

A Comparative Study of in-Game and out-Game Assessment for Storyline-based Games

Andreea Molnar
Computing and Communications Department
Lancaster University
InfoLab21, South Dr, Bailrigg, Lancaster
Email: andreea.molnar@lancaster.ac.uk

Jose Garcia Estrada
School of Computing
University of Portsmouth
Winston Churchill Ave, Portsmouth
Email: jose.garcia@port.ac.uk

Abstract—Serious games have the potential to complement existing teaching methods by motivating and providing a more enjoyable experience for the players or by simulating events that would be otherwise difficult to reproduce in the classroom. Despite their potential, little is known about how the games could be used not only for teaching but also as assessment tools. This research addresses this gap. We present an in-game assessment method which assesses the learning objectives included in the game without the need for a separate intervention. We evaluate the proposed method and we show that there is no statistically significant difference in participants being assessed through a questionnaire outside the game and the integrated game assessment method. Moreover, we looked at whether the player experience has been affected by the changes needed in the game design and the players' preferences for different types of assessment. Most participants preferred being assessed through the game. They also felt that the assessment has overall improved their game experience.

I. INTRODUCTION

Educational games are seen as a way to both motivate and teach in a fun way. They can help students acquire skills that would be otherwise difficult to acquire in real life by experiencing the consequences of different actions without suffering real-world consequences (i.e. a student could learn the consequences of an E.coli outbreak without necessarily experiencing it in real life). There has been a lot of research focused on the potential of games to teach and their effectiveness as a teaching tool [1], [2], [3], [4]. However, less research has been focused on how games could be used not only as a tool for teaching but also for assessment. A systematic literature review of serious games assessment has shown that most games evaluation are performed through questionnaires which are performed as separate intervention from the game [5].

Although there are advantages to performing a separate assessment (e.g. some integrated assessment methods were perceived by the players as intrusive, disrupting the game flow and negatively affecting the player experience [6], off-the-shelf games could be used without need modify them) having the assessment integrated into the game and making use of the game analytics to determine player performance has several advantages. It can remove or seriously reduce the test anxiety without affecting the validity or reliability of the results [7]; engage players with the subject taught by providing feedback

[8]; increasing the validity of the assessment due to the increased motivation provided by the games [9] and providing contextual feedback (for players, teachers or parents) without the need for a separate intervention. Feedback encourages 'deep' learning and promotes the engagement with the educational material [8], [10]. On top of the aforementioned reasons, due to the highly motivational potential, educational games could reach people who are not necessarily willing to learn in traditional settings, hence having a method of assessment seamlessly integrated into the game, might provide a method of automatically evaluating their knowledge.

Existing research that focused on using learning analytics for assessment purposes [11], [12], [13], [14], [15] have not assessed whether the use of learning analytics would provide similar results with the traditional assessment methods or have not focused on determining how changes done in the design of the game to collect existing analytics have affected the user experience. In this paper, we focus on addressing this gap. In doing so, we focus on particular kind of games: interactive digital storytelling based games. These are known due to their ability to let the players change the story based on the user actions.

The rest of this paper is organised as follows. Section II introduces the related work focusing on assessment in games. Section III presents our proposed evaluation mechanism. Section IV presents the game we used as a case study and discusses how the in-game assessment has been seamlessly integrated into the game. Section V presents the study performed to validate the proposed mechanism and discusses the results. Section VI presents our plans for future work. Section VII summarises the paper and draws our conclusions.

II. RELATED WORK

Although the evaluation is considered to be an important component of educational games [16], little is known on how to determine what students learned through educational games without disrupting the player experience and without a need of a separate intervention. Games are rarely adopted in the classroom as teachers perceived them as black boxes [13] through which the assurance of learning is difficult to determine. Currently, there is a lack of methodologies on how to assess learning outcomes in games [13]. Learning analytics

suffer from high complexity and costs [17] and there is little known of what learning analytics could tell the stakeholders [18]. The learning analytics used so far in games to measure the learning objectives are hard to generalize [19], and/or often use simple generic traces that make it difficult to gain insight into the player learning achievements. The assessment of the player knowledge is still mostly through questionnaires and interviews [5], often done as a separate part of the game. Studies which have tried to integrate them have often done it in an intrusive manner disrupting the game flow and leading to negatively affecting the player experience [20].

