A Quality Label for Serious Games

The Dutch Society of Simulation in Healthcare (DSSH) has developed a transparent evaluation framework with an associated quality label for medical serious games. B. Doyen, A. Mert, H.A.W. Meijer, T.C.W. Nijboer, and M.E.W. Dankbaar report.

The development of medical serious games is booming. In the past years, the number of healthcare-related serious games has rapidly increased, just as the number of purposes for which they are being developed. Although several definitions of serious games exist, they can be defined as "interactive computer applications, that are fun to play, challenging and engaging, incorporate some concept of scoring, and teach the user a skill, knowledge, or attitude that can be of use in the real world".[1] These serious games are praised for their attractive, engaging, challenging and motivating nature and some have already shown their effectiveness in various settings, including medical education for both patients and healthcare workers.[2,3]

Despite this premise and the large number of released medical serious games, only a minority of them are successfully implemented and used in the context for which they were designed.[3] This may be attributed to a great variation in quality and a lack of transparency regarding the rationale, relevance and scientific validity of these games.[4] Additionally, it is not always clear to what degree various user groups such as relevant experts (e.g. game, healthcare and educational specialists) and the end-users were involved during development of the game. The latter is very important to make sure potential users appreciate and understand the game and may learn from it. This can make it very difficult for potential users to distinguish good and effective serious games from nice looking, but ineffective or potentially even harmful games. Differentiating between effective and ineffective games is especially relevant in the current era of evidence-based medicine and education, where sound evidence of effectiveness is required before policymakers consider implementing a serious game as a teaching or treatment modality. An additional concern is data-safety, since many of these games store personal and/or even medical data either locally or on external servers. Unfortunately, clear information on these topics is rarely available for potential users, in app stores, or developer websites.

In order to help solve these issues and to stimulate the production of well-developed, scientifically supported and tested serious games, the Dutch Society of Simulation in Healthcare (DSSH) has developed

Above
Figure 1 - CPR Training.
Image credit: Warp Industries (http://warp. industries/work/ cpr-training/).

a transparent evaluation framework with an associated quality label for medical serious games.

The DSSH is a non-profit organization that strives to improve patient safety and the quality of healthcare, by promoting simulation technology and serious gaming and by stimulating high quality scientific research on these topics. The quality label for serious games allows developers to show that their games have been evaluated as useful, relevant and safe and that the scientific validity of the game was tested through high quality research according to common standards in medical science. It also allows healthcare specialists to distinguish high quality serious games from low quality games, thus supporting them in their decision making of implementing a game. The quality label works with a 5-star rating system, in which a higher number of stars represent higher quality and validity of the games. This article will provide an overview of the application, the evaluation process and a few case examples of evaluated games.

Application & Evaluation Process

All applications for the quality label are reviewed by a separate committee on serious games, which consists of a mix of serious game specialists from various professions and backgrounds (amongst others medical and educational). Owners of a medical serious game are able to apply their game for evaluation, regardless of the genre or purpose of their game. Application for assessment of the game is free and there is no limit to how many games an applicant can enter for evaluation. The evaluation process is started by completing an online application form, which collects all the information necessary to systematically assess the serious game. The contents of this form are based on the consensus-based evaluation framework, described elsewhere in detail.[5]

In short, the evaluation framework contains 62 items, which can be divided into five categories (Table 1). Using this information, every entry is evaluated during a committee meeting. To determine what level of rating is awarded, pre-defined evaluation criteria are used. These criteria were designed in an incremental way and range from simple (e.g. plausibility of

Table 1: Evaluation framework categories

Category	Description
1. Game description	This section contains questions about general game data, affiliations and funding or sponsoring information.
2. Rationale	This section is about the purpose of the game, the intended user groups, and the setting in which the game is going to be used.
3. Functionality	This section further elaborates on how the purposes are to be achieved in the game and how effect is measured.
4. Validity	In this section, owners are required to provide evidence that the game was developed under the right circumstances, with involvement of the right (user) groups and to prove different degrees of validity (ranging from face validity until predictive validity).
5. Data protection	The final section handles all issues regarding data safety and protection, to ensure that data is collected and stored correctly and safely.

