Towards a Learning Game Evaluation Methodology in a Training Context: a Literature Review

Fahima Djelil, Eric Sanchez, Benjamin Albouy-Kissi, Jean-Marc Lavest, Adélaïde Albouy-Kissi

Towards a Learning Game Evaluation Methodology in a Training Context: a Literature Review

Fahima Djelil[1], Eric Sanchez[2], Benjamin Albouy-Kissi[1], Jean-Marc Lavest[1], Adélaïde Albouy-Kissi[3]
[1]Laboratoire Institut Pascal CNRS-UMR 6602, Université Blaise Pascal, Clermont-Ferrand, France
[2]EducTice, Institut Français de l'Education, Ecole Normale Supérieure, Lyon, France
[3]Laboratoire ISIT CNRS-UMR 6284, Université d'Auvergne, Clermont-Ferrand, France
fahima.djelil@udamail.fr
eric.sanchez@ens-lyon.fr
benjamin.albouy@udamail.fr
j-marc.lavest@udamail.fr
adelaide.kissi@udamail.fr

Abstract: Given their diverse dimensions and characteristics, learning games are difficult to evaluate. However, evaluation remains the only way to verify if the educational targets are being achieved and to detect any functional vulnerability inside the learning game. Based on a review of literature, in this paper we propose to structure the evaluation process in the design and experimental phases, using analytical and empirical evaluation methods in order to reduce the risk of a weak designed learning game. Indeed, we suggest that a well designed learning game should be useful from a pedagogical point of view, usable from a learner point of view and acceptable from the institution, the teacher/trainer and the learner points of views. For this purpose, we outlined four criteria classes that influence the evaluation process of such learning game. The first class criteria is related to the ludopedagogical environment which gathers learning and entertainment aspects of the learning game, the second class is related to the learner affective and cognitive reactions that are formed when using the learning game, the third class is related to the training context regarding to the curriculum and the learning games' operational conditions, finally, the fourth class is regarding to the learner profile which gathers information about his competencies background and preferences. After introducing the concepts of usability, usefulness and acceptability as well as the current analytical and empirical evaluation practices which exist in the field of game based learning, the analysis and measurement criteria are introduced before associating them with each of these three evaluation dimensions.

Keywords: Game Based Learning, Evaluation, Usability, Usefulness, Acceptability.

1. Introduction

Learning games evaluation practices generally aim to ensure their design quality (Di Loreto & Gouaich, 2010), diagnose their uses (Loh, 2012) and verify their instructional outcomes (Papastergiou, 2009). However, the evaluation of that kind of software is complex and not so well structured (Thomas et al, 2012). Indeed, the evaluation often undergoes search and definition steps for measurement variables and appropriate analysis methods for these variables. In addition, al evaluations do not provide coherent results (Von Wangenheim et al , 2009; Annetta et al, 2011), despite the many existing methods for data collection and analysis which have proved their effectiveness in practice. However, although evaluation is complex and time consuming, it remains the only means to verify the achievement of the training goals and to detect any malfunction inside the learning game. Therefore, a learning game ought to be evaluated prior to being used as a learning material.

The purpose of this paper is to propose an approach to evaluate learning games in a training context, which is based on a review of literature on learning game's evaluation practices, as well as theoretical and experimental models for designing effective instructional games. We addressed three primary questions upon which the paper is structured. First what are the objectives of a learning game evaluation that can help us structuring such process? Second, what are the current evaluation practices and methods that can help us to know at what stage of the development process should we evaluate? Third, what are the primary characteristics of learning games that can be exploited as evaluation criteria?

Furthermore, we are situated in a context of academic training and we mean by learning game *software designed for learning in an entertaining way, well-defined concepts and knowledge within a well-established training context.* A learning game can be an educational game, a serious game or any game based learning software.

2. Usability, usefulness and acceptability: three dimensions for evaluating learning games

The concepts of usability, usefulness and acceptability have been the subject of considerable research and literature in the field of learning environments (Tricot et al, 2003; Barbel, 2003). In the field of game based learning, contrary to usability which is one of the most addressed point in game based learning assessment (Warren et al, 2011; Aikaterini et al, 2013), usefulness and acceptability seem not to be as well widespread. However, we believe that an evaluation based on these three dimensions would simplify the complex relationship that exists between the functional properties of that type of learning software and those of its use and exploitation.

