recommend the game to educate their colleagues (81.8 %).

Conclusions This serious game showed clear discriminatory ability between different levels of expertise in biliary tract disease management and clear teaching capability. It was perceived as appealing and realistic. Serious gaming has the potential to increase adherence to training programs in surgical residency training and medical school.

M. Graafland · S. M. Lagarde · W. A. Bemelman ·
M. P. Schijven (✉)
Department of Surgery, Academic Medical Centre,
PO Box 22600, Amsterdam 1100 DD, The Netherlands
e-mail: m.p.schijven@amc.uva.nl

M. F. Vollebergh · M. van Haperen
Department of Anaesthesiology, Academic Medical Centre,
Amsterdam, The Netherlands

Introduction

Modern surgical training curricula aim to create surgical experts in a time- and cost-efficient manner [1]. The difference between nonexperts and experts in surgery is the latter's ability to recognize and solve problems timely and adequately, the acquisition of which requires prolonged deliberate practice [2]. Patent safety concerns limit possibilities to train in everyday clinical practice. Therefore, educators seek alternative methods [2, 3]. Whereas

technical surgical skills may be trained efficiently using simulators [4], diagnostic reasoning and problem-solving (e.g., in the emergency department or ward) are largely learned in the work place as few alternatives are currently known to exist.

Serious games are "interactive computer applications … with challenging goals, that are fun to play, and supply users with skills, knowledge or attitudes useful in reality" [5]. A well-designed serious game combines psychological factors, design, and technology to engage learners in voluntary training. Games have many effects on the brain, most of which occur at a behavioral level. Serious gaming can influence the brain's adaptive neural plasticity, resulting in both structural [6] and functional [7] changes. Hence, whereas many may perceive video games as simply entertainment, they actually form a potent tool for learning. Players are required to develop and test hypotheses to overcome a given challenge relating to experiential learning [8, 9]. Serious games could thus serve as a powerful tool for cognitive enhancement in surgical training by mimicking clinical problem-solving. To date, games training clinical decision-making outside the operating room are novel, and evidence remains scarce [10].

A serious game was developed to train correct diagnosis and management of biliary tract disease. This serious game is based on a quiz game in which time-pressure and competition is used to trigger the intrinsic motivation of the player [11]. When assessing the value of a game as a serious training instrument, information on its validity is essential. This includes face validity (degree of perceived resemblance between medical constructs represented in the game and reality), content validity (degree to which game content covers the targeted medical construct), and construct validity (degree to which outcome parameters can measure differences between experts and novices) [12]. This study hypothesized that both surgeons and trainees would regard medical constructs to be represented realistically (face validity) and that surgeons would outperform trainees, indicating the game to be both robust and believable (construct validity).

Materials and methods

Participants

Surgeons, surgical residents and medical students from an academic hospital were recruited between September and November 2013. All 41 participants received a standardized instruction tutorial, which did not involve education of surgical content. Participants received a personal invitation with personal login code. Participants played a minimum of one gaming session, averaging five minutes of playtime. Scores were compared between groups of different levels

of expertise: *surgeons* (licensed surgeon), *residents* (surgical trainee), *interns* (has experience as physician at surgical wards, no surgical training), *master students* (medical student in surgical clerkship), and *bachelor students* (medical student, no clinical experience). After playing, participants filled out a questionnaire. Questionnaire results were linked to performance results through login code; results were assessed anonymously.

Serious game

The serious game Medialis (Little Chicken Game Co., Amsterdam, The Netherlands) contains cases on diagnosis and management of patients with biliary tract disease (Table 1). The medical content was validated prior to the study. Four surgical residents (postgraduate years 5 and 6) with experience in gastrointestinal surgery (>50 laparoscopic cholecystectomies as the primary surgeon) checked each case independently. Cases were evaluated and were marked as "valid" or "invalid." When marked invalid, cases were removed or corrected. The results were rechecked by two surgeons (M.S. and S.L.) until all were considered valid. The prototype was then tested for fluency, reading time, and clarity of imaging.

The serious game contained 97 cases: cholecystolithiasis (15), laparoscopic cholecystectomy (15), bile duct injuries (9), biliary pancreatitis (8), choledocholithiasis (5), cholecystitis (4), gallbladder carcinoma (5), adenomyomatosis (3), a variety of minor topics regarding biliary tract disease (33). Learning objectives concerned pathophysiologic and epidemiologic background, workup, treatment, and surgical management based on the residency teaching curriculum [13].