Table 2: Quality label evaluation criteria

Star rating	Quality label evaluation criteria - All requirements are cumulative and must be met -
★	1. The entry is a serious game, which contains all necessary game-elements. 2. There is a safe data storage. 3. The working mechanism and theoretical background of the game must be at least plausible.
★★	4. Face-validity has been checked and confirmed by experts. 5. The underlying mechanism has to be supported by evidence.
★★★	6. (para-) medical, educational/psychological and game development experts have to be involved in the development process. 7. Relevant user tests need to be performed, and the results must be processed into the game.
★★★★	8. The game has been validated in a study by independent experts. At least construct validity (proving that a high score in the game correlates to high scores on proficiency tests in real life) needs to be proven.
★★★★★	9. Predictive validity has been confirmed in a scientific validation study, published in a peer-reviewed journal, hereby proving that the game achieves the set learning goals outside of the game.

the concept) to hard-to-achieve (e.g. published results showing the effectiveness of the game in a peer-reviewed journal).

Because of this, a well-developed serious game with a valid and plausible base may be eligible for a star-rating, even though it has not been fully validated (yet), due to the demanding, time-consuming and costly nature of these validation studies.

The first requirement that an entry needs to comply to is that it must actually be a serious game. The DSSH defines a serious game as the following:

"A serious game is an interactive digital application, characterized by a storyline, a clear goal or objective, that is suitable for the target audience. The objective of the game needs to be relevant to accomplish the learning goal, either in a direct or indirect fashion. Interaction with the player is required, e.g. through direct

player feedback or a scoring system and it needs to be an important element to achieve the goal of the game."

Additionally, all entries must comply with the required safety regulations and must have at least a plausible working mechanism and theoretical background. These three criteria must be fulfilled before the serious game is eligible for a quality label. Further evaluation criteria are described in Table 2.

Once the committee achieves consensus about the evaluation, a jury report is created and the applicants are informed on the rating, with narrative information and feedback about their game. Even if the game is not awarded any stars, feedback is provided. If applicable, applicants are encouraged to further improve their game, in order to achieve a higher level of validity of their game, and obtain a (higher) star rating. If the applicant wishes to

apply for a higher rating, a new application must be filed and the process is repeated, ideally until the highest quality, a five star rating can be awarded.

Examples of Awarded Quality Labels

Using this method, the committee has already awarded ten games with a quality label: three games earned a two star rating, a three star rating was awarded to four entries. One game obtained a four star rating, requiring validation by experts and finally, the full five star rating was awarded to two games, which proved predictive validity in a randomized controlled trial.

The following section features three examples of games which have recently been submitted for review and obtained a quality label:

Case 1: CPR training ★★

Game description: The game (Figure 1) is developed to train the basic knowledge skills required for effectively performing CPR. By using virtual reality, the player is put in the middle of an emergency situation, in which he/she needs to make the right choices to save the patient. Players receive feedback throughout the game (reactions of bystanders) and at the end. A score is awarded reflecting the quality of the performance and the choices that were made. Players are encouraged to replay the game, in order to achieve a perfect maximum score.

Committee evaluation: The underlying methods of CPR training and improving knowledge of the procedure are not only plausible, but have proven to be effective and important in order to achieve a successful resuscitation. Furthermore, the content of the game is based on interna-

tional guidelines. Throughout the development of the game, basic life support experts were involved and also confirmed face validity. However, despite this, a three-star rating could not be awarded, since user tests were not performed in the intended user groups. Further user tests should be organized in target user groups, to investigate how the immersive game affects the user experience and learning effects.

Star rating: This serious game was awarded a two-star rating.

