

Original Paper

Medical Student Evaluation With a Serious Game Compared to Multiple Choice Questions Assessment

Julien Adjedj^{1,2,3}, MD; Gregory Ducrocq^{4,5,6}, MD, PhD; Claire Bouleti^{4,5,6}, MD, PhD; Louise Reinhart², GradDipHlthSc; Eleonora Fabbro³, MSc; Yedid Elbez^{6,7}, MSc Biostat; Quentin Fischer¹, MD; Antoine Tesniere^{2,3,8}, MD, PhD; Laurent Feldman^{4,5}, MD, PhD; Olivier Varenne^{1,2,3}, MD, PhD

¹AP-HP, Hôpital Cochin, Cardiology, Paris, France

²Université Paris Descartes, Paris, France

³iLUMENS, Department of Simulation, University of Sorbonne Paris Cité, Paris, France

⁴Université Paris Diderot, Paris, France

⁵AP-HP, Hôpital Bichat, Cardiology, Université Paris Diderot, Paris, France

⁶FACT, French Alliance for Cardiovascular Trials, DHU FIRE, Cardiology department of Bichat hospital, Université Paris Diderot, Paris, France

⁷URC-Est, AP-HP, Paris, France

⁸AP-HP, Hôpital Cochin, Anesthesiology, Paris, France

Corresponding Author:

Olivier Varenne, MD, PhD

AP-HP, Hôpital Cochin

Cardiology

27 rue du Faubourg Saint Jacques

Paris,

France

Phone: 33 158412750

Email: olivier.varenne@aphp.fr

Abstract

Background: The gold standard for evaluating medical students' knowledge is by multiple choice question (MCQs) tests: an objective and effective means of restituting book-based knowledge. However, concerns have been raised regarding their effectiveness to evaluate global medical skills. Furthermore, MCQs of unequal difficulty can generate frustration and may also lead to a sizable proportion of close results with low score variability. Serious games (SG) have recently been introduced to better evaluate students' medical skills.

Objectives: The study aimed to compare MCQs with SG for medical student evaluation.

Methods: We designed a cross-over randomized study including volunteer medical students from two medical schools in Paris (France) from January to September 2016. The students were randomized into two groups and evaluated either by the SG first and then the MCQs, or vice-versa, for a cardiology clinical case. The primary endpoint was score variability evaluated by variance comparison. Secondary endpoints were differences in and correlation between the MCQ and SG results, and student satisfaction.

Results: A total of 68 medical students were included. The score variability was significantly higher in the SG group ($\sigma^2=265.4$) than the MCQs group ($\sigma^2=140.2$; $P=.009$). The mean score was significantly lower for the SG than the MCQs at 66.1 (SD 16.3) and 75.7 (SD 11.8) points out of 100, respectively ($P<.001$). No correlation was found between the two test results ($R^2=0.04$, $P=.58$). The self-reported satisfaction was significantly higher for SG ($P<.001$).

Conclusions: Our study suggests that SGs are more effective in terms of score variability than MCQs. In addition, they are associated with a higher student satisfaction rate. SGs could represent a new evaluation modality for medical students.

(*JMIR Serious Games* 2017;5(2):e11) doi: [10.2196/games.7033](https://doi.org/10.2196/games.7033)

KEYWORDS

serious game; multiple choice questions; medical student; student evaluation

Introduction

Student evaluation is one of the most important components of a medical educational program and is used for training and for validating degrees and career options. If handled well, it can improve student motivation for learning and provide educators useful feedback. Medical education cannot be limited to book-based knowledge which is defined as the ability to provide an answer from medical literature [1,2]. It also comprises developing medical skills such as the ability to act to obtain medical data and provide good care to patients [3]. Therefore, given the importance of questioning and the deductive process required to reach the right diagnosis and prescribe the right treatment, proper evaluation modalities are needed based on both book-based knowledge and diagnostic skills. High score variability, defined as the highest score in points obtained between students, is also mandatory to allow representative classification and fair career access based on test results in large student populations. On the other hand, the evaluation modality should also allow for an objective, fast, and inexpensive correction. As such, multiple choice questions (MCQs) are currently the most frequently used modality. Medical serious games (SG), based on virtual reality, are emerging as an alternative way of evaluating medical education [4]. However, they have not yet been evaluated in terms of score variability. In this study, we sought to evaluate medical students' test results with an SG compared with MCQs in terms of score variability, score difference, correlation between scores in MCQs and SG, student satisfaction, and finally whether SGs could be of use to learn and evaluate medical skills for medical students.