Existing in-game assessment methods use a combination of questionnaires integrated into the educational game and data generated through the educational gameplay [11], [12], [13], [14] to provide an overview of the student learning outcome. Although none of these studies reported user experience being affected, few studies [11], [2] have actually assessed the effect of the way the assessment was integrated into the player experience. [11] have shown that the user experience with an educational platform game has not been affected if integrated as a quiz. The study done in [15] focuses on interactive digital storytelling based game and has shown that if the questions are integrated seamlessly through the interactive digital storytelling based game, the assessment could improve most of the players' experience. As such this study differs from the above studies by assessing not only the player's experience but also comparing different methods of assessment.

III. USER-CENTRED SEAMLESS EVALUATION FRAMEWORK FOR INTERACTIVE DIGITAL STORYTELLING BASED GAMES

The *User-Centred Seamless Evaluation Framework for Interactive Digital Storytelling based Games* [15], [2] is the framework through which the evaluation of the educational content could be seamlessly integrated into the narrative flow. The framework allows the integration of previous player knowledge of the learning objectives into interactive digital storytelling based games (pre-knowledge evaluation), but also of the player knowledge of the learning objectives after playing the interactive digital storytelling based game (post-knowledge evaluation). The pre-knowledge evaluation is done before the player is exposed to the game mechanics which teach the given learning objectives, whereas post-knowledge evaluation is done to assess the player knowledge after s/he is exposed to the game mechanics. In order for the assessment to be seamlessly integrated, the assessment is done through questions asked by the non-player characters, distributed through the interactive digital storytelling and adapted to the context in which the player is (see Figure 1).

In the interactive digital storytelling based game, the questions and all the narratives are delivered through conversation nodes. A conversation node is a narrative displayed to the player at a certain point in time. Each question offers the player multiple options to choose from. Depending on the choice the player makes, the game changes and the interactive digital storytelling based game provides the player with a different



Fig. 1. Example of an In-Game Assessment Question

feedback. A set of *Conversation Rules* is implemented through the interactive digital storytelling based game and decide which piece of information is displayed to the user and a transition function decides what the player can access at a certain point in the interactive digital storytelling based game.

An action could lead to a change in the room the player is in, displaying multiple conversation nodes, etc. A pre-condition to an interactive digital storytelling based game action could be that the player clicks on a hotspot or selects an action path through the game. This leads to a change in the predefined flow of the interactive digital storytelling based game. This allows for the interactive digital storytelling to adapt both to the player actions and to the user knowledge. Immediate feedback on the learning outcome is provided to the player which could lead to a better player engagement with the educational content by providing immediate feedback [8]. The data collected through the player interactions with the game is also used to provide the player with a score which does not only reflect his game performance but also through his academic improvement through the game.

IV. CASE STUDY: GLOBAL HANDWASHING DAY GAME

We used Global Handwashing Day game [15] as a case study. Global Handwashing Day game is part of the e-bug/edugames4all collection of educational games that teach about microbiology (see Figure 2). Edugames4all consists of several platform [21] and interactive storytelling based games [15] aimed at different age groups and developed for different platforms [21] The content of the game was developed in collaboration with experts in healthcare and education [22].

