

Best Practices for an Effective Design and Evaluation of Serious Games

¹Chiara Eva Catalano, ²Angelo Marco Luccini, ³Michela Mortara

¹ CNR IMATI-Genova (Italy), chiara.catalano@ge.imati.cnr.it

² CEDEP Paris (France), marco.luccini@cedep.fr

³ CNR IMATI-Genova (Italy), michela.mortara@ge.imat.cnr.it

Abstract

There is an increasing awareness about the potential of serious games for education and training in many disciplines. However, research still witnesses a lack of methodologies, guidelines and best practices on how to develop effective serious games and how to integrate them in the actual learning and training processes. The process of integration heavily depends on providing and spreading evidence of the effectiveness of serious games. In this paper we present an overview on the factors that make serious games effective in the perspective of maximizing the learning impact, and discuss the current efforts in evaluating such impact. Such recommendations are the result of an extensive survey of the current proposition of serious games in different application domains.

Keywords: *game design, learning impact, best practice*

1. Introduction

Serious games (SGs), digital games with educational objectives, may be an alternative and effective way to convey new knowledge to people. Such knowledge may refer to several application areas and spans from awareness to formal education, while players/learners cover a very broad category of users, ranging from general public to pupils. Modern theories of effective learning suggest that learning is most effective when it is active, experiential, situated, problem based and provides immediate feedback [1]. Games have the potential to offer learning experiences with these features but pedagogical aspects are often not taken into account when designing serious games. Moreover, there is still lack of methodologies to evaluate the learning impact and then foster a wider adoption of serious games.

This paper aims to provide operative guidelines to maximize the learning impact of serious games and is based on the extensive work done within the GALA project¹ in surveying and analyzing the current proposition of serious games in different application fields, namely Business & Management (B&M), Engineering & Manufacturing (E&M), Health & Fitness (H&F), Security, Safety & Crisis Management (SSC), Humanities & Heritage (H&H) [2], Personal Social Learning & Ethics (PSL&E).

Since 2010 indeed one of the research activities of the project is concerned with understanding the application-specific requirements and the identification of best practices: it is evident that SGs with high learning impact and diffusion, once identified, can be used as reference points, thanks to their successful approaches. A clear map of SGs deployment in the different application fields is being developed: more than 170 SGs in the different sectors identified so far, with continuous updates; at present more than 40 serious games from different sectors have been tested and evaluated, based on literature review and/or expert evaluation (single or focus group) and/or user studies involving collectively more than 400 users.

For each application domain, this extensive survey has highlighted exemplary best practices about how to measure the learning impact and how to design the game as an impact project by adhering to grounded pedagogical principles from the very beginning. We provided in the following a general synthesis of the findings of the specific application areas, after introducing briefly the necessary background.

The paper is organized as follows: first, a brief overview of relevant pedagogical theories is given to introduce the best practices on pedagogically-driven design of SGs. Then, the few available frameworks for SG evaluation are presented and the common ways for evaluating the learning impact are discussed,

¹ Games and Learning Alliance Network of Excellence, FP7 contract n° 258169 www.galanoe.eu

before introducing some open issues on this subject. Finally, we conclude with the main challenges for future research in SGs.

2. Background on pedagogical theories and learning models

Many theories and models about knowledge and learning have inspired educational efforts and instructional design and, therefore, provided the theoretical background of serious games design and evaluation. Among the pedagogical principles and models that are more or less explicitly rooted in the serious games evaluated by each application field, we can mention historically **cognitivism** where learning is seen as “the acquisition or reorganization of the cognitive structures through which humans process and store information” [3].

Bloom’s taxonomy [4] definitely has got a strong cognitive orientation. It was originated by the shared need to provide educators with a common theoretical framework for assessing learners’ achievements which would have helped also in setting up common assessment tools and a language so that assessment results could be effectively compared. The learning objectives fall into three main domains: (1) Cognitive, (2) Affective and (3) Psychomotor. The Cognitive learning domain refers to intellectual capabilities; the Affective learning domain addresses learner’s feelings, motivation and attitudes and their impact on learning; whilst the Psychomotor learning domain focuses on manual and physical skills.

Kirkpatrick’s Four-Level Training Evaluation Model [5] addresses explicitly the effect that a course can have, specifically for training. This model identifies four levels of impact over trainees, namely Reaction (user’s perspective without considering any factor that can have impact on it), Learning (increase of knowledge before / after), Behavior (extent of applied knowledge back on the job) and Results (effect on the business or on the environment). The way to assess whether each level has been reached are also suggested.

With **constructivism** the focus is moved on the individual representation of the world that is formed through own mental schemas and experiences [6]. The key points in constructivism are given by the direct link between learning and experience, by the active role of learners in constructing knowledge for themselves [7] and by the application of knowledge onto realistic problems to solve. The best way of implementing the constructivist approach is to embed the learning content into the context; the learner has to actively deal with it to advance in the game. This is practically common to most of the games reported by the application fields.

Grounded on the previous theories and on constructivism in particular, in 1984 Kolb presented the **experiential theory of learning** as “a perspective on learning that combines experience, perception, cognition and behavior”[8]. According to Kolb, ideas are formed and re-formed through experience. Therefore, learning is a continuous process grounded on experience rather than just an outcome. An excerpt from [9] explains well how such process can be visualized as a 4 phase cycle: “Adults learn through a process that involves a set of sequential steps: (1) obtaining concrete experience; (2) observing and reflecting upon this experience; (3) formulating abstract concepts in response to this reflection and observation; and (4) (actively) experimenting to test the validity of these concepts”. Associated to experiential learning but clearly along the Cognitivist line, Gagné’s work [10] contributed considerably to the development of Instructional Design and of Computer-Based Training). According to Gagné there are five categories of learning: verbal information, intellectual skills, motor skills, cognitive strategies, attitudes; each of them requires a different kind of instruction and of learning experiences. The learning design and methods used depend on the learning objectives and if the latter change also the learning strategy and instruments have to change accordingly, and the learning taxonomies, as like as Bloom’s one, help in defining what are the skills to be learnt and in which sequence [11].