Methods

Study Design

From January to September 2016, we included all volunteer medical students with previous cardiology validation in two

medical schools (University Paris Descartes, Paris, France and University Denis Diderot, Paris, France). Students were randomized in a cross-over design between two groups to avoid order bias. Group 1 started with evaluation by SG and finished with evaluation by MCQs and group 2 performed alternatively. The tests were performed in the examination centers of both medical schools. Both tests lasted 30 minutes and the tests were performed consecutively. The study was approved by the educational committee of both institutions. All students gave their informed consent before inclusion.

Serious Game

We used a clinical case from an SG (Medusims, Paris, France and iLUMENS, Medical Simulation Department, Université Sorbonne Paris Cité, Paris, France). The SG focuses on the management of atrial fibrillation. It represents a cardiologist and a patient within a free tridimensional (3D) environment within a medical office, and is available on computers and tablets (Figure 1). Students play the role of the cardiologist and ask the patients questions using key words, perform a complete clinical examination with electrocardiogram, and require the prescription of additional tests and medical treatment. Points are awarded if the student asks the patient a correct question or performs the appropriate physical examination. There are no negative points due to wrong answers. Besides the free conduct of the clinical questioning, pop-up questions also arise in the SG during electrocardiogram interpretation, risk score calculation and the potential medical treatment in form of MCQs. Points are also awarded for correct answers to these pop-up questions. An automatic and precise correction is given to the student at the end of the game. Results are expressed out of a total of 100 points divided into four subcategories: clinical examination out of 25 points, diagnosis out of 25, risk score calculation out of 30 and medical decision out of 20 (Figure 1).

Figure 1. Serious game illustration (in French).



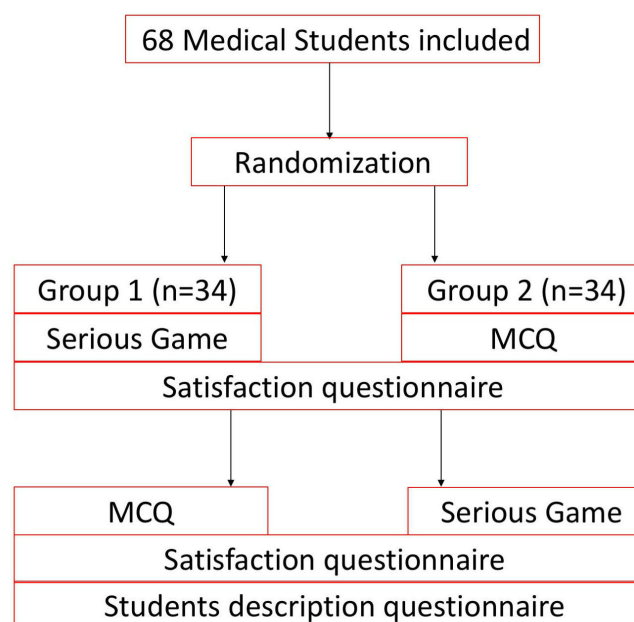
Multiple Choice Questions

We built an online MCQ test of 15 questions based on the SG clinical case with the same clinical and electrographic presentation. Each MCQ presented five possible answers. The student scored full points if all the selected answers were correct, 50% if one answer was incorrect, and 20% if two answers were incorrect. No points were awarded if three or more answers were incorrect. The correction was aligned to the SG correction giving a final score out of 100 points. A translated version of the MCQ test is available in the [Multimedia Appendix 1](#).

Satisfaction and Student Description Questionnaires

Questionnaires to record student characteristics and satisfaction were designed by a psychologist from the Medical Simulation Department of University Paris Descartes (iLUMENS, Paris, France). The student satisfaction questionnaire was filled in immediately after each evaluation using website. The student characteristics questionnaire was filled in online at the end of the study protocol to assess the medical degree and whether the student played video games regularly at the time of the study (gamers) or not (non-gamers; [Figure 2](#)).

Figure 2. Flowchart.