Cases consisted of an image, information describing a clinical problem, and possible solutions (Fig. 1). Participants solved as many cases as possible within one play session. The player had a maximum of 10 s to solve each case. After each attempt, players received feedback focused on their solution. Gaming mechanics included playing against time, competition, and sharing high scores among players on popular social platforms (leaderboards) to increase players' visibility and boost motivation. The latter two options were turned off during this research phase, securing players' anonymity. The performance of the participants was automatically measured through the following parameters: correct, incorrect, or neither correct/incorrect solution (+1, −1, 0 points, respectively); time required solving the case; total score; and total play time.

Questionnaire

The questionnaire (Google Docs; Google, Mountain View, CA, USA) contained items on demographic characteristics (10), realism (5), educational and testing value (8),

Table 1 Description of 3/97 cases

Case	Case description	Question	Answer options	Score
Example 1	Laparoscopy image shows enlarged, hyperemic gallbladder.	What is the diagnosis?	1. Cholecystitis	+1
			2. Porcelain gallbladder	−1
			3. Cystic duct	−1
			4. Liver hemangioma	−1
Example 2	Patient presents with pain in the right upper abdomen. Laboratory results reveal elevated alkaline phosphatase and aspartate transaminase. Bilirubin and C-reactive protein appear normal.	What is the diagnosis?	1. Cholecystolithiasis	+1
			2. Gastric ulcer	−1
			3. Cholangitis	−1
			4. Acute cholecystitis	−1
Example 3	MRI shows choledocholithiasis.	What is the standard treatment?	1. Laparoscopic bile duct exploration	0
			2. ERCP + papillotomy	+1
			3. Percutaneous drainage	−1
			4. All possible	−1

If solutions were correct, players received a point (+1). If solutions were incorrect, players lost a point (−1). In case of a question that had "best" and "next best" solving strategies, players did not gain or lose points when choosing the "next best" strategy (0)
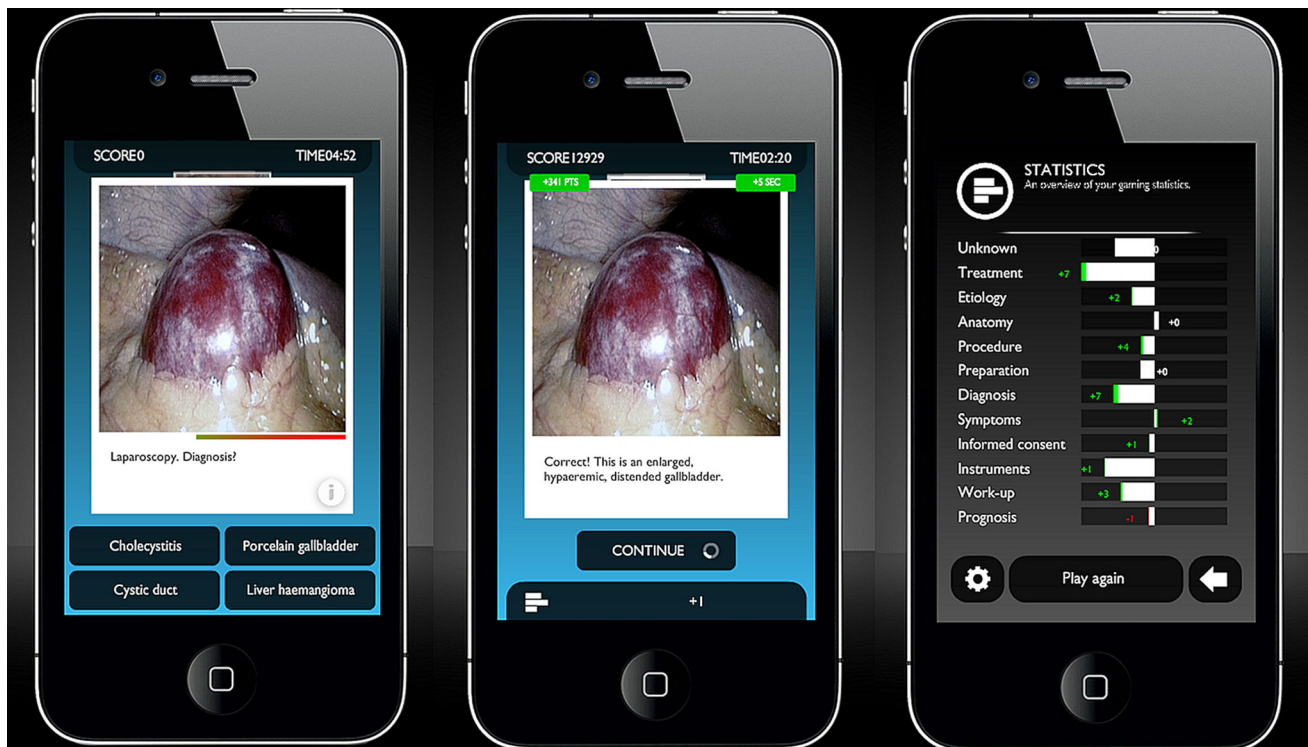