Finally, Nonaka’s knowledge transformation theory [12] that tries to model how the process of knowledge conversion takes place. Its **SECI model** [13] foresees four main phases (from which derives the acronym SECI) across which knowledge is transformed in a learning organisation: *socialization*: knowledge transfer from expert to learner; *externalization*: conversion to an understandable and interpretable form via concepts and models creation; *combination*: new explicit knowledge is combined with other and already existing explicit knowledge; *internalization*: conversion back to individual tacit knowledge: understanding explicit knowledge.

3 Best practices for pedagogically-driven design

As anticipated, the extensive monitoring and analysis of the SG proposition in various domain has let us identify this set of best practices about how to design games with the intent of generating a learning impact. Now we briefly introduce some of the games that exhibit features of interest in these regards.

In the B&M sector, we encounter mainly simulation games, a genre which is well suited to make the player appreciate the consequences of his/her decisions and therefore suits the training of soft skills. In particular, *EagleRacing* [9] is a video-based management simulation enabling players to experience a number of collaboration dilemmas, face complex team decision making tasks as well as better understand the opportunities and limitations of collaboration technologies in organizations. It provides the basis for highly interactive, collaborative learning experiences which can be deployed both online and offline, among diverse and distributed individuals and teams, or even in a blended deployment by running the individual stage online and one or all the team-based ones offline in a face-to-face (F2F) workshop. *WhatADay* [14] is played individually online and is completely virtually facilitated. The game trains management competences along different dimensions, such as (1) allocating the right time and attention to the issues at hand (and well understanding and prioritising the actions to be undertaken); (2) opportunity and problem management; (3) people and motivation management. *Synergy* [15] addresses a complex business situation challenging the change readiness of an organization. The Synergy learning experiences is deployed in a F2F workshop framework. The players have to design collaboratively in their team a strategy and are given the possibility to implement a variety of tactics, that is, actions, in order to convince all the members of the management board. Each action has a direct consequence on the evolution of the game that they can directly observe. The game ends according to the kind and frequency of actions taken during the gaming session. *MetaVals* [16] is played individually and in pairs. The players have to pass a pre-test for assessing their previous knowledge about the finance topic first, then, they can access individually the game that is represented by a multiple choice questionnaire where they are asked to classify six different items or short sentences into different possible answers. After this stage, the collaborative phase starts and the players are invited to play in pairs on the twelve items they have been exposed to. *MetaVals* can be played both in offline F2F workshops or online.

Similarly to B&M, games for E&M aim to train soft skills but in specific environments (e.g. factories) therefore even if the main genre remains simulation, some are set in 3D environments. *TARGET* [17] for instance is based on a coffee machine manufacturer company that has to adopt global sustainable manufacturing strategy. In this scenario a new sustainable manager is hired, who directly reports to the general manager. To develop soft-skills like communication, the game requires for instance reading the expression of a Non-Player Character and positioning the player avatar in the best distance from the NPC to instil trust, thus 3D becomes relevant. *Innov8* [18], the IBM Business Process Management (BPM) 3D simulation game, aims at giving a better understanding of how effective BPM impacts an entire business ecosystem, and demonstrating how process improvements may help build a smarter planet. *Shortfall* [19] is a complex simulation of a supply chain for automobiles. Each competing supply chain has three distinct tiers: materials, parts, and cars. The different tiers compete as teams against each other, but also require each other for success. Each player on a team is given a direct role in the decisions made during his or her turn. Siemens promotes *PlantVille* [20] as an innovative 3D gaming platform that gives players the opportunity and challenge of running a virtual factory, complete with evaluation of key performance indicators, allocation of scarce capital funds, and the ability to improve process efficiency with the purchase and installation of (naturally) more Siemens equipment.

SGs for SS&CM range from interactive videos to 3D, immersive role play games and aim at training different profiles of professionals to react in the right way in complex scenarios. *Afghanistan Pre-Deployment* course [21], consists of three modules organized with a set of slides plus few pages of questions to answer interactively. The presentation is static (photos + text). The modules focus on getting knowledge of the operational framework and the main actors working in Afghanistan; acquiring key aspects of geography, history, economy, cultural issues in various situations, and language basics.. *Boarders Ahoy!* [22], is a user-friendly, immersive, 3D, multiplayer, first-person-perspective role game developed either for PC or for Android, addressed to instruct NATO security teams on how to conduct ship boarding operations (e.g., questioning crews) and search a suspect merchant vessel for illegal goods or weapons of mass destruction. *Sigur* [23] is a 3D role game in scenarios of industrial hazardous environments for professional training, where emphasis is given on the collaboration dimension. Several

scenarios are available (onboard activities, working at heights, lifting operations, first aid after scald burn or electric burn, etc.) and the session includes simultaneous evaluation debriefing. *RescueSim* [24] deals with professional training in rescue applications and addresses players at operational, tactical and resource management levels. It allows emergency crews to experience an incident involving things and/or people, assess the situation and determine the best response strategy, then implement it and observe the consequences of their decisions. Points of interest include high fidelity and high versatility, high flexibility for the instructor before and during use, different deployment styles (one-to-one training, classroom training, or multi-user team training).