Objectives and Endpoints

The primary objective of the study was to compare the students' scores at the MCQs and SG tests. The related primary endpoint was therefore the score variability calculated as a variance for each test. The secondary endpoints were student satisfaction with semiquantitative questions expressed from 0 (no, not at all) to 5 (yes, entirely) and the correlation between the test results. Subgroup analysis was performed for SG results between gamers and non-gamers.

Statistical Analysis

Summary descriptive statistics are reported as mean and standard deviation, median (inter quartile range), or counts (%), as appropriate. We used the *t* test and Mann-Whitney test to evaluate the difference of continuous variables as appropriate. Fisher exact test was used for variance comparison. For score differences, the paired *t* test was used. Finally, the correlation coefficient was calculated between the results of SG and MCQs using the Pearson R^2 correlation test. All analyses were

performed with SPSS 21.0 (IBM Inc), R 3.3.1 (R Project for Statistical Computing) and Prism GraphPad 7.0 (GraphPad Software Inc).

Results

Main Student Characteristics

A total of 68 medical students were included (34 in each group), of which 29 were male (43%) and the mean age was 23(SD 1) years. Students were in their 5th [4-6] year of medical school. All the students owned a cellphone and a personal computer, and subscribed to an Internet connection; 31 (46%) owned a tablet and 21 (31%) a video game console. A past experience of video games was reported by 60 (88%) of students and 22 (32%) were currently playing video games for an average duration of 1.6(SD 3.0) hours per week. The main characteristics of the population according to the allocated group of randomization are detailed in Table 1. There were no significant differences in student characteristics between groups 1 and 2.

Table 1. Student characteristics.

| Student description N=68 | Overall | Group 1 n=34 | Group 2 n=34 | Comparison between groups 1 and 2 (<i>P</i> value) |
|--|-----------|-----------------|-----------------|--|
| Sex (male), n (%) | 29 (43) | 17 (50) | 12 (35) | .22 |
| Age in years, mean (SD) | 23 (1) | 23 (1) | 23 (1) | .26 |
| Year of medical school, mean (SD) | 4.7 (1.0) | 4.7 (0.8) | 5.1 (0.9) | .08 |
| Cardiology internship within the past 12 months, n (%) | 34 (50) | 16 (47) | 18 (53) | .74 |
| Owns a cell phone with Internet connection and social network account, n (%) | 67 (99) | 33 (97) | 34 (100) | >.99 |
| Owns a tablet, n (%) | 31 (46) | 13 (39) | 18 (54) | .20 |
| Owns a computer with Internet connection possession, n (%) | 68 (100) | 34 (100) | 34 (100) | >.99 |
| Owns a video game console, n (%) | 21 (31) | 14 (42) | 7 (20) | .07 |
| Past video game experience, n (%) | 60 (88) | 28 (83) | 32 (94) | .26 |
| Age in years at first video game experience, mean (SD) | 9 (3) | 9 (3) | 9 (3) | .51 |
| Currently playing video games, n (%) | 22 (32) | 14 (40) | 8 (26) | .31 |
| Hours of video game per week, mean (SD) | 1.6 (3.0) | 1.9 (3.7) | 1.3 (2.1) | .65 |

Test Results

The score variability expressed as variance of the students' results was significantly higher in the SG group ($\sigma^2=265.4$) compared with MCQs group ($\sigma^2=140.2$; $P=.009$), as illustrated in Figure 3. The overall results for each test were significantly lower for SG (mean 66.1, SD 16.3 points) compared with MCQs (mean 75.7, SD 11.8 points; $P<.001$). For both the SG and MCQs, the results were better when the student had already performed the other test before: 62.0 (SD 15.2) points for the

SG when it was performed first versus 70.2 (SD 16.5) points when performed second ($P=.02$); 67.4 (SD 8.9) points for the MCQs when it was performed first versus 83.9 (SD 8.1) points when performed second ($P<.001$). No correlation was found between the results of the two tests: $R^2=0.048$ ($P=.58$; Figure 4). No significant difference was observed between gamers (22/68; 32%) and non-gamers (46/68; 68%) for SG results, respectively 65.8 (SD 13.3) versus 66.2 (SD 17.4) points ($P=.71$).

Figure 3. Result's histogram.