Szilas & Sutter Widmer (2013) state that usability of a learning game reflects the possibility of its use in a training context and depends as any other learning software, on the quality of its ergonomics. According to Warren et al (2011) usability testing is the process of examining those components in a software that impact the ability of the user to complete successfully those tasks that the designer intends them to be able to complete. Therfore, evaluating usability helps to ensure that the game is playable and that its instructional content is correct before putting it in classrooms (Warren et al, 2011).

Software usefulness underlines its relevance and appropriateness according to the user high-level objectives (Senach, 1999). In the field of computing learning environments, user higher-level objectives are related to the learning goals (Nogry et al, 2004). In this sense, evaluation of usefulness helps to verify the impact of the learning software on the learner's knowledge and skills (Barbel, 2003; Nogry et al, 2004). This can be also applied to learning games in order to verify the learning progression and to ensure the achievement of the learning targets (Sanchez, 2011b).

Acceptability is related to the mental representation of the learning software value comprising attitudes and opinions about the learning artifact (Tricot et al, 2003). The value of this mental representation affects considerably the use of the learning artifact (Tricot et al, 2003). Szilas & Sutter Widmer (2013) underline the impact of the exploitation context to determine this value. According to Sanchez (2011b), acceptability is determined by the judgement value of the student, the trainer and the institution regarding to the learning game usability and usefulness.

3. Analytical and empirical evaluation methods for learning games

Analytical and empirical evaluation methods are widely known, both in human machine interfaces and computing learning environments (Senach, 1999; Tricot et al, 2003; Nogry et al, 2004). However, this terminology has, to our knowledge, not yet been adopted in the field of learning games, although practices are really present.

Many researchers have presented different grids and models based on game design factors and games characteristics devoted to the design and analysis of learning games (Garris et al, 2002; Kiili, 2005 ; De Freitas & Oliver, 2006; Amory & Seagram, 2007; Annetta et al, 2011; Sanchez, 2013). These models provide exhaustive design guidelines before being used to evaluate the effectiveness of the resulting learning game, as has been done in (Marne et al, 2012; Marfisi et al, 2012). Applying these models in the design phase of the development process helps to detect any functional limitation of the learning game before the implementation phase (Di Loreto & Gouaich, 2010). This is typically analytical evaluation of learning games.

Empirical evaluation of a learning game consists in examining experimental data arising from activity traces, questionnaires, interviews, observational notes, audio and video records …etc. This kind of evaluation practices is regarding to the use of the learning artifact and requires mock-ups, prototypes or final designed software to be available as well as user testers. Such related works have been conducted by many researchers (Navarro & Van Der Hoek, 2007; Papastergiou, 2009; Von Wangenheim et al, 2009; Dunleavy & Simmons, 2011 ; Sanchez, 2011c ; Loh, 2012). The collected empirical data is then analyzed statically or automatically using appropriate algorithms in order to get significant data helping to diagnose the learner competencies (Thomas et al, 2012), verify pedagogical effectiveness of the learning game such as the achievement of the instructional goals (Papastergiou, 2009 ; Von Wangenheim et al, 2009; Loh, 2012) and to detect any functional weakness inside the learning game (Brown, 2011; Manin et al, 2006).

Based on this brief summary, we concluded that evaluating usability, usefulness and acceptability can be achieved using analytical and empirical evaluation methods which are complementary. This should also be performed along the design and the experimental phases of the learning game development process.

4. Key criteria for evaluting learning games

Based on a review of literature, we concluded that learning games evaluation criteria can be described in terms of four categories: ludopedagogical environment, learner affective and cognitive reactions, training context and learner profile. These categories are described below, and related criteria are detailed according to addressed research.