Among these games, all the interactive digital storytelling based games, including Global Handwashing Day game, follow a problem-based learning approach [23]. The game aims to reinforce the importance of hygiene, focusing on micro-organisms transmission. During the game, the player is a detective that has to tackle a mystery regarding the poisoning of a famous actor. Through the game, the player gathers evidence from different scenes that could be related to the

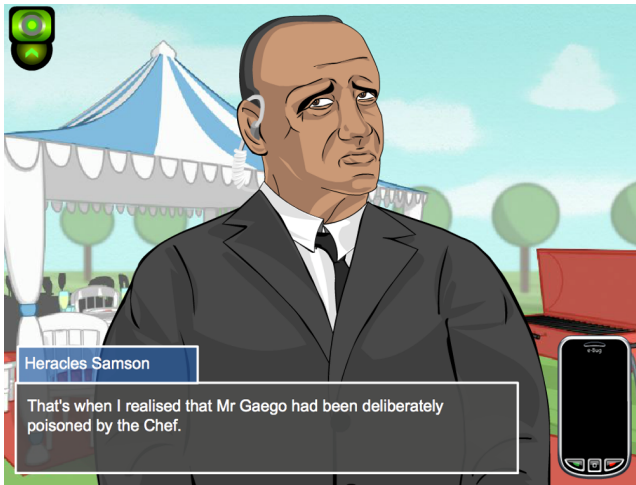


Fig. 2. Global Handwashing Day Game.

investigation. The player questions different characters for information. The player must be able to use the evidence collected during the interrogation, as well as see who the actual suspects are. Based on the evidence gathered in the testimonies, the player has to decide who the guilty party is.

As the game is not linear, it allows players to explore different parts of the game and allows choosing different potential suspects during the investigation. Not all the paths the user can take through the game lead to an answer and they are not all mandatory for solving the mystery. While trying to solve the mystery, the player is taught about microbiology.

The game has been shown to be effective in delivering the learning outcomes [15] and as such it was considered a good case study for the integrated assessment. Small changes have been done through the game to improve its usability and to collect the data necessary to assess the learning performance. The proposed assessment has been integrated into the game based on the above framework and data about the learning objectives performance was collected. In this case, the score the students obtained in the game was just based on the user learning performance and not on how well they advance in the game.

V. EVALUATION

The first aim of this study was to determine whether the integrated assessment provides similar results with the paper-based assessment. In doing so we organised a quasi-experimental study [24]. We divided the participants into *control group* and *experimental group*. The control group played the game without the assessment and was assessed only through the questionnaire. The experimental group played the game modified to collect and interpret data regarding their learning performance. In the study we assessed whether (a) the control group questionnaire assessment was statistically significant different from the experimental group in-game assessment and (b) whether the experimental group knowledge of the learning objectives as evaluated by a questionnaire

before playing the game, was similar to the one determined by the game before they were exposed to the game mechanics.

The second aim of the study was to determine whether the players preferred this method of assessment as opposed to the traditional one and in what way it has affected the user experience. To do so, after the experimental group players finished the game they participated in an interview. During the interview, the *experimental group* participants were asked whether they would prefer this method to a questionnaire based assessment.

A. Methodology

A quasi-experimental [24] study was organised. Written informed consent was obtained from all participants before starting the study. The participants were divided into two groups: a control group and an experimental group. All the participants regardless of the group were asked to complete a questionnaire at the beginning of the study collecting demographic data and assessing their knowledge of the learning objectives covered in the game:

- **LO-1:** Microbes found in food can transfer to humans
- **LO-2:** Separate utensils should be used for raw meat, and vegetables
- **LO-3:** Bacteria from raw meat can make a person sick
- **LO-4:** Food cooked properly should be free of bacteria
- **LO-5:** Vomiting viruses are unpleasant but usually not dangerous
- **LO-6:** Vomiting viruses can spread through sneezing, coughing or just particles of vomit that are in the air after someone is sick
- **LO-7:** Vomiting viruses and E.coli can spread through bad hygiene
- **LO-8:** It is not always necessary to take medicine when dealing with E.coli and vomiting viruses infections
- **LO-9:** E.coli is commonly found in the lower intestine
- **LO-10:** E.coli can spread through the 'faecal-oral' route or poor food preparation hygiene
- **LO-11:** If eaten, bacteria from raw meat can make a person sick

The knowledge was assessed through multiple choice questions covering the learning objectives above. In order not to affect the validity of the results, for the experimental group, the data collected about the initial student knowledge through the Global Handwashing Day game was evaluated before the player being exposed to the game mechanics. For the control group, the evaluation of the player knowledge was performed before playing the game by completing a questionnaire but also through the in-game assessment.