In the H&H domain we found interesting applications for virtual and augmented visits as well as 3D adventures in historically recreated scenarios. *Thiatio* [25] recreates a 3D virtual environment where the player acts as a museum curator, who has to arrange an exhibition on several topics collecting paintings from different virtual museums. To fulfil their task, players have to learn iconography, perspective, lighting and shading, the eras of Art, etc. *ICURA* [26][27] is a 3D adventure in which the player learns about Japanese culture and etiquette. It can be used as a tool to gather information about the Japanese language, behavioral rules and culture in a playful way, either for pre-trip planning or for raising interest in another culture. *Signets of Power* [28] is a point-and-click adventure, recreating the power struggle between the church, the nobility and the Crown in Aarhus, Denmark, depicted in 1458. The *Playing History* [29] game series is about experiencing engaging personal stories in the larger world history. The series places the player in historically significant and interesting time periods, where he/she will get the chance to be part of history in the making. The genre is 3D adventure and can be played on the browser. *MediaEvo* [30] is a virtual / augmented treasure hunt through the town of Otranto suitable for heritage awareness and historical reconstruction.

In PSL&E we face many social awareness games aiming at a large-scale impact and focusing on raising empathy and awareness of specific conditions related to others. *Elude* [31] aims to raise awareness for depression, to inform about this dangerous illness and to enhance friends' and relatives' understanding of people suffering from depression about what their loved ones are going through. *Elude* portrays depression metaphorically. The various parts of the game-world represent emotional landscapes that correspond to different moods with the gameplay changing according to mood changes. *FearNot!* [32] is an interactive drama/video game that teaches children strategies to prevent bullying and social exclusion. The software uses innovative psychology inspired character AI. *Elect BiLAT* [33] provides students an immersive and compelling training environment to practice their skills in conducting meetings and negotiations in a specific cultural context. Like *FearNot!* it uses an affective agent architecture, which is used to promote the believability of the NPCs behaviors. *MeTycoon* [34] aims to stimulate young people's thinking about their careers, allowing them to discover jobs that may not have previously thought of as a possible career. Players have the opportunity to explore and discover the ever changing nature of the job market, and realize that career progression does not always occur in a linear manner.

In the next sections, the best practices evidenced by these games are illustrated.

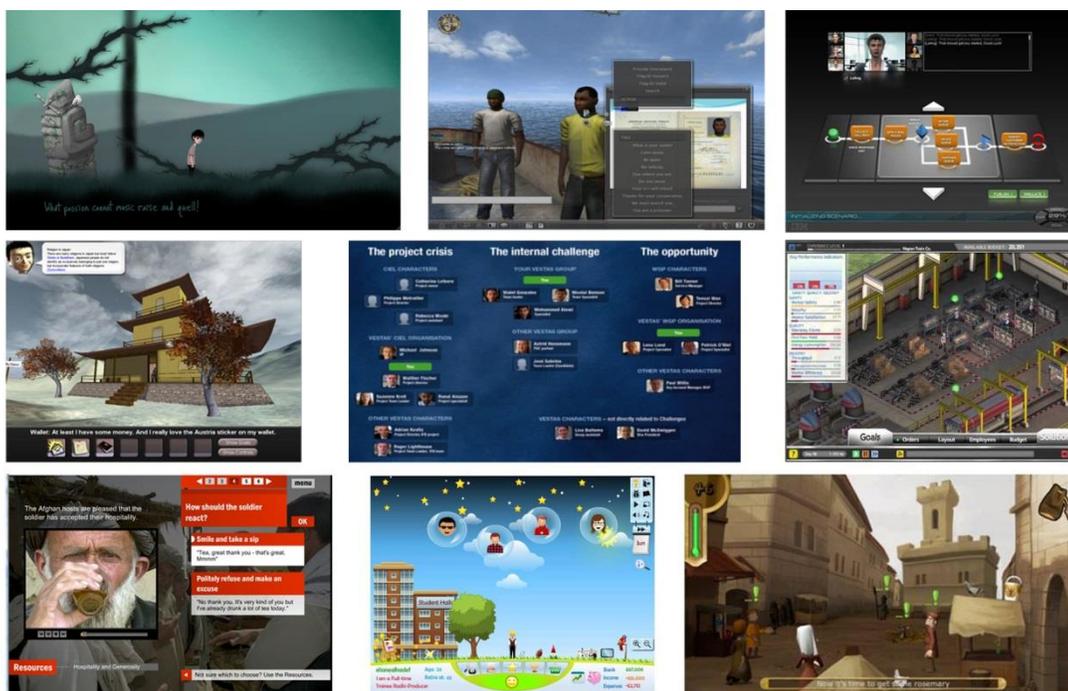


Figure 1. Screenshots from some of the games mentioned in this paper; from top left to bottom right: Elude, Borders Ahoy!, Innov8, Icura, WhatADay!, Plantiville, Afghanistan pre-deployment course, MeTycoon, PlayingHistory-the plague.

3.1. Situate the learning

Situating the learning generally refers to the fact that a suitable deploying environment and proper interactions have to be designed in order to fit the context of use at best. In practice, this may be applied in several ways. For instance, an immersive 3D environment could be powerful and effective only if it serves the learning objectives properly. The positive examples provided by Thiatro, Afghanistan Pre-Deployment course, Borders Ahoy!, Siqur, RescueSim and by TARGET confirm that building a meaningful context help learners a lot and fosters recall and new knowledge application to real life easier [35]. As a counterexample, the use of the 3D environment in online deployment of Innov8 does not really add any value to the game and very doubtfully to the learning itself. The goal of the game is to learn about business process management and there is practically nothing constructive in wandering back and forth through the 3D world.