Fig. 1 Serious game (screenshots). The player is presented with a case (*left*) with four resolutions. After making a choice, the player is presented with feedback (*center*), points, and extra session time. After a session, the player can review his or her statistics (*right*)

perceived desirability and preferred user groups (7), perceived user experience (8) and game implementation (5). A medical psychologist checked the items on formulation and consistency. Participants could elaborate through open text boxes. Statements were scored on five-point Likert scales (1, disagree; 2, slightly disagree; 3, neither agree nor disagree; 4, slightly agree; 5, agree). Results were compared between three groups: expert (surgeons), surgical trainee (interns and residents), and novice (bachelor- and master-degree students). A median score >3.49 was considered a positive opinion toward the statement.

Statistical analysis

Performances of user groups with clinically relevant levels of expertise were compared (surgeons, surgical residents,

interns, master-degree students, bachelor-degree students). Proportions of correctly solved cases and mean case time were compared with parametric and nonparametric tests, respectively. Score improvements during repeated sessions were compared through Wilcoxon signed rank testing. IBM Statistical Package for Social Sciences version 20 was used (IBM, Armonk, NY, USA).

Results

Characteristics

The 41 participants completed a median of 70 cases (interquartile range (IQR) 39–87). Demographic characteristics are described in Table 2. Age differences between the groups were statistically significant (one-way ANOVA with post hoc Bonferroni, $p < 0.01$). A significantly higher proportion of master and bachelor students were female (Chi square compared to other groups, $p = 0.03$). There were no significant differences in videogame experience between groups.

Face validity

Overall, 34 participants completed the face validity questionnaire (one trainee and five students failed to complete the questionnaire). The majority found the presentation of images, radiology, and clinical situations realistic (88.2, 97.1 and 91.2 %, respectively) (Table 3). In total, 60.6 % stated that decisions in the serious game are based on realistic elements, and 66.7 % found their clinical experience to be helpful.

The majority believed the serious game to be useful for learning disease background (77.6 %), decisions during the workup and treatment (82.4 %), perioperative decisions (64.7 %), knowledge on medical technology (55.9 %), and risk management (53.0 %) (Supplementary Table 1). Participants found it useful for testing clinical decision-making (76.5 %) and for monitoring trainees' progress (64.7 %). The majority considered training with this serious game desirable (77.5 %) and more fun than classic training (94.1 %) (Supplementary Table 2) 0.2 The majority of surgeons and residents found it useful for training students (58.9 %) and residents (94.1 %). Surgeons found the serious game also useful for training surgeons (100 %). Interestingly, surgical residents disagreed with surgeons on this item (Kruskal–Wallis, $p = 0.02$).

The majority of gamers had enjoyed the game, considering it fun (91.2 %) and challenging (85.3 %). Furthermore, 44.1 % said they felt involved during game play (Supplementary Table 3). The majority did not feel frustrated (55.9 %), or bored (79.4 %), and 44.1 % did not feel easily distracted. Experts and trainees found the serious game less frustrating than did novices (Kruskal–Wallis, $p = 0.07$).

