Method for Evaluating the Quality of Serious Games in Medical Education

Francisco Anderson Mariano da Silva[1][a], Wellington Candeia de Araújo[2][b], Thiago Prado de Campos[3][c] and Tiago Silva da Silva[1][d]
[1]*Institute of Science and Technology (ICT), Federal University of São Paulo, Brazil*
[2]*Center for Science and Technology, State University of Paraíba (UEPB), Brazil*
[3]*Federal University of Technology, Paraná, Brazil*

Keywords: Serious Games, Medical Education, Usability Evaluation, User Experience, Motivation, Knowledge Acquisition.

Abstract: The integration of serious games into medical education, particularly in surgical training, has proven to be a promising approach for enhancing skills and knowledge acquisition. This study introduces the MAQJSEM (Method for Evaluating the Quality of Serious Games in Medical Education), a comprehensive evaluation method designed to address critical dimensions such as motivation, user experience, usability, and knowledge acquisition. The development process involved a systematic literature review, followed by validation through a pilot study and expert evaluations. MAQJSEM was applied to a mobile application focusing on surgical training, and its evaluation revealed the method's robustness and practicality in assessing serious games within this context. Notable findings include the importance of incorporating emotional and immersive elements, as well as clear instructions and intuitive usability features. Expert feedback led to the refinement of dimensions and items, enhancing the clarity and relevance of the method. MAQJSEM contributes significantly by offering a validated and adaptable tool for improving serious games in medical education. The method supports developers and educators in creating engaging and pedagogically effective tools, fostering skill development and preparation for real-world challenges in the medical field.

1 INTRODUCTION

This paper presents an innovative method for evaluating serious games in medical education, with a special focus on surgical training. Serious games, introduced by Clark C. Abt in the 1970s, have become established as pedagogical tools that combine motivational and interactive elements, fostering engagement, immersion, and continuous feedback (Abt, 1987). In healthcare, events like *Games for Health* highlight their potential in training professionals (Drummond & Tesnière, 2017). However, their effectiveness depends on rigorous evaluations to ensure quality and identify areas for improvement.

This study aims to answer the research question: *"What are the relevant aspects for evaluating serious games in medical education, especially surgical training, and how can a method be developed and validated to measure their quality?"* To address this, we propose the Method for Evaluating the Quality of Serious Games for Medical Education (MAQJSEM), designed to assess key dimensions such as motivation, user experience, usability, and knowledge acquisition.

The method was developed based on theories of motivation (Keller, 1987), user experience (Savi, 2011), usability (Petri et al., 2019; Nielsen & Molich, 1990), and studies on serious games in medical education, particularly *Graafland, Schraagen, & Schijven (2012)* and *Meijer et al. (2019)*. Validation

[a] https://orcid.org/0000-0001-9797-7552
[b] https://orcid.org/0000-0003-2102-7993
[c] https://orcid.org/0000-0003-2102-7993
[d] https://orcid.org/0000-0003-1038-4004

da Silva, F. A. M., de Araújo, W. C., de Campos, T. P. and da Silva, T. S.
Method for Evaluating the Quality of Serious Games in Medical Education.
DOI: 10.5220/0013433800003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 2*, pages 605-613
ISBN: 978-989-758-749-8; ISSN: 2184-4992

was conducted in two stages: (i) a pilot study, in which the method was applied to a serious game with feedback from gaming and medical experts, and (ii) a face and content validation study, ensuring that the assessment criteria were relevant and comprehensible to the target audience.

Key results include the incorporation of dimensions such as immersion and emotions, restructuring questions for clarity, and adding items related to tutorials and sound—critical factors for engagement and learning. This approach enhances the evaluation of serious games and lays a foundation for future studies, fostering advancements in the use of these tools in medical education.

This paper is structured as follows: Section 2 presents related work, highlighting previous approaches to evaluating serious games in medical contexts. Section 3 details the methodology for developing and validating MAQJSEM. Section 4 presents the proposed method, while Sections 5 and 6 describe studies conducted to evaluate and validate it. Finally, Section 7 concludes with reflections on the study's implications and future research directions.