4.1. Ludopedagogical environment related criteria

We define a ludopedagogical environment as a computing environment which gathers all the interactive components of the learning game in which learners will actively construct knowledge. This environment has a set of pedagogical and entertaining characteristics that help learners to enhance their motivation and accomplish instructional objectives. Essential characteristics we retain as evaluation criteria are the following:

- *Goals and rules:* the learning game should have specific and clear goals that match with the instructional objectives describing the targeted skills and knowledge (Marfisi et al, 2012; Sanchez, 2013). The achievement of these goals depends on a set of rules which consist on constraints the learner has to fulfill (Garris et al, 2002; Kiili, 2005).
- *Representational modes*: consist on the internal representational world of the learning game, which in our context we use to mean the representational learning resources that may include dynamic graphics, sound effects, text…etc. This may also include metaphors (Sanchez, 2011a) and narration (Dondlinger, 2007). The representational modes should be designed in a way that leads to enhanced motivation and performance (Garris et al, 2002).
- *Appropriate and progressive difficulty levels*: the learning game activities should be neither too easy nor too difficult to perform and should employ progressive difficulty levels (Garris et al, 2002; Sanchez, 2011a ; Marfisi et al, 2012). This leads to a progressive learning and increases learner interest in targeted training (Kiili, 2005; Annetta et al, 2011).
- *Interaction:* the learning game should be highly interactive providing an enjoyable playing experience (Dondlinger, 2007). Interaction should be easy and instructive in order to help learners to understand new perspectives, prevent errors and enhance ease of use (Dondlinger, 2007; Marfisi et al, 2012).
- *Feedback*: is provided when the learner interacts with the learning game (Dondlinger, 2007). This needs to be immediate, instructive and relevant to the game goals in order to provide enough guidance helping learners to consider strategies and make decisions (Annetta et al, 2011). Feedback provides also an assessment of progress that drives the learner to expend more effort and to focus attention on his tasks (Garris et al, 2002).
- *Failures and mistakes*: learners should be encouraged to persevere despite failure. Failures and mistakes have no real-world consequences (Garris et al, 2002; Marfisi et al, 2012; Sanchez, 2013), they are rather turned into learning moments where learners have to question and explore different options to achieve their tasks (Annetta et al., 2011).
- *Prologue:* consists on an introduction clearly communicating the learning game's goals, this helps the learner to early understand the game's goals in order to facilitate its use (Annetta et al, 2011).

4.2. Learner's affective and cognitive reactions related criteria

As learners interact with the learning game, a set of emotions, feelings and attitudes are triggered (Garris et al, 2002). This is inherent to the game's characteristics which are incorporated in the ludopedagogical environment (Amory, 2007). These affective reactions include judgments such as enjoyment and interest as well as behaviors such as greater persistence and motivation. These reactions are also cognitive because they lead to the achievement of training objectives and specific learning outcomes (Garris et al, 2002). These reactions are mainly represented by self-efficacy, entertainment, immersion, decision making, strategy developing, interest and motivation.

- *Self-efficacy*: this feeling develops when a learner succeed in performing tasks to the extent that he is allowed to select strategies and make decisions (Dondlinger, 2007 ; Annetta et al,

2011). This helps learners to be more confident and evokes a sense of control that leads to increased motivation and greater learning (Garris et al, 2002; Kiili, 2005).

- *Entertainment*: one central characteristic of games is that they are fun and a source of enjoyment. This is inherent to the design quality of the ludopedagogical environment including richness and variety of information presented to the learner (Marfisi et al, 2012). This is related also to whether the learning content and the difficulty levels are suited to the learner's prior knowledge, expectations and characteristics (Garris et al, 2002).
- *Immersion*: occurs when a learner experiences a high degree of control and is being absorbed to the extent that he may lose a sense of time and self (Annetta et al, 2011). This is also inherent to the design quality of the ludopedagogical environment (De Freitas & Oliver, 2006).
- *Decision making and strategy developing*: the learner is allowed to make decisions and set strategies when solving problems and performing tasks (Sanchez 2011b).
- *Interest*: is a subjective learner judgment regarding whether the game is fun, engaging and suited to the learner's expectations (Garris et al, 2002). This is inherent to the design quality of the ludopedagogical environment and the relevance of the learning content (Marfisi et al, 2012).
- *Motivation:* consist on the learner's engagement and implication in the game's tasks on which he becomes more involved and devotes more time and effort (Garris et al, 2002; Annetta et al, 2011).