The majority of participants considered the serious game an addition to regular surgical training (79.4 %), although not in an obligatory fashion (Supplementary Table 4). The majority (81.8 %) would recommend it to colleagues, and 58.8 % would play in their free time. A majority (55.9 %) considered it a useful contribution to patient safety.

Surgeons were overall more positive [median 3.89, interquartile range (IQR) 0.47] about the game than residents (median 3.57, IQR 0.39) or students (median 3.68, IQR 0.82) (Kruskal–Wallis, $p = 0.04$). The student group showed a wider range of attitudes than other groups.

Construct validity

All participants completed the first play session, 22 completed a second (5 residents, 2 interns, 8 master-degree

Table 2 Demographic characteristics of the participants

Group characteristics	Bachelor	Master	Intern	Resident	Surgeon
Group size	8	14	6	6	7
Age (mean ± SE)	20.7 ± 0.9	23.9 (±0.5)	28.1 (±0.5)	31.0 (±0.7)	40.3 (±2.1)
Gender					
Male	33 %	31 %	67 %	83 %	71 %
Female	67 %	69 %	33 %	17 %	29 %
No. of patients with biliary tract disease treated (*n*)					
0	100 %	77 %	33 %	0 %	0 %
1–100	0 %	23 %	67 %	100 %	29 %
>100	0 %	0 %	0 %	0 %	71 %
Video-gamer					
Last month	50 %	30 %	0 %	20 %	40 %
Past	83 %	40 %	60 %	100 %	100 %

The groups were combined into students, trainees, and experts for the face validity analysis

Table 3 Participants' opinions on representation of important medical constructs in the serious game (medians in bold)

Representation	Student (n = 17)			Trainee (n = 11)			Expert (n = 6)			p*
	Median	P25	P75	Median	P25	P75	Median	P25	P75	
Images	**5.0**	4.0	5.0	**4.0**	4.0	5.0	**4.5**	3.8	5.0	
Radiology	**5.0**	4.0	5.0	**4.0**	4.0	5.0	**4.0**	4.0	4.3	
Clinical situations	**4.0**	4.0	5.0	**4.0**	4.0	4.0	**4.0**	4.0	5.0	
Decisions are based on realistic cues	**3.0**	2.5	4.0	**4.0**	4.0	4.0	**4.0**	4.0	4.3	0.008
My experience helps me to solve cases	**3.0**	2.0	4.0	**4.0**	4.0	5.0	**5.0**	4.8	5.0	0.001

*Kruskal–Wallis test

students, 7 bachelor-degree students), and 10 completed a third (3 residents, 1 intern, 3 master-degree students, 3 bachelor-degree students).

Figure 2 shows the combined proportion of correct and not correct/incorrect choices during first session for each study group. Surgeons (mean 0.77, SD 0.09) solved more cases than residents (mean 0.67, SD 0.05), interns (mean 0.60, SD 0.09), master-degree students (mean 0.50, SD 0.10), and bachelor-degree students (mean 0.39, SD 0.03). One-way analysis of variance (ANOVA) with post hoc Bonferroni correction showed differences between surgeons versus interns, master-degree students, and bachelor-degree students to be statistically significant ($p < 0.01$). Differences between residents versus master-degree students and bachelor-degree students were statistically significant ($p < 0.01$), as were differences between interns and bachelor-degree students and differences between master-degree and bachelor-degree students ($p = 0.00$ and $p = 0.035$, respectively).

Figure 3 shows the average time required by participants to solve first-session cases. Surgeons required a median of 8.47 s (IQR 0.72), residents 8.41 s (3.16 IQR), interns 7.71 s (4.02 IQR), master-degree students 8.99 s (1.66 IQR), and bachelor-degree students 8.20 s (2.26 IQR). The differences were not significant (Kruskal–Wallis, $p = 0.73$).

Figure 4 shows that participants improved their performance during their second gaming session (median proportion of 0.72 (IQR 0.23) compared to 0.48 (IQR 0.22) during the first (Wilcoxon signed rank, $p = 0.000$)). Intragroup comparison showed a statistically significant improvement within two sessions for residents [median 0.66 (IQR 0.10) in session 1 and 0.84 (0.21) in session 2; $p = 0.043$]; master-degree students [median 0.48 (IQR 0.19) in session 1 and 0.77 (0.13) in session 2; p = 0.012]; bachelor-degree students [median 0.38 (IQR 0.04) in session 1 and 0.56 (0.23) in session 2; $p = 0.043$].