2 RELATED WORKS

Several studies propose methods to evaluate serious games, each addressing different aspects. Rocha, Bittencourt, and Isotani (2015) developed a questionnaire for self-assessment and learner reactions regarding simulation and training. Rocha (2017) outlined criteria for balancing content, simulation, game elements, and evaluation. Schroeder and Hounsell (2015) introduced SEU-Q to assess serious games as tools, while Petri, von Wangenheim, and Borgatto (2019) proposed MEEGA+ for educational games in science.

Other works expand evaluation dimensions. Oliveira and Rocha (2020) developed "Avalia JS" for serious game planning. Rodrigues et al. (2021) assessed a game targeting childhood obesity, while Pires et al. (2015) created an instrument for evaluating playfulness in healthcare games. Feitosa (2018) designed a quiz-based game for dental biosafety, and Santos (2018) introduced PAJDE, focusing on learning potential in educational games. These studies highlight approaches and the need for tailored evaluation tools.

Campos (2024) applied a similar validation approach, developing UUXE-ToH to assess touchable holograms. The study emphasized iterative refinements and expert feedback to ensure quality applicability.

While these studies provide valuable insights, none specifically address the pedagogical challenges of surgical training. MAQJSEM was developed to fill this gap by integrating validated elements from methods—focusing on motivation, user experience, usability, and knowledge acquisition—while adapting them to surgical education. Its validation through expert feedback has positioned it as a robust tool for assessing serious games in training, contributing to advancements in surgical education.

3 METHODOLOGY

This section outlines the development and validation process of the MAQJSEM method. The research followed three key stages: a literature review to identify essential evaluation criteria, the formulation of the method based on established theoretical frameworks, and validation through a pilot study and content and face analysis by experts. This ensured the method's accuracy, applicability, and alignment with medical education needs.

This research aimed to develop and validate the MAQJSEM, an evaluation method focusing on surgical training. The methodology followed a structured process (Figure 1), including:

3.1 Systematic Literature Review (SLR)

Conducted according to Kitchenham et al. (2009), complemented by consultations with theses and dissertations. This step identified gaps and incorporated contributions from key authors such as Keller (1987), Savi et al. (2011), Bedwell et al. (2012), Rocha (2014), Petri et al. (2019), Graafland et al. (2012), and Meijer et al. (2019).

3.2 Method Development

The method was developed based on Costa (2011) and DeVellis (2017), emphasizing clear definitions, item selection, and validation processes.

3.3 Validation Process

3.3.1 Pilot Study

As highlighted by Benassi, Cancian, & Strieder (2023), pilot studies are a crucial tool in perception research, as they help identify flaws, challenges, and opportunities for improvement, which can then be addressed before the main research phase. The goal

was to identify and correct flaws in question construction and ensure coherence across dimensions. To address any potential unfamiliarity with serious games, the researcher participated in an introductory lecture designed to contextualize the evaluation process. Feedback from this phase led to adjustments in dimensions, question clarity, and the structure of the method.

3.3.2 Face and Content Validation Study

It ensures that questionnaire items are representative of the domain and understandable to the target audience. Content validation involves expert analysis to confirm coverage of theoretical constructs, while face validation assesses clarity and appropriateness from respondents' perspectives through quantitative and qualitative feedback. These procedures enhance the instrument's reliability and validity before larger studies, following psychometric assessment standards.

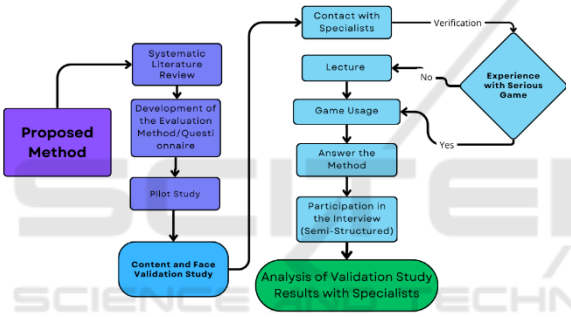

Figure 1: Items construction and validation process.