For what concerns the appropriateness of the interactions among players, they may affect also the choice of the learning style. In EagleRacing it is possible to set different kinds of communication among the players depending on the choice on the deploying style. Real time communication such as verbal interaction happens typically during face-to-face facilitated workshops or even online when web conferencing is chosen for collaborative decision making. However, the latter is not the standard case, since web conferencing is usually adopted for online facilitation (briefing and de-briefing). In online deployment asynchronous communication among players is usually set through emails, blogs and wikis. These interaction modes have been included in the design of the learning experience developed through the game in order to foster group dynamics and use different (and sometimes unusual) ways of collaborating, taking advantage of different time and spatial constraints (i.e. groups and group members typically distributed worldwide: spatial dispersion and different time zones).

Moreover, there exists a trade-off between the high level of contextualization required by corporate organizations that use SGs for the competence development of executive and senior managers and engineers (a common need in B&M and E&M application fields) and the academic natural approach of offering a general solution, good for “most of the cases” without any specificities in particular. Again, this trade-off has to be solved at design level: whenever possible, all relevant stakeholders should

intervene at design level in order to provide the right rate of contextualization so that the resulting outcome fully satisfies the user's needs. Clearly, this approach happens typically in B&M and E&M where the targeted audience is quite specialized and the investment in education/training is a key asset.

Noticeably, a special case of high contextualization is provided by *metaphorical games* such as Elude whose aim is to raise empathy and awareness of specific conditions related to others (depression) without flooding the player with large quantities of information to which (s)he has no relation with. Also the use of Artificial Intelligence in game design is strongly affected by the rate of contextualization vs. generalization as highlighted in games targeting PSL&E. In FearNot! and Elect BiLAT affective agents characterize the believability of Non Playing Characters (NPC) trying to promote cultural adaptability. The same happens for some B&M games such as WhatADay and Synergy where key actors in a corporate organization are modelled in NPCs so that their behaviors deeply contribute to the realism of a crisis or a change management situation the learners have to address, respectively, and for SSC games.

3.2. Minimize the cognitive load

This basic rule follows Rogers' [36] and Gagné's [10][11] experiential learning recommendations for setting ideal learning conditions but has been also confirmed by the analysis of the application domains. In fact, the condition of removing the cognitive load contributes to keep high the level of engagement, of attention and of game playability.

The risk of a steep learning curve is common to all games in all application fields. Either the GUI and navigation modes or the complexity of the learning objectives embedded into the game-plot can represent considerable difficulties that players may meet and that would affect their engagement and motivation. In applications there are good and bad examples: for instance, Shortfall and Thiatro are reported to have a good design since players are introduced gradually to increasingly difficult learning tasks and the game interface is user friendly enough not to make them waste time in understanding how to navigate it. On the contrary, inherently complex game design as in Plantville can have a negative impact on both gaming and learning since the players are lost in trying to understand the goals and how to play. Another unsatisfactory implementation of progressive difficulty embedded into the game is represented by Siquir, a very complex game. Whenever the embedded complexity in game design is relevant, as in B&M, E&M and SSC, facilitation in a structured workshop framework may be necessarily required. For example, in the context of MetaVals the cognitive load related to the use of the interface has been reduced to allow learners to focus on the decision-making process present in each of the phases of the game.

In many cases the need of involving immediately the learner in the game flow emerged clearly, therefore, the actual challenges are how to set up *user-friendly interfaces* so that the player has not to waste time understanding what to do as well as how to set the right pace for progressively introducing the learning content.

3.3. Engage the learner constructively/experientially

This twofold aspect of engagement per se and building learning in a constructivist/experiential way is certainly key in most of the serious games in the application domains. *Learner's engagement* has to be triggered and kept high during the game session (and beyond whenever facilitated workshops are foreseen). Without motivated learners having actual learning is difficult, as reported by Rogers [36] and confirmed by the state of the art: Thiatro and ICURA, RescueSim, EagleRacing, MeTycoon are all games that focus on learners and aim at developing around them rich experiences where they are engaged in building their new knowledge (e.g. by role playing, solving puzzles, achieving a task, coming to a decision collaboratively).

In order to support this, it is good practice to remove any useless redundancy (e.g. avoiding repeating the same walks over and over again as in Signets of Power, ICURA and Innov8, to provide hints at the right time for preventing blocks in the game flow (e.g. as a negative example Siquir in SSC) and to choose the appropriate interface and interaction mode (e.g. reducing text information in Playing History). Instead, the provision of maps that gives the snapshot of current achievements progress could be helpful (e.g. by visualizing Points of Interest as in MediaEvo and the influence trend in the management board in Synergy). Any kind of real-time feedback and of self-evaluation is an important factor that supports engagement and follow the experiential learning best practices. In addition, a variety of scenarios or the

introduction of random elements that may affect the game development should prevent repetitiveness and too deterministic action flows that would help players in predicting and anticipating the flow of events and likely missing part of the learning goals (e.g. Siquir and Boarder Ahoy!).

Again, another important asset for supporting motivation and engagement and fully in line with constructivist and experiential approaches is the *fun component* in game design. This is not mandatory in all application fields (in particular in the B&M, E&M and SSC where the word “game” is not widely and openly accepted and any association to playability and fun often tagged as “not serious” – that is why in those domains often “simulation” replaces the concept of game). However, in H&H and in PSL&E fun could be the key to make a game not only successful in terms of popularity, but also in terms of effectiveness as learning experience.