4.3. Training context related criteria

We mean by training context the particular context where learning will take place using the learning game. This may include several factors such as the availability of specific resources and technologies, the organization of the curriculum, the time devoted to the use of the learning game, the use of additional learning resources and the presence of trainers/teachers (De Freitas & Oliver, 2006; Marfisi et al, 2012).

4.4. Learner profile related criteria

This criteria category focuses upon attributes of the learner. This may include the age and level, as well as specific information of how he learns including training background, styles and preferences (De Freitas & Oliver, 2006). This may also include a prior experience use of learning games.

5. Analytical and empirical evaluation of the usability, usefulness and acceptability of a learning game

5.1. Usability evaluation

Learning game usability is related to the design quality of the ludopedagogical environment as well as some specific learner's reactions trigged when using the game. Therefore, analytical usability evaluation should be accomplished during the design phase in order to check whether criteria related to ease of use as well as learner's enjoyment are being incorporated into the learning game. An empirical evaluation would help to verify how the learning game is easy to use, and how the ergonomic and entertaining aspects are suited to the learners characteristics. This involves user testers and should be achieved in preference from an early development stage such as a prototyping phase. Table 1 lists criteria retained for this purpose.

Table 1: Analytical and empirical criteria for usability evaluation

Criteria	Analytical evaluation	Empirical evaluation
Representational modes	Check their adequacy to the learner preferences, age and level	Verify the learner satisfaction and subjective judgment
Interaction	Check if the interaction will be fun and easy to understand	Verify the learner satisfaction and subjective judgment
Entertainment	Check if entertaining features are being incorporated into the learning game	Verify the learner satisfaction and subjective judgment
Prologue	Check if it is designed so that it can help learners to understand the game's goals	Verify its impact on ease of use

5.2. Usefulness evaluation

Utility is related to the learning objectives. Thus, evaluating utility is regarding to those criteria that trigger cognitive reactions and enhance learning. Analytical evaluation helps to check if these criteria are being considered during the design process, while empirical evaluation will verify the relevance of these criteria to the expected learning as well as the achievement of the instructional targets during the experimental phase. Table 2 lists criteria retained for this purpose.

Table 2: Analytical and empirical criteria for usefulness evaluation

Criteria	Analytical evaluation	Empirical evaluation
Goals and rules	Specify the learning objectives to match with the game's goals; Check the adequacy of game's rules to the learning objectives	Verify the learning effectiveness (analyze those skills that have been acquired and those that have not, test comprehension, memorizing, and learning mastery)
Decision making and strategy developing	Check if decision making and strategy developing features are being incorporated into the learning game	Verify how decision making and strategies developing may impact the learning effectiveness
Representational modes	Check their adequacy to the learning objectives	Verify their relevance to the learning objectives
Interaction	Check its adequacy to the learning objectives	Verify its relevance to the learning objectives
Feedback	Set a feedback that will be immediate and instructive	Verify its impact on decision making and strategy developing
Difficulty levels	Set appropriate and progressive difficulty levels	Verify how appropriate difficulty levels impact the learning progression
Failures and mistakes	Ensure that failures and mistakes will not be hardly sanctioned	Verify how this can impact the learning progression
Self-efficacy	Set game sequences where learners can experience control	Verify how this can impact the learning progression and enhance performance
Immersion	Set an appropriate immersion level	Verify how this can impact the learning progression

5.3. Acceptability evaluation

Learning game acceptability is related to the exploitation context and concerns its relevance to available technologies and the instructional curriculum. It relates to the learner, trainer/teacher and the institution judgment on the learning game value. Therefore, analytical acceptability evaluation should be accomplished at an early design stage in order to check whether the available resources are suited to the learning game requirements as well as the learner's characteristics. Empirical evaluation should verify how the learning game can enhance the learner's interest and motivation. This will ensure the development of a learning game suitable to the expectations of the learner, the trainer/teacher and the institution. Criteria retained for this purpose are presented in table 3.