4 MAQJSEM

The MAQJSEM method is grounded in a robust theoretical foundation, integrating approaches and methodologies from the reviewed literature. Central to its development are the ARCS model (Attention, Relevance, Confidence, and Satisfaction) for evaluating motivation, usability heuristics, and studies on user experience in games. It also incorporates categories for identifying serious games and the MEEGA+ method, designed for assessing educational games in fields such as Computer Science and surgical skill training. Contributions from the medical field were included, emphasizing knowledge evaluation and surgical practices, bridging educational and practical dimensions.

As a result of the systematic literature review (SLR), 70 items were developed and organized into four main dimensions: Motivation, User Experience, Usability, and Knowledge Acquisition. These items form the basis for a comprehensive evaluation framework that assesses both student motivation and the effectiveness of educational content delivered by serious games.

4.1 Dimensions of the MAQJSEM

The development of the evaluation method was structured around four main dimensions, each addressing specific aspects of the serious games experience in medical education:

Motivation: Evaluates how the game keeps students engaged, focusing on the attractiveness of challenges, feedback and rewards, and social interaction. **User Experience:** Analyzes aesthetics, design, immersion, and the connection with characters and scenarios. **Usability:** Examines ease of use, including intuitive interfaces and responsive controls. **Knowledge Acquisition:** Assesses the game's impact on learning medical concepts, developing surgical skills, and preparing for real-world situations.

Figure 2 presents a visual representation of these dimensions and how they interconnect within the evaluation method.

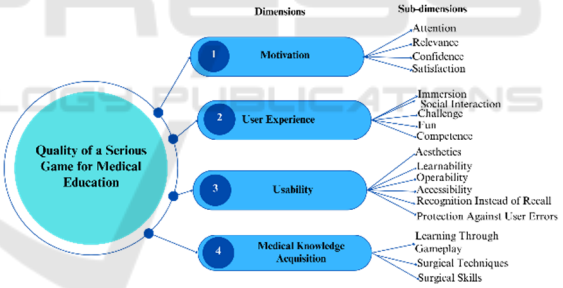

Figure 2: Method Design.

The creation of items for evaluating game quality was based on models and research previously discussed, such as Keller (1987), Savi et al. (2011), Petri et al. (2019), Nielsen and Molich (1990), and Graafland et al. (2012–2019). These sources particularly emphasized the application of serious games in medicine.

Table 1 summarizes the key elements and authors that informed the development of the evaluation categories, which were refined and translated into specific questions tailored to medical education through serious games.

Table 1: Key Elements and Authors for Development.

Categories	Authors and Quality Indicators
Motivation	Keller (1987)
User Experience	Savi *et al*. (2011), Bedwell *et al.* (2012)
Usability	Petri et al. (2019), Nielsen & Molich (1990)
Knowledge Acquisition	Graafland et al. (2012, 2013, 2014, 2015, 2017, 2018) **and** Meijer et al. (2018, 2019, 2021, 2022).
	Rocha (2014)

Source: Author, 2024

5 PILOT STUDY

This section outlines the pilot study to evaluate the MAQJSEM method's initial effectiveness, identify improvements, and assess its feasibility in real-world applications.

5.1 Instruments and Procedures

The pilot study assessed the preliminary feasibility and clarity of the MAQJSEM method before full-scale validation, identifying potential ambiguities in the first version of the questionnaire[1], refine its structure, and ensure its applicability in evaluating serious games for medical education.

The participant (P1), a specialist with a bachelor's and master's degree in Computer Science and in the final semester of a Ph.D. in the field, contributed over 15 years of experience in the development and evaluation of accessibility, usability, and user experience for websites and mobile applications. This expertise ensured precise and applicable results, strengthening the study's robustness.

The game *Operate Now: Hospital* (Azerion Casual, 2017) was selected due to its focus on surgical scenarios and task-based gameplay, aligning with the method's objectives. Though not a medical student, the participant's usability expertise provided critical insights into the questionnaire's clarity and structure. Future studies will extend validation to medical students and professionals to further refine the method in real training contexts.

The study involved the participant downloading, installing, and using the game over a three-day period, followed by an evaluation using the MAQJSEM method. Additionally, the participant provided feedback on the method itself, further supporting its refinement.

The evaluation was conducted via a Google Form containing 70 items distributed across four dimensions. Each item was assessed on a five-point Likert scale (from strongly disagree to strongly agree). At the end of each dimension, an open-ended question allowed the participant to suggest improvements. After completing the form, the participant participated in a semi-structured interview to clarify doubts and provide additional insights.