Case 2: LAKA ★★★

Game description: The game (Figure 2) was developed to help patients with chronic fatigue or chronic pain, and teach them about mindfulness and meditation. In this serious game, players control an avatar, which they have to guide through various troublesome situations, by making choices and interacting with other characters. Based on their performance, players earn a number of puzzle pieces, reflecting how well they handled the situations. The gameplay consists of an adventure-like story, interrupted by mini-games and meditation exercises.

Committee evaluation: The game is meant to be played as an addition to intensive, multidisciplinary therapy. The principles of mindfulness and meditation on which the main mechanism of the game are based, have been shown to be effective in literature. These principles

have been translated to game mechanics, with the aid of a multidisciplinary expert team. The research group organized extensive user tests with both experts and target user groups, which provided feedback and confirmed face validity.[6] Although the study results were published in a clear paper, construct and predictive validity have not been confirmed yet. In order to achieve this, additional studies with (independent) experts are required. However, the concept of the game looks promising and the committee is interested in further research and validation of this game.

Star rating: This serious game was awarded a three-star rating.

Case 3: AbcdeSIM ★★★★★

Game description: This game (Figure 3) is developed to teach healthcare workers how to resuscitate a critically ill patient according to the ABCDE-principle. The player is a physician in an online simulated emergency department, where the right choices must be made in a timely manner, in order to save the patient. The game features several cases, varying in difficulty.

Committee evaluation: The effectiveness of the theoretical background on which the game was based (ABCDE), has been proven in earlier studies. These principles have been successfully transferred to game-mechanics, using the input of medical, educational and game-experts. Extensive user tests have been undertaken with experts and users, to prove face-validity and assess the opinions of the target group. Additionally, a randomized clinical trial was performed and the results published in peer-reviewed journal.[7] The scientific article described the

effectiveness of this game on clinical competency and knowledge, establishing predictive validity. Although the committee would like to remark that the observed effects may not be completely accountable to playing the game, but also due to extra training time, AbcdeSIM has proven to be an interesting and promising new way of training healthcare workers in core competencies.

Star rating: The committee has awarded this game a five-star rating.

Conclusion

Serious games for healthcare are potentially useful, but often lack transparency regarding their goals, expertise involved in the design and their effectiveness and validity, hampering their implementation in current medical practice. This article described the transparent 5-star based quality label for serious games, developed by the DSSH, and the underlying evaluation framework. At the moment, ten games have already been awarded with a DSSH quality label and the evaluation system appears feasible for clinical practice. The importance and credibility of the label will become apparent in due time, when more games are evaluated and implemented.

About the Authors

The Committee on Serious Gaming of the Dutch Society of Simulation in Healthcare (DSSH) consists of a mix of serious game specialists from various professions and aims to support the collaboration between soft- and hardware developers, game designers, ICT professionals and medical professionals in healthcare. With its 'quality label for serious games', the committee tries to promote and facilitate initiatives contributing to the development, validation and implementation of 'Serious Gaming' in healthcare. **MTM**

REFERENCES

1. Andreatta, P., et al., Simulation-based mock codes significantly correlate with improved pediatric patient cardiopulmonary arrest survival rates. Pediatr Crit Care Med, 2011. 12(1): p. 33-8.
2. Draycott, T., et al., Does training in obstetric emergencies improve neonatal outcome? BJOG, 2006. 113(2): p. 177-82.
3. Haerkens, M.H., et al., Crew Resource Management in the Intensive Care Unit: a prospective 3-year cohort study. Acta Anaesthesiol Scand, 2015. 59(10): p. 1319-29.
4. Hayden, J., et al., The NCSBN National Simulation Study: A longitudinal, randomized controlled study, replacing clinical hours with simulation. Journal of Nursing Regulation, 2014. 5(2): p. 1-64.
5. Kohn, L.T., J. Corrigan, and M.S. Donaldson, To err is human: building a safer health care system, ed. I.o.M. (U.S.). 2000, Washington, DC: National Academic Press.
6. Moffatt-Bruce, S.D., et al., What Is the Return on Investment for Implementation of a Crew Resource Management Program at an Academic Medical Center? Am J Med Qual, 2017. 32(1): p. 5-11.