Table 3: Analytical and empirical criteria for acceptability evaluation

Criteria	Analytical evaluation	Empirical evaluation
Training context	Check if required resources and technologies are available	Relevance of the game's activities to the expectations of the learner, the trainer/teacher and the institution
Learner profile	Check whether the learner profile is being considered during the design process	Relevance of the game's activities to the learner profile
Interest	Learner's interest is not possible to check at an early stage	Behavior observation and usage analysis; Verify the learner satisfaction and subjective judgment
Motivation	Learner's motivation is not possible to check at an early stage	Behavior observation and usage analysis

6. Conclusion

We have presented our approach to evaluate learning games in which we propose that specific game's characteristics can trigger the evaluation process. This approach is based on a review of literature and focuses upon four category criteria namely the ludopedagogical environment in which learners will actively construct knowledge and accomplish instructional objectives, the learner's affective and cognitive reactions including subjective judgment and behavior that lead to the achievement of the learning objectives, the training context in which the learning game will be used and finally the learner's profile that gathers attributes regarding to his learning background, preferences and characteristics.

Moreover, we situate the evaluation process at the design and experimental phases of the learning game development process and point the importance of analytical and empirical methods. We propose also to structure evaluation on three dimensions namely, usability, usefulness and acceptability which represent the main objectives of such evaluation. We have assigned to each dimension the relevant criteria while taking into account analytical and empirical evaluation considerations.

It is the authors' belief that this approach is a valuable tool for the development of learning games and can be useful to game designers to design and evaluate their own games. However, there are some basic issues of how to apply this methodology to evaluate learning games. To examine this, further experimentations are needed. In our current and future work, we design and develop a virtual reality game for learning object-oriented programming for beginner students in a context of academic training. This methodology is meant to serve as a guide in the design and the development of this learning game. The four category criteria presented in this paper, will be used at an early design stage to structure the evaluation of the usability, usefulness and acceptability of the learning game. We will verify in an analytical way the incorporation of the presented criteria into the learning game. We also intend to use activity traces, questionnaires, interviews, as well as observational notes to conduct an empirical evaluation with user testers once a first prototype is implemented. This will help us to examine the degree to which our current approach is applicable, and how it can help in the design of learning games which are relevant to the expectations and needs of the learner, the trainer/teacher and the institution.

Acknowledgement

This work is funded by the French ministry of higher education and research within Tactileo project (http://project.tactileo.net).

References

Aikaterini, K., Apostolos, M. and Thrasyvoulos, T. (2013) Game based learning in mathematics: teachers' support by a flexible tool*, In Proceedings of the 7th European Conference on Games Based Learning.*

Amory, A. (2007) Game object model version II: a theoretical framework for educational game development, *Educational Technology Research and Development,* Vol 55, No. 1, pp 51-77.

Amory, A. and Seagram, R. (2007) Educational game models: conceptualization and evaluation, *South African Journal of Higher Education,* Vol 17, No. 2, pp 206-217.

Annetta, L., Lamb, R. and Stone, M. (2011) Assessing Serious Educational Games, *Serious Educational Game Assessment,* Springer, pp 75-93.

Barbel, P. (2003) Liens Entre Utilisabilité et Utilité d'un Logiciel Educatif en Situation d'Enseignement, In Proceedings of *Environnements Informatiques pour l'Apprentissage Humain*, pp 405-412.

Brown, R. (2011) Assessment Using After-action Review, *Serious Educational Game Assessment,* Springer, pp 119-129.

De Freitas, S. and Oliver, M. (2006) How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?, *Computers & Education,* 46(3), pp. 249-264.

Di Loreto, I. and Gouaich, A. (2010) An Early Evaluation Method for Social Presence in Serious games*, In Proceedings of the 2nd International Conference on Computer Supported Education.*

Dondlinger, M. J. (2007) Educational Video Game Design: A Review of the Literature, *Journal of Applied Educational Technology,* Vol 4, No. 1, pp 21-31.