Discussion

Valid serious games have the potential to shorten surgical trainees' learning curves in clinical reasoning and problem-
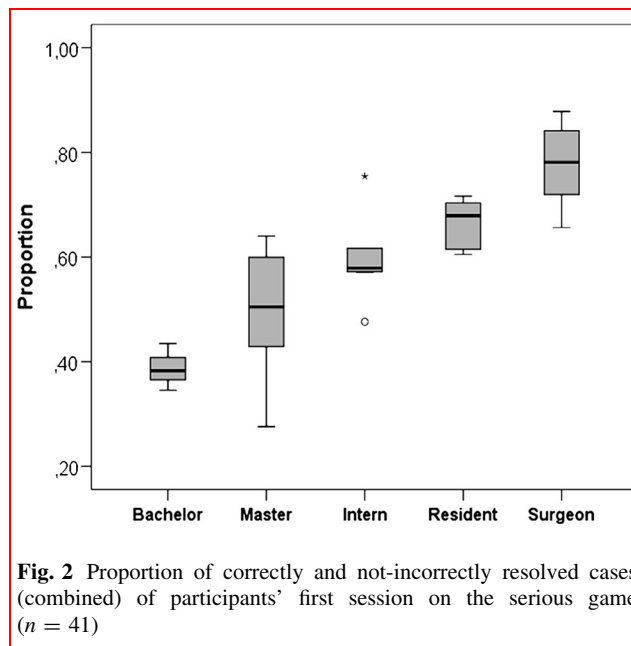
Fig. 2 Proportion of correctly and not-incorrectly resolved cases (combined) of participants' first session on the serious game (n = 41)
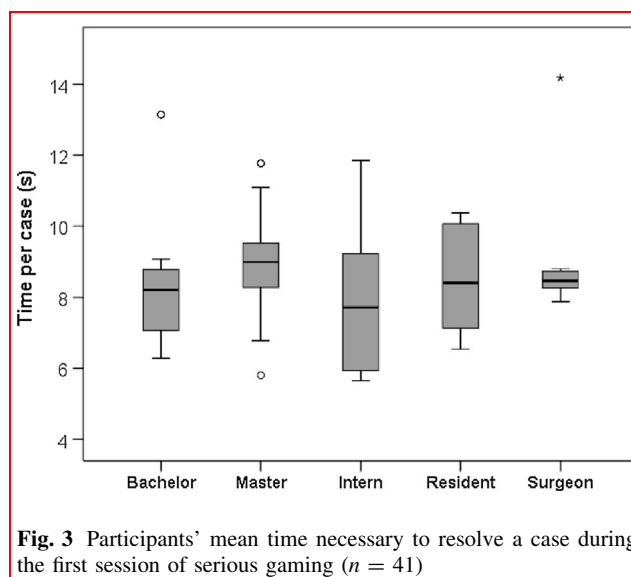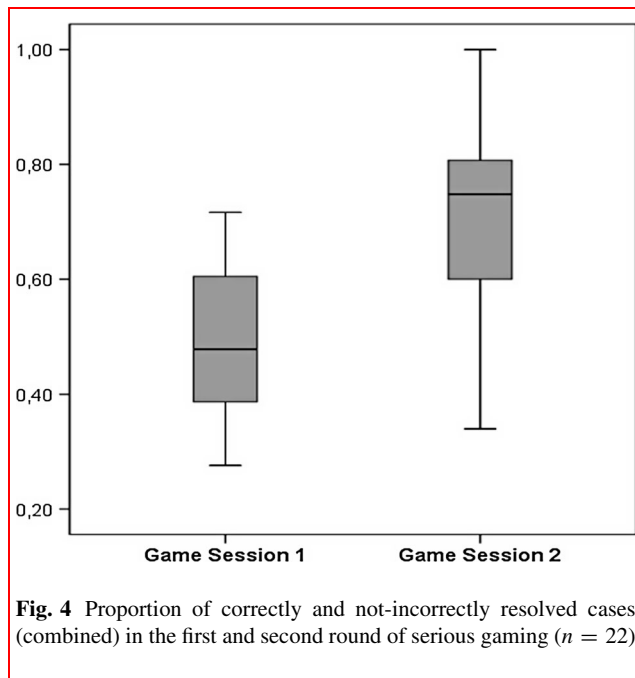
Fig. 3 Participants' mean time necessary to resolve a case during the first session of serious gaming (n = 41)