In B&M, E&M and SSC fun could be a useful add-on too. But what is important in these domains is to guarantee a full engagement, if not even a state of immersion, so that the player is constantly focused on the learning task. However, again in the experiential learning spirit, the goal is not to reach a winning state eventually, but to *learn effectively from a rich and realistic experience and in a safe environment*. Therefore, even the frustration felt by players failing their tasks is part of the learning triggers that is typically used by facilitators during the game session itself (whenever the player realizes that a decision made has brought to negative consequences on the game development), but likely more in the debriefing session just after. RescueSim provides another example where emotional tension helped by the accurate realism keeps high the level of engagement and makes the game stressful and successful.

3.4. Facilitate the learning task

In this case the learning experience is set into a facilitated workshop framework that fits the experiential learning approach and consequently the constructivist one. Indeed,

- (1) the briefing phase has the goal to raise the attention of the learners by exposing them to their learning objectives, game rules and recall of prior knowledge;
- (2) the facilitation during the gaming session builds on the previous stage and helps in providing the actual learning guidance;
- (3) the debriefing phase is designed to consolidate the learning not only via facilitated debating and assessment of the game learning experience but also by transferring it to the ordinary reality.

It is clear that facilitated workshops fit well for application fields as B&M, E&M and SSC where subject domain complexity is high enough to require efforts to support learning and the serious games on their own do not have embedded the necessary mechanisms to support reflection (e.g. in Plantville). However, the actual challenge in workshops is given by the role of the facilitators since, especially in the game session, they have not only to contribute to reduce the cognitive load given by the learning task, but have also to minimize their impact on the flow the players are immersed in. For this reason, it is very important to consider carefully whether the learning objective and the subject domain require facilitation or not at the very beginning of the game design. If it is the case, the relevant stakeholders (e.g. educators, facilitators, domain experts) have to be involved from the beginning of the project and the three stages of the workshop have to be considered as a learning process continuum of activities. This second aspect could be pushed to the limit recommending that the game is not based on any a-priori knowledge, but conversely it should introduce the basics of new knowledge and increment the associated difficulty gradually. In fact, this could bring to the quite defying design challenge of *fully integrating the workshop phases into the game session only*.

3.5. Flexibility, reusability, exploitability

One of the best ways of keeping high the level of engagement is presenting a wide variety of situations and scenarios which may bring value well beyond the original scope of the game. Such flexibility is typical of rich and complex games as in B&M and in SSC where the same game scenario can be used to address different learning objectives (e.g. EagleRacing and WhatADay in B&M) or present multiple scenarios for the same goals (e.g. in RescueSim it is possible to edit the scenarios).

This kind of flexibility offers the possibility of reusing the serious games in different contexts. EagleRacing, Synergy, WhatADay are all good examples of games that have been successfully deployed in different industries. Beyond the quality of the design, most of the merit is likely due to the fact that they address soft skills such as collaboration, change management and crisis management that can be

applied into every business and management sector. Another example of reusability is provided by SSC with the *Boarders Ahoy!* game as it can be successfully deployed in other contexts such as anti-piracy or cultural awareness.

Finally, exploitability is related to the two previous points but also to the concepts of portability and interoperability, web-based deployment, use of apps for mobile deployment, distribution via (open) market places are all good examples of features that make a game successfully exploitable (e.g. *Boarders Ahoy!*).

All features here recapped are strongly depending on the game design. It is possible that a game can be found useful in contexts which were not foreseen originally. However, when this happens, it is probably due to the richness and complexity of the scenarios. Likewise, the technological choices made at design level are preliminary necessary conditions for successful diffusion.

4. Evaluation of the learning impact

As anticipated, despite the growing interest in digital game-based learning and the significant efforts in researching and evaluating SGs, considerable weaknesses remain, including a lack of comprehensive frameworks for comparative evaluation ([37] and references therein). While some game-based learning models have been developed in the literature, they do not specifically tackle *the evaluation of the learning impact produced in the learner by playing (serious) games*.

The TILT, CIAO! and Flashlight frameworks [37][39][40][41] for instance have been designed to evaluate the integration of technology into teaching. More specifically, the Four-Dimensional Framework [42] supports teachers and tutors to evaluate the potential of using a game in their practice (the process of evaluation that a tutor will undertake in advance of game selection and use). Similarly, the Game Object Model [43] is devised to evaluate the use of computer games in the classroom (“educational fitness”), besides supporting the conceptualization and design principles of educational games. Finally, the RETAIN model [44] is conceived as a design tool to be utilized during the creation of an educational game. It provides a set of aspects to be taken into account, as the game characteristics which ideally are responsible of generating a learning impact, but it is not focused on how to evaluate each aspect (a group of experts assigned scores in the examples provided in [44]).

Focusing on the objective estimation of the learning gain produced by a SG, we can just report the current practices reported by the specific projects in the literature and highlight the main challenges towards a more effective methodology for evaluating the learning impact of SGs.

4.1 Current practices in state-of-the-art serious games

All the approaches to evaluation are generally very similar: they usually take care of using Bloom’s taxonomy to position their educational objectives, in particular with respect to the cognitive gain (in PSL&E all three learning categories are considered) and, even if not explicitly, they can be mapped on the Kirkpatrick’s methodology.