The interview was recorded, transcribed, and analyzed using Atlas.ti software to identify emerging patterns and categories that informed the instrument's refinement. Content analysis (Bardin, 2016), a widely used qualitative research technique, was employed to code and interpret the data.

After the pilot study, the method's structure was reorganized, with wording and content adjustments to enhance clarity and relevance in evaluating serious games for medical education, resulting in the post-pilot version[2].

5.2 Results and Discussion

The quantitative evaluation reinforced the method's robustness and highlighted areas for improvement. P1's responses showed strong agreement with key aspects of the method. For instance, in the Motivation dimension, 70% of responses were in the "Strongly Agree" category for items addressing attention-capturing activities and the relevance of skills taught. In the UX dimension, 90% "Strongly Agreed" on the aesthetic appeal and engaging graphics of the game. Similarly, 90% "Strongly Agreed" that instructions and tutorials were clear and effective in Usability. In Knowledge Acquisition, 80% of responses confirmed the game reinforced medical concepts and provided learning.

The interview with P1 provided valuable insights and suggestions for refining the method. P1 emphasized the clarity of the method's objectives, describing them as well-defined and comprehensive. Key feedback included recommendations to address emotions, game efficiency, and UX, along with adjustments to items related to immersion, usability, and emotional aspects. These inputs guided the reclassification of items, removal of unnecessary questions, and enhancements to question clarity.

The method's clarity was praised, particularly its transparency in evaluating criteria such as motivation

[1] https://figshare.com/s/739ddf3f3245fd6a3c27

[2] https://figshare.com/s/4854c73dfebc9206a652

and usability. However, P1 recommended reallocating items to better align with categories and improving questions targeting medical and surgical education. Motivation was emphasized as critical for maintaining user engagement and fostering learning, with suggestions to link concentration and engagement to outcomes.

P1 highlighted UX as a key focus area, recommending adjustments to aesthetics-related items and the inclusion of questions about emotions to enhance game immersion. Accessibility was also addressed, with suggestions to incorporate items on fatigue and usability to account for diverse user profiles. Engagement was deemed vital, with feedback encouraging a stronger connection between concentration, immersion, and learning.

Knowledge acquisition, the core goal of serious games, was assessed for its role in developing surgical skills. The relationship between confidence, game progression, and learning was emphasized. Immediate feedback was noted as particularly important for reinforcing learning, with a suggestion to prioritize rapid responses to user actions. Socialization was also highlighted as a valuable factor for knowledge sharing among students and healthcare professionals. Graafland and Schijven (2018) reinforces this, noting the efficacy of serious games in clinical simulations and skill development. P1 further suggested involving medical field experts for more precise evaluations in future stages.

6 CONTENT AND FACE VALIDATION STUDY

6.1 Instruments and Procedures

This study was approved by the Research Ethics Committee, adhering to ethical guidelines to ensure participant safety and research integrity. Conducted between September 1 and October 30, 2024, participants were selected via email invitations sent to specialists in serious games, computing, and medicine. Despite invitations to female specialists, all responses came from male participants. This may reflect a gender imbalance in serious game development or the availability of female researchers for this stage. Future studies should investigate this issue, examining gender representation in medical educational game research and promoting greater diversity among evaluators.

Interviews were conducted via Zoom between September 20 and October 27, 2024. Participants unfamiliar with serious games attended an introductory lecture, while experienced ones installed and evaluated *Operate Now: Hospital*, followed by a semi-structured interview.

The process involved completing an Google Forms questionnaire to evaluate the post-pilot version of MAQJSEM. The form had 70 items across four dimensions: motivation, user experience, usability, and knowledge acquisition. Each item was rated on a Likert scale, with open-ended questions at the end of each dimension for improvement suggestions.

Following the questionnaire, participants engaged in semi-structured interviews to clarify doubts and provide further insights. The interviews were recorded, transcribed, and analyzed using Atlas.ti software, identifying key categories and patterns. This iterative process ensured the method was refined to meet evaluators' expectations and practical needs.

6.2 Results and Discussion

This subsection presents expert assessments, interview insights, and key decisions that refined the post-validated version.