Dunleavy, M. and Simmons, B.(2011) Assessing Learning and Identity in Augmented Reality Science Games, *Serious Educational Game Assessment, Springer,* pp 221-240.
Garris, R., Ahlers, R. & Driskell, J. E. (2002) Games, motivation, and learning: A research and practice model, *Simulation & gaming,* Vol 33, No.4, pp 441-467.
Kiili, K. (2005) Digital game-based learning: Towards an experiential gaming model, *The Internet and higher education,* Vol 8, No. 1, pp 13-24.
Loh, C. S. (2012) Information Trails: In-process Assessment of Game-Based Learning, *Assessment in Game-Based Learning,* pp 123-144.
Manin, N., George, S. and Prévot, P. (2006) Using virtual learners' behaviours to help the development of educational business games, In Proceedings of *Innovative Approaches for Learning and Knowledge Sharing*, pp 287-301.
Marfisi, I., George, S., Tarpin-Bernard, F. and Prévot, P. (2012) Comment évaluer la qualité d'un Learning Game pendant sa conception ?, In Proceedings of *Technologies de l'information et de la communication pour l'enseignement*, pp 80-90.
Marne, B., Wisdom, J., Huynh-Kim-Bang, B. and Labat, J.M. (2012) The six facets of serious game design: a methodology enhanced by our design pattern library, *21st Century Learning for 21st Century Skills*, Springer, pp 208-221.
Navarro, E. O. and Van Der Hoek, A.(2007) Comprehensive evaluation of an educational software engineering simulation environment, In Proceedings of the *20th Conference on Software Engineering Education & Training*, pp 195-202.
Nogry, S., Jean-Daubias, S. and Ollagnier-Beldame, M. (2004) Evaluation des EIAH: une nécessaire diversité des méthodes, In Proceedings of *Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie*, pp. 265-271.
Papastergiou, M. (2009) Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation, *Computers & Education,* pp 1-12.
Sanchez, E. (2011a) When Games Meet Learning, *ICT and Informatics in a Globalized World of Education*, pp 9-13.
Sanchez, E. (2011b) *Key Criteria for Game Design. A Framework,* Rapport du MEET project, Europen Commission.
Sanchez, E. (2011c) Usage d'un jeu sérieux dans l'enseignement secondaire. Modélisation comportementale et épistémique de l'apprenant, *Revue d'intelligence artificielle,* 25(2), pp. 203-222.
Sanchez, E. (2013) A model for the Design of Digital Epistemic Games, In *Proceedings of the World Conference on Computers in Education,* pp 257-264.
Senach, B. (1999) *Evaluation ergonomique des interfaces homme-machine: une revue de la littérature,* Rapport de recherche INRIA.
Szilas, N. And Sutter Widmer, D. (2013) L'évaluation rapide de jeux d'apprentissage : la clef de voûte de l'ingénierie ludo-pédagogique (Instructional Game Design), In *Actes de l'Atelier « Serious games, jeux épistémiques numériques » Méthodologies de recherche pour l'étude des interactions*, pp 24-28.
Thomas, P., Labat, J.M., Muratet, M. and Yessad, A. (2012) How to Evaluate Competencies in Game-Based Learning Systems Automatically?, *Intelligent Tutoring Systems,* Springer, pp 168-173.
Tricot, A., Plégat-Soutjis, F., Camps, J-F. , Amiel, A., Lutz, G. and Morcillo, A. (2003) Utilité, Utilisabilité, Acceptabilité: Interpréter les Relations Entre Trois Dimensions de l'Evaluation des EIAH, In Proceedings of *Environnements Informatiques pour l'Apprentissage Humain*, pp 391-402.
Von Wangenheim, C. G., Thiry, M. And Kochanski, D. (2009) Empirical Evaluation of an Educational Game on Software Measurement, *Empirical Software Engineering,*Springer, pp 418-452.
Warren, S., Jones, G. and Lin, L. (2011) Usability and Play Testing, *Serious Educational Game, Springer,* pp 131-146.