Pre/post-game questionnaires and interviews are certainly the most widely used instruments for capturing pre-existing knowledge, on the one hand, and what has been learnt during the game session, on the other hand. Multiple-Choice Questionnaires (MCQ) fit only to assess the lower levels of Bloom’s taxonomy and certainly will not go beyond Level 2 in Kirkpatrick’s model of learning evaluation. In order to go higher in both scales, it is necessary to introduce open questions so that learners can find a less rigid framework to provide their feedback. However, in H&H, SSC and B&M we find as rather common practice just to use only Post-game Questionnaires in debriefing sessions (*Playing History*, *EagleRacing* and *WhatADay*) and therefore addressing only Level 1 in Kirkpatrick’s model and leaving further assessment to organizations and educational institutes.

In E&M making use of Scenario Questions for assessing the right level in Bloom’s taxonomy seems very appropriate. What they propose consists in making the learners read the scenarios before playing the game and inviting them to answer to the same set of questions that will be answered after the game session. Any gap or improvement in the quality of the answers should provide enough information to position correctly the cognitive learning in the taxonomy (but not necessarily only that dimension). In general, questionnaires pertain on the personal enjoyment of the gaming experience (self-assessment at Level 1 in Kirkpatrick’s model) and support the self-assessment on the learning outcomes (Level 2 in Kirkpatrick’s model) which should be better addressed by specific interviews.

Actually, there is the risks of “selective attention” and “confirmation bias” according to which people tend to favor information that confirms their belief or hypotheses [45]. Therefore, it should be recommended to change the order of the questions between pre-test and post-test, or, even better, to disperse the final questions among many other less or no relevant ones at with respect to the learning goals into the pre-game questionnaire. In fact, the player may be affected by patterns in the questions and brought to pay attention only to the few aspects of the game, related to the questions of the pre-game questionnaire.

However, since learning is a process and its experience may be actually revealed only into practice, direct observation has a key role in the evaluation. In the case of game sessions in workshops, the facilitator will be the observer and the behavior and the progress of the player can be monitored in order to see if the learning goals are achieved and when exactly. Noticeably, if the game represents a highly realistic situation, the learning experience could be considered as an actual learning-by-doing effort and therefore, even the direct observation during the game session may be a substantial measure for assessing learning not only at Level 2 but also at Level 3 in Kirkpatrick’s model.

Generally, control groups could be useful at all levels since they can provide an effective way of measuring the learning effectiveness of a serious game by comparing and contrasting the learning gap between the two groups. It is clearly desirable that the group playing the game benefits of the gaming experience and performs better in tests on the job and other assessment when the learning should be already consolidated.

Talking about collected data, all the ones that are related to tests, closed questions and in particular to MCQs are good for deriving statistical analysis findings that could be very useful whenever a game is deployed at large scale. All the other data collected via open questions, interviews and direct observations are good for qualitative evaluations. However, another category of collected data that can be profitably used to assess the level of learning and is suitable for quantitative analysis is given by interaction and activity logs that can track players’ behaviors while playing the game and assess their performance in real time [46]. As a matter of fact, this may be a direct source of information about the level of acquired knowledge during the game session depending on how it is effectively possible to relate an action performed by the player to the achievement of a learning objective. This can be easier if not even trivial when the game is structured in layered levels and access to a level is the sign of an acquired expertise. The task could be tricky in case of games dealing with soft skills and of activities not automatically referring to specific achievements. WhatADay makes use of logs in combination with post-game self-assessment questionnaire in order to provide a comparison and contrast between actual and perceived behavior during the game session.

Last but not least, in B&M a way of evaluating the learning impact of a game over time and which supports perfectly Level 3 in Kirkpatrick’s model can be devised: setting up Post Game Learning Communities (PGLC), that is, aggregating the community of a game around a collaborative tool could to observe and track behavioral changes in learners that have played a game over time. Players are asked to report and share their experiences on the job about the learning addressed in the game session. These kinds of communities respond perfectly to the experiential learning and constructivist principles because of the continuous relationship among the community peers (i.e. the other players) that is a stimulus and an asset for competence development. Facilitators may be part of the community too and keep their role of support and guidance. It is apparent how this instrument is an important asset for organizations since it also represents a learning repository where know-how would be inherited by alumni and preserved within the organization itself. Certainly PGLC suits well in any context where there is facilitation in workshops, but could be likely applied without facilitation in social games and whenever communities of players can be gathered together.

4.2 Challenges

Designing properly the *learning evaluation metrics is an impact project*. Aside the learning objectives of the serious game, it is key to identify as early as possible all the goals that are not domain specific and that may be collectively recapped as intangible values to be captured (e.g. raising awareness, supporting motivation), together with the kind of data to be collected (based on game deployment, on privacy and ethical settings and requirements, and on the availability of data analysis tools / instruments) in order to define what are the suitable learning metrics.

The consequent challenge is designing a *serious game as an integrating part of a learning program*. Actually, if the serious game is an integrating part of a learning program, it acquires a more significant status and value to the eyes of the learners. However, particular care should be posed to the early involvement of all stakeholders not only in the design phase but also in the evaluation one. If the game is facilitated, if the learning objectives are complex and require training of the trainers, if it is possible to bring the serious game from a deeply contextualized setting to a more general one and addressing different domains, it is very likely that not only the maximization of the learning impact will be pursued but also that effective instruments for measuring the actual learning gain will be put in place. Moreover, these conditions are the ideal background for trying to go higher than Level 2 in Kirkpatrick's model, reaching Level 3 quite comfortably, before aiming at further heights.