6.2.1 Quantitative Results

Nine experts answered the MAQJSEM evaluation form. The quantitative data from the experts' evaluations showed that, in general, experts classified the questions in each dimension as relevant for that dimension's evaluation. On average, for each question, eight (8.2) experts indicated that it was relevant. 25 questions were unanimously classified as relevant. The question indicated as not relevant by the most experts was Q18, with 3 indications. Eight questions were classified as not relevant twice.

In the group of questions in the **Motivation** dimension (Figure 3a), 6 questions were unanimously identified as relevant, 5 were classified as not relevant once, and two questions were classified as not relevant twice. Regarding the **UX** dimension (Figure 3b), three questions were unanimously classified as relevant; 10 questions were classified as not relevant once; 5 questions were identified as not relevant twice; and 1 question was identified as not relevant by three experts. In the group of questions in the **Usability** dimension (Figure 3c), 7 questions were unanimously identified as relevant, and the other 6 had only one classification as not relevant.

Figures 3: a) Motivation, b) User Experience and c) Usability.

Regarding the **Knowledge Acquisition** dimension (Figure 4), 7 questions were unanimously classified as relevant; 16 questions had only one classification as not relevant, and only one question was classified by two experts as not relevant.

Figure 4 - Knowledge Acquisition.

In summary, the evaluation by nine experts confirmed the overall relevance of the MAQJSEM questions, with most receiving unanimous or near-unanimous agreement. While a few questions, such as Q18, were flagged as less relevant, the findings highlight the questionnaire's effectiveness in addressing its intended dimensions, with opportunities for minor refinements.

6.2.2 Qualitative Results

The data obtained through the semi-structured interviews allowed the identification of central categories for evaluating the method. The main categories identified include general and specific suggestions, method objectivity, method utility, proposals for including new statements, and dimensions for emotional and immersive experience.

In the **general and specific suggestions** category, P5 stated that "the instructions on how to apply the method should be more detailed, especially for users unfamiliar with serious games." P6 noted that "dimensions such as motivation are well-addressed, but the usability dimension could include more items about customization and accessibility."

In the **method objectivity** category, P1 highlighted that "the method's objectives are well-defined and comprehensive, making it easier for users to understand and apply." However, P2 suggested that "some questions could be rephrased to better address surgical learning."

In the **method utility** category, P3 mentioned that "the method is practical and well-structured, but it would be interesting to include more elements that directly connect the game's practice to real work environments." P4 reinforced that "the method meets evaluation demands but could include additional criteria to measure emotional experience."

In the **proposals for including new statements** category, P7 suggested that "it would be interesting to include questions related to the impact of sound on user experience, as this can influence immersion." P8

recommended adding questions about the effectiveness of tutorials: "Tutorials are important for guiding new users, and the method could explore this aspect further."

In the **dimensions for emotional and immersive experience** category, P9 emphasized that "questions focusing on how emotions are evoked during gameplay could provide insights into user engagement and learning retention." P10 added that "immersion levels should be assessed based on how realistic the scenarios feel and how much they encourage concentration and focus during gameplay."

These findings from the semi-structured interviews provided fundamental results for adjustments to the method, ensuring it became more robust and aligned with the needs and expectations of evaluators and future users.

6.3 Processing and Decisions

After analyzing the validation study, several improvements were implemented to enhance the method's comprehensiveness and effectiveness. To address overlooked aspects, new sentences were incorporated, focusing on critical areas such as sound quality and tutorial effectiveness. For example, sentences were added to assess whether the sound quality supports the intended experience and whether the tutorial effectively guides users in navigating the game and performing tasks confidently.

Redundant sentences were removed to streamline the questionnaire, making it more concise and reducing respondent fatigue. Examples of excluded sentences include those that overlapped in purpose, such as statements related to the improvement of surgical skills and the game's visual aesthetics. This optimization ensured that each question uniquely contributed to the evaluation without unnecessary repetition.

To provide targeted feedback, questions related to knowledge acquisition were divided into conceptual and practical aspects. This distinction allowed for a more precise evaluation of theoretical understanding versus hands-on application, reflecting the dual focus of knowledge development in serious games.