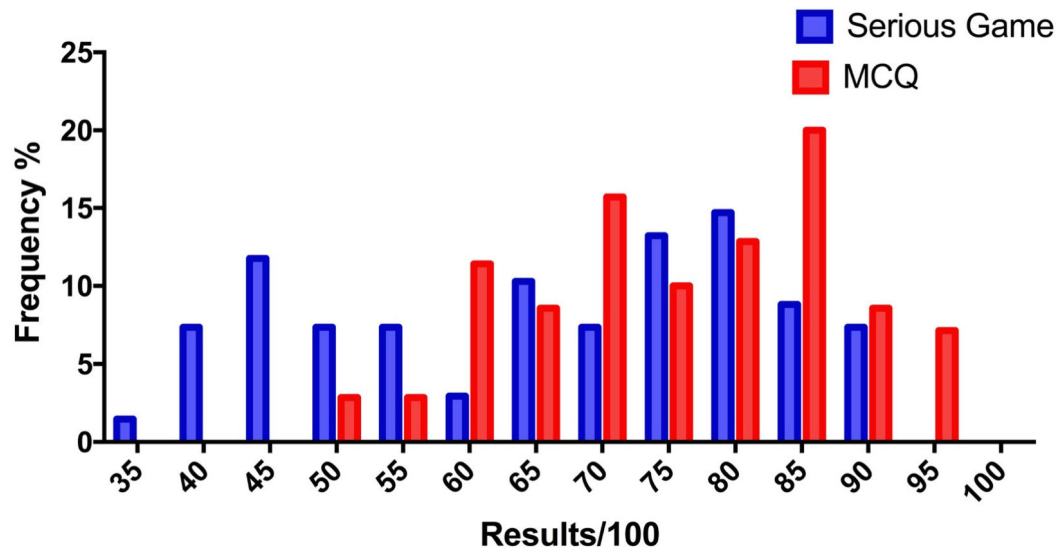


Figure 4. Individual test results in the left panel (A) and correlation coefficient in the right panel (B).

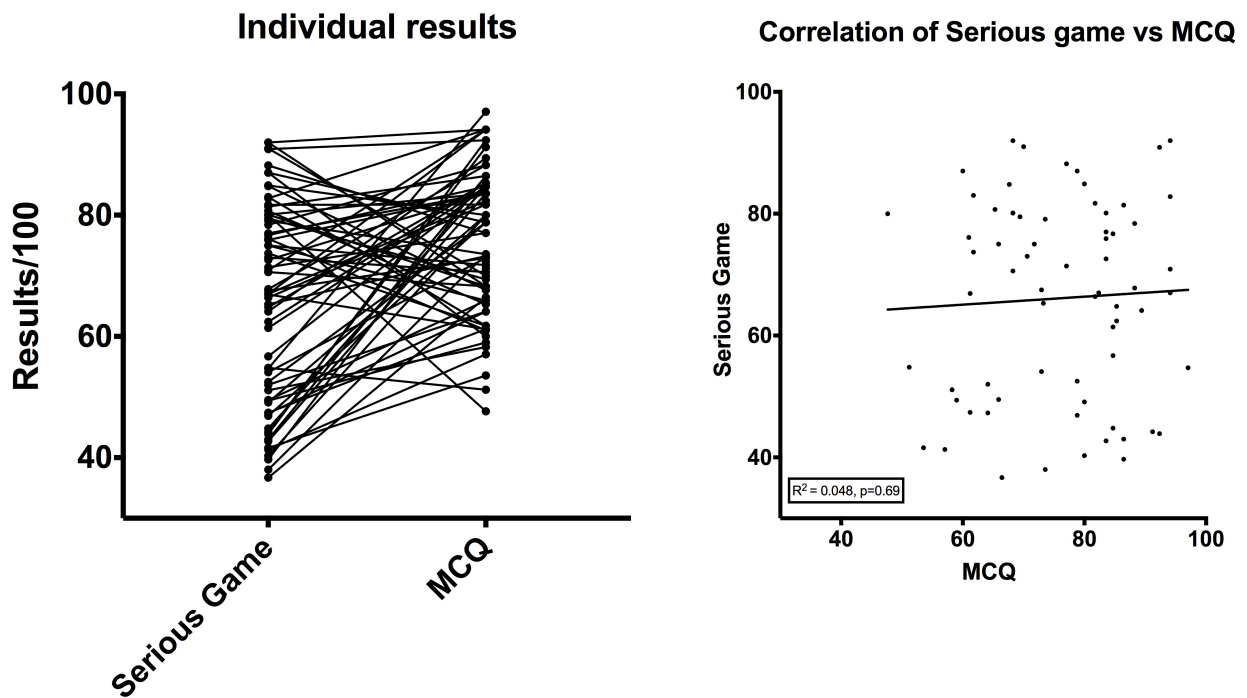


Table 2. Satisfaction questionnaire: results are expressed as mean (SD) of numeric ordinal variable from 1 (no, not at all) to 5 (yes, entirely).