One of the most criticized but still implemented way for collecting feedback and data is the use of questionnaires. If not carefully designed, questionnaires are ignored or refused any time that a responder is not really obliged to answer them. For instance, it has been reported in business environments that executive managers delegated the task of answering to questionnaires to their assistants, which clearly affects the quality of the evaluation based on the collected data (not exactly what surveyors would expect indeed). This could be one of the reasons why only Post-Game Questionnaires are offered to players of B&M serious games, but it is just a suspect. However, often there are no other alternatives.

Educators and game designers could cooperate to try to find such alternative instruments, by *gamifying the questionnaires*, or by assessing learning achievements and collecting the related data in a different, hopefully totally transparent, way. Even the latter may represent another challenge as the actual risk of interrupting the flow while measuring/assessing could be quite high as it gets along with the risk of increasing the cognitive load of the game. Both consequences might be likely more undesirable than the original issue eventually.

Another partially addressed issue is the *measurement of the learning impact over time*. Learning retention and the ability of *sustaining learning actionability* in the medium and long period should be better investigated. In B&M there are examples of the use of PGLC in order to give a playground where the collective learning experience of the community of players may be collected, stored and shared (therefore, not dispersed). Even in this case, likely more sophisticated techniques than just recording and exchanging among community members their experiences of putting learning into practice should be devised. However, this could be a promising first step and the observation and collection of the first rounds of experiences could suggest to design and embed new ways of assessing and measuring the learning impact.

5. Conclusions

In this paper the best practices derived from the survey of current proposition of serious games coming from many application domains have been listed and grouped according to their pedagogical or assessment-oriented relevance. In all the best practices common traits among the different areas have been presented and discussed.

From the analysis of the best practices, a set of recommendations has been provided which can be considered the assets on which the maximization of the learning impact can be pursued as well as its effective evaluation.

In particular, we believe that the advent of new technologies could bring value and open to new perspectives for both the maximization of the learning impact and for its evaluation. In fact, on the serious game design side, it is reasonable to expect advances for what concerns the game mechanics and the accurate modelling of systems and in particular of NPCs. AI will play an important role and even reshape the way facilitation is conceived so far. Virtual facilitators could become the assistants supporting learning efforts by preserving playability, that is, becoming themselves part of the game. Player's engagement would not be affected but purposely sustained. This way, the concept of facilitated workshop itself would be transformed and the serious game could become, in a future perspective, a self-contained holistic learning experience.

Additional technological contributions can come along with the increased AI. For instance, substantial advancement can be expected also from Semantic Web (WEB3.0) which would allow to support AI itself and create more flexible, adaptable and customizable serious games: games played on the web can benefit of getting the right data and content and can support the player in the learning task

in real time. Serious game design would be certainly affected as it should allow to make the game “porous” towards the web and extremely reactive.

Moreover, the Internet of Things (IoT) is taking place and, together with the success of mobile technology, can really come up on massive scale to what is called *pervasive computing* (WEB4.0). Again, the impact on serious game design will rely on the concept of mobility brought to extreme so that players would be able to play online anytime and anywhere. Any appliance would be able to communicate with the mobile device where the game is running. But this scenario may suggest an even greater challenge: how to keep high the level of engagement and of attention. Certainly, new natural paradigms about human interaction have to be defined, but very likely also of learning experience.

New research should investigate more about how to sustain the learning over time and how to make it more actionable. Social Network Analysis (SNA) [47] and learning analytics can become an important asset in this and not only for social games, but also for all the community of learners that can be gathered around collaborative/social platforms, as proposed with the PGLC. Hopefully, SNA should support both aspects: maximization and evaluation of the learning impact.

Anyway, not all of these scenarios will certainly happen or are even devisable. However, they will represent possible challenges that will be able to influence the way serious games will be designed and the way learners may play and acquire new knowledge.

Acknowledgments

This research was supported by the European Commission in the Seventh Framework Program Network of Excellence “GALA: Games and Learning Alliance” (<http://galanoe.eu>), contract no. 258169, FP7-ICT-2009.4.2.

6. References

- [1] Boyle, E. A., Connolly, T. M. & Hainey, T., “The role of psychology in understanding the impact of computer games”, *Entertainment Computing*, vol. 2, pp. 69-74, 2011. <http://dx.doi.org/10.1016/j.entcom.2010.12.002>
- [2] Mortara, M., Catalano, C. E., Bellotti, F., Fiucci, G., Houry-Panchetti, M. & Petridis, P., “Learning cultural heritage by serious games”, *Journal of Cultural Heritage*, <http://dx.doi.org/10.1016/j.culher.2013.04.004>, in press.
- [3] Good, T. L. & Brophy, J. E., “Educational psychology: A realistic approach”. (4th ed.). White Plains, NY: Longman.
- [4] Bloom, S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R., “Taxonomy of educational objectives: the classification of educational goals”, *Handbook I: Cognitive Domain* New York, Longmans, Green, 1956.
- [5] Kirkpatrick, D. L., “Evaluating Training Programs: The Four Levels”. 2nd Edition, Berrett-Koehler Publishers, Inc, San Francisco, 1998.
- [6] Bartlett, F. C., “Remembering: A Study in Experimental and Social Psychology”. Cambridge University Press.
- [7] Devries, B. & Zan B., “When children make rules”, *Educational Leadership*, 61 (1), pp. 64–67, 2003.
- [8] Kolb, D., “Experiential Learning”, Englewood Cliffs, Prentice Hall, pp.20-38, 1984.
- [9] Angehrn, A. A., & Maxwell, K., “EagleRacing: Addressing Corporate Collaboration Challenges Through an Online Simulation Game”, *Innovate, Journal of Online Education*, Vol. 5, Issue 6, Aug/Sept 2009, <http://www.insead.edu/facultyresearch/research/EagleRacing.cfm>.
- [10] Gagné, R., “The Conditions of Learning”. (4th ed.), New York: Holt, Rinehart & Winston, 1985.
- [11] Gagné, R., Briggs, L., & Wager, W., “Principles of Instructional Design”, (4th ed.), Boston, Wadsworth Publishing, 1992.
- [12] Nonaka, I., “A dynamic theory of organizational knowledge creation”, *Organization Science*, Vol. 5, No. 1, pp. 14-37, 1994. <http://dx.doi.org/10.1287/orsc.5.1.14>