Fig. 4 Proportion of correctly and not-incorrectly resolved cases (combined) in the first and second round of serious gaming ($n = 22$)

solving, increasing time efficiency in surgical training. They provide direct feedback on decisions made by the player by distributing rewards and punishments. They assist in learning at different paces and allow practice to the point of mastery and automaticity. Repetition is a precondition for long-term potentiation. Whereas gamers repeat actions as they play, their strengthening of synaptic connections induces memory storage and learning. Medialis is the first serious game to show face, content, and construct validity for training clinical decision-making relevant to surgeons outside the operation theater. Quality of decision-making appears to be a valid assessment parameter in the game (proportion correct), whereas the time required to solve cases is not.

According to the study results, serious gaming could be of considerable use to younger surgical residents, enhancing their level of functioning and saving valuable time that can be spent otherwise, such as in surgery itself. Moreover, structured assessment of skills and competencies is becoming more relevant to modern curricula. New curricula have introduced systems of entrustable professional activities and statements of awarded responsibility to measure and assess trainee's competencies in specific clinical activities [14, 15]. A serious game that validly measures specific levels of competencies could facilitate this entrustment process. Such serious games could also play a role in training and recertification of surgical specialists.

Learning in serious games is based on two principles. Gameplay closely mimics the problem-solving ability required in diagnostic and therapeutic reasoning (i.e., testing and readjusting hypotheses), thereby leading to experiential learning [8, 9]. Learning in this study occurs over time, and participants significantly increase their scores in consecutive sessions. Second, high levels of engagement in gameplay are thought to draw players into a state of "flow" [9]. As such, players are completely absorbed, ignoring all external stimuli and focusing their attention solely on the gameplay and thus on prolonged deliberate practice [9]. Flow occurs mostly when the level of difficulty is adjusted optimally to the level of skill of the player. Results show that this serious game complies with these principles, as 44 % of participants feel engaged when gaming (67 % of the surgeons) even though two of the main competitive game mechanics were turned off during the study. Of the students, 8 of 17 found the game frustrating (compared to 1 of 11 trainees and 1 of 6 experts). According to the "flow theory," the level of difficulty is too high for some students, which could be adjusted by incorporating difficulty levels.

The majority of surgical residents and surgeons have experience with video games and are therefore likely to embrace using a video game in surgical training. Players are enthusiastic about dealing with everyday medical content in an engaging way. Positive statements by experienced users indicate that serious games have the potential to overcome the appreciation problems of simulators in surgical training. Although generally considered to be both efficient and attractive for teaching surgical skill, residents do not practice voluntarily on simulators in their free time [16].

Serious games currently experience a development surge, even though randomized trials on learning outcome of individual games remain scarce [10, 17]. Recent studies show promising use of game mechanisms optimizing adherence to surgical skills training. Verdaasdonk et al. showed that "gamifying" simulator exercises by adding real-time competition and reward systems significantly increased exposure time and determination to play in 31 surgical professionals [11]. Badurdeen et al. showed that performance of 20 surgical trainees on commercially developed, off-the-shelf video games with motion-sensing controllers (Nintendo Wii[TM]) correlated with laparoscopic simulator performance [18]. Jalink et al. showed that altering the software of laparoscopic simulators into a custom-made video game environment, controlled by laparoscopic handles, shows similar correlations [19]. Next to training technical skills, this study is the first to describe a serious game's value for training clinical disease-specific decisions outside the operation theater.

Limitations of this study include the following. Analysis of subjective opinions may induce a Hawthorne effect. Exposure to questioning could have affected participants' behavior and thus manipulated results [20]. By distributing questionnaires and game access online anonymously,

social desirability effects have been partially dealt with. Second, a power analysis could not be performed beforehand because an estimated effect size from previous research was lacking. Therefore, lack of power cannot be excluded. The magnitude of differences in performance and limited variances within experienced groups ensured statistical significant differences in this study.