The questionnaire used to assess students knowledge was validated by an expert in microbiology. Both the control and experimental group played the Global Handwashing Day game until the end, the only difference between the two groups being that the student's knowledge was being tracked by the experimental group. The game took around 1h to complete.

We use the Student t-test [25] to analyse the statistically significant difference between the control and experimental

TABLE I
OVERALL STUDENT PERFORMANCE: AVERAGE, MEAN, STANDARD
DEVIATION AND STANDARD ERROR

Assessment Type	Average	Mean	δ	Std. Error
Questionnaire	9	9.10	0.968	0.21
In-Game	9	9.35	1.04	0.23

group, considering confidence interval of 95%. The same method was used to whether there is any statistically significant difference between the player’s answers before playing the game and their knowledge determined by the game before they were exposed to the game mechanics.

The participants in the experimental group were asked to volunteer for an interview after playing the game. The interview was analysed using inductive thematic analysis [26]. In the analysis of the data we followed five stages: familiarisation, identifying a thematic framework, indexing, charting and mapping and interpretation [27].

B. Participants

A total of 40 people took part in the study. The participants were all adults (19 to 40 years old, mean age 28, std. 9.13). Most of the participants (72%) were male. The participants were equally divided between the two groups (i.e. 20 participants in the experimental group and 20 participants in the control group). They all volunteered to take part in the study. The participants were offered a small incentive to take part in the study.

Before playing the game the participants were asked to complete a questionnaire assessing their knowledge of the learning objectives covered in the game. To avoid differences in participants knowledge between the two groups we compared their answers in the questionnaire. This has shown that there was no statistically significant difference ($p=0.64$, $t=-0.45$, $mean=-0.30$, $std. deviation=2.92$) between the two groups knowledge of the learning objectives covered in the game.

C. Results: Assessment Comparison - Within Study

To measure the overall learning performance we compare the experimental group participants responses to an online questionnaire covering the learning objectives taught in the game with those seamlessly integrated into the game. The questions asked outside the game were defined in an abstract manner while the ones in the game are adapted to the game context, although whenever possible the questions were kept as abstract as possible to ensure information transferability. Both the online questionnaire and the in-game integration of the learning objectives were used and were validated in previous studies [15].

The results show no statistically significant difference in the results of the player questionnaire and player performance considering a 95% confidence interval ($p=0.437$, $t=-0.794$, $\delta=1.41$, $x=-0.25$). The average number of questions answered correctly, mean and standard deviation among the two survey answers are presented in Table II.

TABLE II
CONTROL AND EXPERIMENTAL GROUP COMPARISON

Group Type	Mean	St. Deviation	Std. Error
Control	9.40	2.72	0.60
Experimental	9.35	1.04	0.23

D. Results: Assessment Comparison - Between Study

We compare the results of the control group on the questionnaires with the experimental group as assessed through the game. The results show no statistically significant difference in the knowledge evaluation results between the experimental and the control group considering a 95% confidence interval ($p=0.94$, $t=0.77$, $mean=0.050$, $std. deviation=2.91$). Table II presents detail results about the differences between the two groups.

E. Results: Qualitative Feedback on Assessment Preference and Player Experience

After the study, the experimental group participants took part in a short interview which aimed to understand players preferences for how the assessment is integrated and their experience with the assessment in the game. The results of the interview showed that most of the participants (75%) prefer the assessment to be integrated into the game. They felt it was more contextually appropriate and relevant (e.g. it "was exploring using real-world situation"), that the questions asked through the game played make the players "feel involved" and they were seen as an incentive to continue playing the game and think about the narratives. Some of the participants felt that they were more relaxed being assessed through the game rather than having a quiz at the end. Overall, the players who preferred this kind of assessment felt that it help with their experience.