The inclusion of medical experts in the validation process ensured that the content aligned with practical realities and professional standards. This collaboration refined the questionnaire to meet the specific needs of evaluating serious games in medical education. As a result of this study, the questionnaire was refined and reached its **post-validation version**[3].

[7] https://figshare.com/s/378ade76028fd661a165

A limitation of MAQJSEM is the absence of direct comparisons with established methods in systems engineering, which often include robust frameworks for evaluating interactive software. Models such as ISO/IEC 25010, used to assess software quality, could be adapted for evaluating serious games in medical contexts.

Although MAQJSEM was validated using a mobile-based serious game, its structured evaluation framework allows for adaptation to different technological contexts, including Virtual Reality (VR) and Augmented Reality (AR) applications. The core dimensions of motivation, user experience, usability, and knowledge acquisition apply across various training environments, making the method flexible for different serious gaming platforms. Future studies should explore its effectiveness in immersive VR and AR simulations, particularly for surgical training, where hands-on experience and spatial awareness are crucial. Such adaptations would further validate the robustness and applicability of MAQJSEM in diverse medical education settings.

7 CONCLUSIONS

This study developed and validated MAQJSEM, a method for evaluating the quality of serious games in medical education, focusing on surgical training. Through a systematic approach based on literature review, pilot study, and expert validation, the method was refined to ensure clarity, precision, and practical applicability. MAQJSEM integrates motivation, user experience, usability, and knowledge acquisition, providing a robust tool to assess and optimize the pedagogical impact of serious games.

The results highlight the potential of serious games as allies in medical education, emphasizing the importance of designs that meet students' pedagogical and contextual needs. The method offers developers a valuable instrument to ensure their educational solutions fulfill medical training requirements, particularly in complex fields such as surgery.

7.1 Limitations

Limitations include the small validation sample, absence of female experts, and focus on a specific mobile application. Future studies should explore MAQJSEM in other technological contexts, such as augmented (AR) and virtual reality (VR), and assess its adaptability to a broader range of serious games.

The lack of complementary studies validating the questionnaire remains a limitation. Future research should conduct additional validations to strengthen the method's reliability and applicability across different scenarios.

7.2 Future Work

Moving forward, future research will focus on expanding the application of MAQJSEM to a broader range of serious games in medical education, including those designed for different specialties and technological platforms such as VR and AR. Additionally, a large-scale validation study involving medical students and professionals will be conducted to assess the instrument's effectiveness in real-world training environments. This will help refine the method further and ensure its adaptability to various educational and clinical contexts.

As a contribution, MAQJSEM advances the literature by proposing a specific evaluation method for serious games in medical education, combining methodological rigor with practical applicability. This method can positively impact the development of effective educational games, supporting the training of well-prepared physicians.

REFERENCES

ABT, Clark C. Serious games. University press of America, 1987.

Bardin, L. (2016). *Análise de Conteúdo*. Edições 70.

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study. *Simulation & Gaming, 43*(6), 729-760. .

Benassi, C. B. P., Cancian, Q. G., & Strieder, D. M. (2023). Estudo piloto: Um instrumento primordial para a pesquisa de percepção da ciência. *Ensino e Tecnologia em Revista, 7*(1), 210-225.

Campos, T. P., Damasceno, E. F., & Valentim, N. M. C. (2024). Evaluating Usability and UX in Touchable Holographic Solutions: A Validation Study of the UUXE-ToH Questionnaire. *International Journal of Human–Computer Interaction*, 1-21.

Costa, F. D. (2011). *Mensuração e desenvolvimento de escalas: aplicações em administração*. Rio de Janeiro: Ciência Moderna, 90-106.

Dantas, C. D. C., Leite, J. L., Lima, S. B. S. D., & Stipp, M. A. C. (2009). Teoria fundamentada nos dados: Aspectos conceituais e operacionais: Metodologia possível de ser aplicada na pesquisa em enfermagem. *Revista Latino-Americana de Enfermagem, 17*, 573-579.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: Theory and Applications.* Sage Publications.

Drummond, D., Hadchouel, A., & Tesnière, A. (2017). Serious games for health: Three steps forwards. *Advances in Simulation, 2*, 3. https://doi.org/10.1186/s41077-017-0036-3.