Third, a limitation of the face validity test is the relatively high loss of participants who failed to complete the questionnaire (one expert, one trainee, and five students). In face validity testing, the opinions of experts and trainees are considered most important. Fallout in these groups was relatively limited. Sex differences between the groups are in conformity with the proportions in the general medical student population [21], the surgical trainee population, and surgeons [11, 19].

Conclusions

This study is the first of its kind to demonstrate face, content, and construct validity of a medical serious game on clinical decision-making. It indicates the potential ability of Medialis to train and assess surgical trainees and professionals in an entertaining way. Future research should determine long-term sustainability of learning outcomes and transfer of skills to clinical practice. Considering the enthusiasm of licensed surgeons, the alleged "digital gap" between surgical residents and surgeons on mHealth applications appears to be closing fast.

Ethical approval This study did not require approval of the local research ethics committee.

Conflict of interest The authors have no conflicts of interest or financial ties to disclose.

References

1. Dunnington GL, Williams RG (2003) Addressing the new competencies for residents' surgical training. Acad Med 78:14–21

2. Andersen DK (2012) How can educators use simulation applications to teach and assess surgical judgment? Acad Med 87:934–941

3. Bell RH (2009) Why Johnny cannot operate. Surgery 146:533–542

4. Thijssen AS, Schijven MP (2010) Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? Am J Surg 199:529–541

5. Michael DR, Chen S (2006) Serious games: games that educate, train, and inform, 1st edn. Thomson Course Technology, Boston

6. Kühn S, Romanowski A, Schilling C et al (2011) The neural basis of video gaming. Transl Psychiatry 1:e53

7. Anguera JA, Boccanfuso J, Rintoul JL et al (2013) Video game training enhances cognitive control in older adults. Nature 501:97–101

8. Kolb D (1984) Experiential learning: experience as the source of learning and development. Prentice Hall, Englewood Cliffs

9. Kiili K (2005) Digital game-based learning: towards an experiential gaming model. Internet High Educ 8:13–24

10. Graafland M, Schraagen JM, Schijven MP (2012) Systematic review of serious games for medical education and surgical skills training. Br J Surg 99:1322–1330

11. Verdaasdonk EGG, Dankelman J, Schijven MP et al (2009) Serious gaming and voluntary laparoscopic skills training: a multicenter study. Minim Invasive Ther Allied Technol 18:232–238

12. Gallagher AG, Ritter EM, Satava RM (2003) Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 17:1525–1529

13. Hamming J, Borel Rinkes IHM, et al (2009) Structured Curriculum for Surgery for Reflective Professionals. [Structuur Curriculum Heelkunde voor Reflectieve Professionals (Dutch). Opleidingsplan Heelkunde (Internet). Dutch Surgical Society (Nederlandse Vereniging voor Heelkunde).] Available at: http://knmg.artsennet.nl

14. Hauer KE, Kohlwes J, Cornett P et al (2013) Identifying entrustable professional activities in internal medicine training. J Grad Med Educ 5:54–59

15. Boyce P, Spratt C, Davies M et al (2011) Using entrustable professional activities to guide curriculum development in psychiatry training. BMC Med Educ 11:96

16. Chang L, Petros J, Hess DT et al (2007) Integrating simulation into a surgical residency program: is voluntary participation effective? Surg Endosc 21:418–421

17. Akl EA, Kairouz VF, Sackett KM et al (2013) Educational games for health professionals. Cochrane Database Syst Rev 3:CD006411

18. Badurdeen S, Abdul-Samad O, Story G et al (2010) Nintendo Wii video-gaming ability predicts laparoscopic skill. Surg Endosc 24:1824–1828

19. Jalink MB, Goris J, Heineman E et al (2014) Construct and concurrent validity of a Nintendo Wii video game made for training basic laparoscopic skills. Surg Endosc 28:537–542

20. Schwartz D, Fischhoff B, Krishnamurti T et al (2013) The Hawthorne effect and energy awareness. Proc Natl Acad Sci USA 110:15242–15246

21. Anonymous (2012) Feminisation in the workplace (Dutch). In: Arts en Auto (magazine). Utrecht, VvAA Association, pp 16–19. Available at: http://www.artsenauto.nl/wp-content/uploads/2012/07/AA08-p016–019.pdf