The participants stated that they felt that it was useful to have a final point to reflect upon what they learned through the game and see the educational results of playing the game. The game as it now does not provide the players with an overview of how much their knowledge has improved after playing the game.

VI. FUTURE WORK

In our future work, to ensure that the results are generalisable, we want to extend this study to a larger sample of participants. We also want to determine whether this type of assessment can be successfully integrated and evaluated across other interactive digital storytelling based games. We want in the future to evaluate the effect of user experience in the interactive digital storytelling based game as opposed to a similar game without the seamless evaluation integrated. We are also planning to research methods of using learning analytics which could generalise across different games genres.

A lack of visualisations of the player’s educational performance through the game was mentioned by some of the participants as an issue when it comes to this kind of assessments. Therefore in the future, we want to change the game design

so that it could better display the player performance through the game.

VII. CONCLUSIONS

This research presented a method of integrating assessment in interactive digital storytelling based games. As a case study, we used the Global Handwashing Day [15], an interactive digital storytelling based game that teaches about microbe transmission. A quasi-experimental study was performed with 40 participants divided into two groups: control group - for which the assessment was evaluated through a questionnaire- and experimental group - for which the assessment was performed by interpreting the data collected through the game.

The results of this study showed that there is no statistical significance between the assessment as performed in the game (experimental group) and the one administered through questionnaires (control group). Moreover, no statistically significant difference was obtained for the assessment performed for the experimental group through questionnaires or in the game assessment. Most of the participants in the experimental group reported preferring the assessment integrated into the game as opposed to a separate assessment through questionnaires.

ACKNOWLEDGMENT

The authors would like to thank: those who have volunteered to take part in the study; Research and Development Fund at the University of Portsmouth who funded the study; DG SANGO for funding the initial implementation of interactive digital storytelling; and Dr. Patty Kotskova for her ongoing support.