Feitosa, M. L. B. (2018). Biossegurança em odontologia: Criação e validação de um serious game do tipo quiz direcionado para profissionais de saúde bucal. *Universidade do Estado do Rio de Janeiro*. https://www.bdtd.uerj.br:8443/handle/1/5932.

Gorbanev, I., Agudelo-Londoño, S., González, R. A., Cortes, A., Pomares, A., Delgadillo, V., Yepes, F. J., & Muñoz, Ó. (2018). A systematic review of serious games in medical education: Quality of evidence and pedagogical strategy. *Medical Education Online, 23*(1), 1438718. https://doi.org/10.1080/10872981.2018.1438718.

Graafland, M., Schraagen, J. M., & Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *The British Journal of Surgery, 99*(10), 1322-1330. https://doi.org/10.1002/bjs.8819.

Graafland, Maurits; Schijven, Marlies. How serious games will improve healthcare. *Digital health: scaling healthcare to the world*, p. 139-157, 2018.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2-10.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering: A systematic literature review. *Information and Software Technology, 51*(1), 7-15.

Meijer, H. A. W., Graafland, M., Obdeijn, M. C., Goslings, J. C., & Schijven, M. P. (2019). Face validity and content validity of a game for distal radius fracture rehabilitation. *Journal of Wrist Surgery, 8*(5), 388-394. https://doi.org/10.1055/s-0039-1688948.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 249-256.

Azerion Casual. (2017). Operate Now Hospital - Surgery. Google Play Store. https://play.google.com/store/apps/details?id=com.spilgames.OperateNow2&hl=pt_BR&gl=US.

Oliveira, R. N. R., & Rocha, R. V. (2020). Modelo conceitual para planejamento da avaliação em jogos sérios. *SBGames Proceedings.* https://www.sbgames.org/proceedings2020/EducacaoFull/209743.pdf.

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2019). MEEGA+: Um modelo para a avaliação de jogos educacionais para o ensino de Computação. *Revista Brasileira de Informática na Educação*. http://br-ie.org/pub/index.php/rbie.

Pires, M. R. G. M., Göttems, L. B. D., Silva, L. V. S., Carvalho, P. A., Melo, G. F., & Fonseca, R. M. G. S.

(2015). Desenvolvimento e validação de instrumento para avaliar a ludicidade de jogos em saúde. *Revista da Escola de Enfermagem da USP, 49*(4), 658–664. https://www.scielo.br/j/reeusp/a/XTDYzBp8Lvgt7VZBHGFBGCN/?format=pdf&lang=pt.

Rodrigues, J. D., Tibes-Cherman, C. M., Aragão, R. B., Filho, H. T., Zem-Mascarenhas, S. H., & Fonseca, L. M. M. (2021). Avaliação de serious game em programa de enfrentamento da obesidade infantil. *Acta Paulista de Enfermagem, 34*, eAPE00102. https://www.scielo.br/j/ape/a/fwLKnwtq6j5RQp46VK9VWdJ/.

Rocha, R. V. (2017). Critérios para a construção de jogos sérios. *Simpósio Brasileiro de Informática na Educação.* http://ojs.sector3.com.br/index.php/sbie/article/view/7623/5419.

Rocha, R. V., Bittencourt, I. I., & Isotani, S. (2015). Avaliação de jogos sérios: Questionário para autoavaliação e avaliação da reação do aprendiz. *SBGames Proceedings.* http://www.sbgames.org/sbgames2015/anaispdf/artesedesign-full/147637.pdf.

Savi, R. (2011). Avaliação de jogos voltados para a disseminação do conhecimento. *Tese de doutorado.* Universidade Federal de Santa Catarina.

Schroeder, R. B., & Hounsell, M. S. (2015). SEU-Q: Um instrumento de avaliação de utilidade de jogos sérios ativos. *ResearchGate.* https://www.researchgate.net/publication/312212577_SEU-Q.

Santos, W. S. (2018). Um modelo de avaliação para jogos digitais educacionais. *Universidade do Estado da Bahia.* http://repositoriosenaiba.fieb.org.br/bitstream/fieb/895/1/William%20de%20Souza%20Santos.pdf