| Questions | Serious game | Multiple choice questions | P value |
|--|--------------|---------------------------|---------|
| Did you encounter difficulties to answer the questions? | 2.18 (1.14) | 2.21 (1.14) | .89 |
| Were you able to concentrate while answering the questions? | 3.93 (0.99) | 3.71 (1.06) | .15 |
| Do you think that this test is close to clinical reality? | 4.21 (0.75) | 2.68 (0.88) | <.001 |
| Did you find this test stressful? | 2.51 (1.05) | 2.30 (1.17) | .24 |
| Did you understand the goal of the test? | 4.24 (0.75) | 3.97 (0.94) | .10 |
| Do you consider that this kind of test represents a proper evaluation? | 3.91 (0.87) | 3.04 (1.02) | <.001 |
| Are you satisfied with your test performance? | 3.05 (1.09) | 3.22 (0.98) | .41 |
| Did you think that your knowledge progressed after this test? | 3.56 (1.09) | 2.42 (0.99) | <.001 |
| Are you satisfied with this type of evaluation? | 3.88 (1.42) | 2.98 (1.53) | <.001 |

Table 3. Assessment of serious games as a tool to learn medicine. Results are expressed as mean (SD) of numeric ordinal variable from 1 (no, not at all) to 5 (yes, entirely).

| Assessment of serious games as a tool to learn medicine | Serious game |
|---|--------------|
| Educational quality | 4.86 (0.35) |
| Feeling of connection or attachment to the serious game | 3.60 (1.19) |
| Possibility of playing with other students | 3.26 (1.18) |
| Possibility of comparing results with other students | 3.44 (1.33) |
| Fun | 3.37 (1.16) |
| Original, innovative or new | 3.90 (0.98) |
| Possibility to adapt level of difficulty | 4.36 (0.68) |
| Availability on smartphone | 4.00 (1.07) |

Satisfaction Analysis

The satisfaction questionnaires showed a significantly higher overall self-reported satisfaction for the SG compared with the MCQ test. Students reported that the SG was closer to clinical practice, represented a proper evaluation and that they felt to have learned more with the SG than with MCQs, thus representing a better evaluation modality ($P<.001$ for all). Conversely, students did not experience significant differences between the two test modalities in terms of understanding, answering the questions, performance satisfaction or stress generated by the test (P value non-significant for all; [Table 2](#)).

Serious Games as a Tool to Learn and Evaluate Medical Skills

The questionnaire was also designed to evaluate whether students thought that SGs could be an interesting tool to learn and evaluate medical skills. Most of the students thought that it could be. The highest ranking points (>4) were educational quality, the possibility of adapting the level of difficulty of an SG and the availability on smart phone ([Table 3](#)).

Discussion

Principal Findings

This study is the first to compare an SG to MCQs in terms of score variability for medical students. This study demonstrates

that the SG was associated with a higher score variability and lower mean score compared with MCQs. Moreover, the SG was associated with significantly higher student satisfaction compared with MCQs. Most medical student evaluation to date is based on MCQ tests which are performed on a large student population. Student grading might therefore be difficult with a sizable proportion of students scoring the same and limited score variability between them. We believe that tests evaluating a large population of medical students should include overall results variability and be of high student satisfaction. For these reasons, we sought to evaluate medical students with a simulation based on an SG compared to MCQs. MCQs evaluate medical knowledge by the means of closed questions, but medical skills and competence are better assessed by on site (bedside) evaluation or simulation [5-8]. While the lines between SGs and simulations are somewhat blurred, an SG represents a virtual world. It is generally played alone with completion based on a score while a simulation is performed on site with an instructor or in a group without score. Several studies have reported higher student satisfaction with simulation compared to MCQs [9,10]. SGs have several potential advantages over simulations to evaluate medical students. Simulation programs are expensive and time-consuming which limits access. Although production of medical SGs is expensive, once the 3D environments have been created it is less expensive to build new SGs using the same environment. Furthermore, they can be easily shared throughout a large medical student community.