REFERENCES

- [1] T. Hainey, T. M. Connolly, E. A. Boyle, A. Wilson, and A. Razak, "A systematic literature review of games-based learning empirical evidence in primary education," *Computers & Education*, vol. 102, pp. 202–223, 2016.
- [2] A. Molnar and P. Kostkova, "Interactive digital storytelling based educational games: formalise, author, play, educate and enjoy!-the edugames4all project framework," in *Transactions on Edutainment XII*. Springer, 2016, pp. 1–20.
- [3] B. Schmitz, R. Klemke, J. Walhout, and M. Specht, "Attuning a mobile simulation game for school children using a design-based research approach," *Computers & Education*, vol. 81, pp. 35–48, 2015.
- [4] J. Torrente, B. Borro-Escribano, M. Freire, Á. del Blanco, E. J. Marchiori, I. Martínez-Ortiz, P. Moreno-Ger, and B. Fernández-Manjón, "Development of game-like simulations for procedural knowledge in healthcare education," *IEEE Transactions on Learning Technologies*, vol. 7, no. 1, pp. 69–82, 2014.
- [5] A. Calderón and M. Ruiz, "A systematic literature review on serious games evaluation: An application to software project management," *Computers & Education*, vol. 87, pp. 396–422, 2015.
- [6] V. Wendel, S. Göbel, and R. Steinmetz, "Seamless learning in serious games-how to improve seamless learning-content integration in serious games," in *CSEdu (1)*, 2011, pp. 219–224.
- [7] V. J. Shute, L. Wang, S. Greiff, W. Zhao, and G. Moore, "Measuring problem solving skills via stealth assessment in an engaging video game," *Computers in Human Behavior*, vol. 63, pp. 106–117, 2016.
- [8] R. Higgins, P. Hartley, and A. Skelton, "The conscientious consumer: Reconsidering the role of assessment feedback in student learning," *Studies in higher education*, vol. 27, no. 1, pp. 53–64, 2002.
- [9] K. L. McClarty, A. Orr, P. M. Frey, R. P. Dolan, V. Vassileva, and A. McVay, "A literature review of gaming in education," *Gaming in education*, pp. 1–36, 2012.
- [10] F.-H. Tsai, C.-C. Tsai, and K.-Y. Lin, "The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment," *Computers & Education*, vol. 81, pp. 259–269, 2015.
- [11] P. Kostkova and A. Molnar, "Educational games for creating awareness about health issues: The case of educational content evaluation integrated in the game," in *Medicine 2.0 Conference*. JMIR Publications Inc., Toronto, Canada, 2014.
- [12] A. Molnar, J. Virseda, and V. Frias-Martinez, "Insights from educamovil: involving teachers in creating educational content for mobile learning games," *Journal of Interactive Learning Research (JILR)*, vol. 26, no. 2, pp. 209–221, 2015.
- [13] A. Serrano-Laguna, B. Manero, M. Freire, and B. Fernández-Manjón, "A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes," *Multimedia Tools and Applications*, pp. 1–23, 2017.
- [14] M. Taub, N. V. Mudrick, R. Azevedo, G. C. Millar, J. Rowe, and J. Lester, "Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with crystal island," *Computers in Human Behavior*, 2017.
- [15] A. Molnar and P. Kostkova, "Seamless evaluation integration into ids educational games," in *International Conference on the Foundations of Digital Games*, 2013.
- [16] S. Liu and W. Ding, "An approach to evaluation component design in building serious game," in *International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 2009, pp. 141–148.
- [17] M. S. El-Nasr, A. Drachen, and A. Canossa, "Game analytics," *New York, Sprint*, 2013.
- [18] M.-T. Cheng, L. Rosenheck, C.-Y. Lin, and E. Klopfer, "Analyzing gameplay data to inform feedback loops in the radix endeavor," *Computers & Education*, vol. 111, pp. 60–73, 2017.
- [19] Á. Serrano-Laguna, J. Torrente, P. Moreno-Ger, and B. Fernández-Manjón, "Application of learning analytics in educational videogames," *Entertainment Computing*, vol. 5, no. 4, pp. 313–322, 2014.
- [20] M. Trevathan, M. Peters, J. Willis, and L. Sansing, "Serious games classroom implementation: Teacher perspectives and student learning outcomes," in *Society for Information Technology & Teacher Education International Conference*, vol. 2016, no. 1, 2016, pp. 624–631.
- [21] A. Molnar and P. Kostkova, "Learning about hygiene and antibiotic resistance through mobile games: Evaluation of learning effectiveness," in *Proceedings of the 2018 International Conference on Digital Health*. ACM, 2018, pp. 95–99.
- [22] D. Farrell, P. Kostkova, J. Weinberg, L. Lazareck, D. Weerasinghe, D. M. Lecky, and C. A. McNulty, "Computer games to teach hygiene: An evaluation of the e-bug junior game," *Journal of Antimicrobial Chemotherapy*, vol. 66, no. suppl 5, pp. v39–v44, 2011.
- [23] D. Farrell, P. Kostkova, D. Lecky, and C. McNulty, "Teaching children hygiene using problem based learning: the story telling approach to games based learning," in *CEUR Workshop Proceedings*, vol. 498. CEUR Workshop Proceedings, 2009.
- [24] D. T. Campbell and J. C. Stanley, *Experimental and quasi-experimental designs for research*. Ravenio Books, 2015.
- [25] W. J. Dixon, F. J. Massey *et al.*, *Introduction to statistical analysis*. McGraw-Hill New York, 1969, vol. 344.
- [26] G. Guest, K. M. MacQueen, and E. E. Namey, *Applied thematic analysis*. Sage, 2011.
- [27] A. Srivastava and S. B. Thomson, "Framework analysis: a qualitative methodology for applied policy research," 2009.