SGs also increase the realism of clinical situations [11] and evaluate both medical knowledge and competences via simulation and unguided actions in a 3D environment [12]. Importantly, we believe that SGs exemplify the human desire to play and to master challenges. Besides their potential use as an evaluation tool, SGs might also be an interesting way to train and to teach medicine. The student is drawn “into the game” making medical knowledge and skills more easily transmitted and retained. This use of SGs as a learning tool is supported by our study: students gave a high score for the educational quality of SGs. They particularly appreciated the possibility of accessing SGs with their smartphone and the personalized difficulty feature. Finally, SGs also offer the advantage of self-assessment.

Medical education encompasses both medical knowledge and reasoning skills. Although it is simple to develop MCQs to test medical knowledge, it becomes much more challenging to evaluate reasoning skills and global medical skills with MCQs. Interestingly, our study did not find any correlation between the two sets of test results, suggesting that success in MCQs does not predict success in SGs and vice versa. This finding might suggest that good results in an SG are different from pure medical knowledge evaluation and that medical skills might increase result variability since the medical knowledge tested were similar in both tests. If SGs are considered to be closer to medical practice, this finding questions the effectiveness of MCQs in evaluating medical students [10,13]. This finding possibly suggests that the tests evaluate different reasoning skills and abilities to perform.

We acknowledge several limitations in our study. Our study compared two different test modalities evaluating a relatively small number of medical students in managing a cardiology clinical case. Therefore, further studies are needed to confirm our findings in larger student populations and in other medical fields. Although we found an order bias in our study—the second test was associated with better results because of similar questions, retention of the students’ answers, and indirect access to the corrections—we believe that randomization in two similar groups allowed us to draw reliable conclusions. As specific questionnaires were designed for this study, no pretest was conducted. Nevertheless, we believe that the questionnaires are valid, since each student acted as his own control in this study, interpreting the questions in the same way when evaluating two different test modalities. Finally, our sample consisted of volunteer students and we cannot rule out the fact that they might have a particular interest in SGs. This might also limit the generalization of our conclusions to the whole population of medical students.

Conclusions

SGs potentially represent a new evaluation modality for medical students. Our study suggests that they are more effective in grading medical students with a higher variability of performance. In addition, SGs seem to be associated with higher student satisfaction compared to MCQs.

Acknowledgments

The authors thank Dr G Kanellopoulos and G Durand-Viel for their help and support with this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Multiple choice questions test translated in English.

[\[PDF File \(Adobe PDF File\), 256KB-Multimedia Appendix 1\]](#)

References

1. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. *J Med Internet Res* 2016;18(1):e2 [FREE Full text] [doi: [10.2196/jmir.4807](https://doi.org/10.2196/jmir.4807)] [Medline: [26729058](https://pubmed.ncbi.nlm.nih.gov/26729058/)]
2. Menet A, Assez N, Lacroix D. Cross analysis of knowledge and learning methods followed by French residents in cardiology. *Arch Cardiovasc Dis* 2015;108(8-9):420-427 [FREE Full text] [doi: [10.1016/j.acvd.2015.03.001](https://doi.org/10.1016/j.acvd.2015.03.001)] [Medline: [25921839](https://pubmed.ncbi.nlm.nih.gov/25921839/)]
3. Ryall T, Judd BK, Gordon CJ. Simulation-based assessments in health professional education: a systematic review. *J Multidiscip Healthc* 2016;9:69-82 [FREE Full text] [doi: [10.2147/JMDH.S92695](https://doi.org/10.2147/JMDH.S92695)] [Medline: [26955280](https://pubmed.ncbi.nlm.nih.gov/26955280/)]
4. Graafland M, Dankbaar M, Mert A, Lagro J, De Wit-Zuurendonk L, Schuit S, et al. How to systematically assess serious games applied to health care. *JMIR Serious Games* 2014;2(2):e11 [FREE Full text] [doi: [10.2196/games.3825](https://doi.org/10.2196/games.3825)] [Medline: [25654163](https://pubmed.ncbi.nlm.nih.gov/25654163/)]
5. Gordon J, Oriol N, Cooper J. Bringing good teaching cases “to life”: a simulator-based medical education service. *Acad Med* 2004;79(1):23-27. [Medline: [14690993](https://pubmed.ncbi.nlm.nih.gov/14690993/)]
6. Scouller K. The influence of assessment method on students' learning approaches: multiple choice question examination versus assignment essay. *Higher Education* 1998;35(4):453-472. [doi: [10.1023/A:1003196224280](https://doi.org/10.1023/A:1003196224280)]
7. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Med Educ* 2007 Nov 28;7:49 [FREE Full text] [doi: [10.1186/1472-6920-7-49](https://doi.org/10.1186/1472-6920-7-49)] [Medline: [18045500](https://pubmed.ncbi.nlm.nih.gov/18045500/)]

8. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med* 2008 Jan;23(Suppl 1):46-49. [doi: [10.1007/s11606-007-0283-4](https://doi.org/10.1007/s11606-007-0283-4)] [Medline: [18095044](https://pubmed.ncbi.nlm.nih.gov/18095044/)]
9. Wahlgren C, Edelbring S, Fors U, Hindbeck H, Ståhle M. Evaluation of an interactive case simulation system in dermatology and venereology for medical students. *BMC Med Educ* 2006;6:40 [FREE Full text] [doi: [10.1186/1472-6920-6-40](https://doi.org/10.1186/1472-6920-6-40)] [Medline: [16907972](https://pubmed.ncbi.nlm.nih.gov/16907972/)]
10. Solymos O, O'Kelly P, Walshe CM. Pilot study comparing simulation-based and didactic lecture-based critical care teaching for final-year medical students. *BMC Anesthesiol* 2015 Oct 21;15:153 [FREE Full text] [doi: [10.1186/s12871-015-0109-6](https://doi.org/10.1186/s12871-015-0109-6)] [Medline: [26490826](https://pubmed.ncbi.nlm.nih.gov/26490826/)]
11. Johnsen HM, Fossum M, Vivekananda-Schmidt P, Fruhling A, Slettebø Å. Teaching clinical reasoning and decision-making skills to nursing students: design, development, and usability evaluation of a serious game. *Int J Med Inform* 2016 Oct;94:39-48. [doi: [10.1016/j.ijmedinf.2016.06.014](https://doi.org/10.1016/j.ijmedinf.2016.06.014)] [Medline: [27573310](https://pubmed.ncbi.nlm.nih.gov/27573310/)]
12. Virvou M, Katsionis G, Manos K. Combining software games with education: evaluation of its educational effectiveness. *J Educ Techno Soc* 2006;8:54-65.
13. Duque G, Fung S, Mallet L, Posel N, Fleischer D. Learning while having fun: the use of video gaming to teach geriatric house calls to medical students. *J Am Geriatr Soc* 2008 Jul;56(7):1328-1332. [doi: [10.1111/j.1532-5415.2008.01759.x](https://doi.org/10.1111/j.1532-5415.2008.01759.x)] [Medline: [18482292](https://pubmed.ncbi.nlm.nih.gov/18482292/)]

Abbreviations

MCQs: Multiple choice questions

SG: Serious games

SD: Standard Deviation

Edited by M Focsa; submitted 22.11.16; peer-reviewed by J Creutzfeldt, Q Liu; comments to author 16.12.16; revised version received 08.02.17; accepted 27.02.17; published 16.05.17

Please cite as:

Adjedj J, Ducrocq G, Bouleti C, Reinhart L, Fabbro E, Elbez Y, Fischer Q, Tesniere A, Feldman L, Varenne O
Medical Student Evaluation With a Serious Game Compared to Multiple Choice Questions Assessment

JMIR Serious Games 2017;5(2):e11

URL: <http://games.jmir.org/2017/2/e11/>

doi: [10.2196/games.7033](https://doi.org/10.2196/games.7033)

PMID: [28512082](https://pubmed.ncbi.nlm.nih.gov/28512082/)

©Julien Adjedj, Gregory Ducrocq, Claire Bouleti, Louise Reinhart, Eleonora Fabbro, Yedid Elbez, Quentin Fischer, Antoine Tesniere, Laurent Feldman, Olivier Varenne. Originally published in *JMIR Serious Games* (<http://games.jmir.org>), 16.05.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Serious Games*, is properly cited. The complete bibliographic information, a link to the original publication on <http://games.jmir.org>, as well as this copyright and license information must be included.