- [13] Nonaka, I, Toyama, R., & Konno, N., "SECI, Ba, and leadership: a unified model of dynamic knowledge creation", *Long Range Planning*, 33, pp. 5- 34, 2000. [http://dx.doi.org/10.1016/S0024-6301\(99\)00115-6](http://dx.doi.org/10.1016/S0024-6301(99)00115-6)
- [14] WhatADay, <http://www.alpha-simulations.com>.
- [15] Synergy, <http://www.calt.insead.edu/eis/Synergy>.
- [16] Popescu, M., Romero, M. & Usart, M., "Using Serious Games in adult education- serious business for serious people - the MetaVals game case study", In Proceedings of ICVL 2012 - 7th International Conference on Virtual Learning, Special Award "ICVL 2012", (<http://metavals.eu>).
- [17] TARGET, EU Collaborative project, FP7, Grant Agreement N° 231717 (<http://www.reachyourtarget.org/>).
- [18] Innov8, <http://www-01.ibm.com/software/solutions/soa/innov8/index.html>.
- [19] Carriere, J.D., "Shortfall: An Educational Game on Environmental Issues in Supply Chain Management", M.S. Thesis, Mechanical & Industrial Engineering, North-eastern University, Boston, MA. 2003.
- [20] Plantville, www.plantville.com/game.
- [21] Tesei, A., Barbieri, A., Roceanu I., & Beligan, D., "Evaluation of "Cultural Awareness – Afghanistan Pre-deployment": a User Study", In Proceedings of the GALA Conference, Paris (France), 23-25 October 2013.
- [22] Boarders Ahoy!, <http://www.act.nato.int/qboarders-ahoyq-wins-peoples-choice-award-at-iitsec-2010>.
- [23] Siquir, <http://www.siquir.com>.
- [24] RescueSim, <http://www.rescuesim.com>.
- [25] Froschauer, J., Merkl, D., Arends, M. & Goldfarb, D., "Art History Concepts at Play with ThIATRO", *ACM Journal on Computing and Cultural Heritage*, Special Issue on Serious Games for Cultural Heritage, Vol. 6, Issue 2, 2013.
- [26] Frochauer, J., Seidel, I., Gartner, M., Berger, H. & Merkl, D., "Design and evaluation of a serious game for immersive cultural training", In Proceeding of the 16th International Conference on Virtual Systems and Multimedia (VSMM 2010), pp. 253– 260, 2010. <http://dx.doi.org/10.1109/VSMM.2010.5665978>
- [27] Mortara, M.M., Catalano, C.E., Fiucci, G. & Derntl, M., "Evaluating the Effectiveness of Serious Games for Cultural Awareness: the Icura User Study", In Proceedings of the GALA Conference, Paris (France), 23-25 October 2013.
- [28] Signets of Power, <http://static.magtenessegl.dk/static/index.html?language=english> .
- [29] Playing History, www.playinghistory.eu.
- [30] MediaEvo, <http://www.mediaevo.unifg.it/>.
- [31] Elude, <http://www.gamesforchange.org/play/elude/>.
- [32] Aylett, R., Vala, M., Sequeira, P. & Paiva, A., "FearNot! an emergent narrative approach to virtual dramas for anti-bullying education", in Proceedings of the International Conference on Virtual Storytelling, Saint Malo, France, 2007. http://dx.doi.org/10.1007/978-3-540-77039-8_19
- [33] Kim, J., Hill, R.W., Durlach, P., Chad Lane, H., Forbell, E., Core, M., Marsella, S., Pynadath, D. & Hart, J., "BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context", *International Journal of Artificial Intelligence in Education*, vol. 19, pp. 289-308, 2009.
- [34] MeTycoon, <http://playgen.com/play/me-tycoon/>.
- [35] Dalgarno, B. & Hedberg, J., "3D learning environments in tertiary education", In Proceedings of the 18th Conference of the Australasian Society for Computers in Learning in Tertiary Education. Ascilite, pp. 253-262, 2001.
- [36] Rogers, C. & Freiberg, H.J., "Freedom to Learn" (3rd ed.), New York: Merrill, 1993.
- [37] Mayer, I., Bekebrede, G., Warmelink, H. & Zhou, Q., "A Brief Methodology for Researching and Evaluating Serious Games and Game-Based Learning", in Connolly, T., Hainey, T., Boyle, E., Baxter, G. & Moreno-Ger, P. (Eds.) *Psychology, Pedagogy, and Assessment in Serious Games* (pp. 357-393). Hershey, PA: Information Science Reference <http://dx.doi.org/10.4018/978-1-4666-4773-2.ch017>

- [38] Oliver, M., "An introduction to the evaluation of learning technology", *Educational Technology & Society*, 3 (4), 20-30.
- [39] Ehrmann, S., "Studying teaching, learning and technology: a tool kit from the Flashlight programme", *Active Learning*, 9, 36-39.
- [40] Jones, A., Scanlon, E., Tosunoglu, C., Ross, S., Butcher, P., Murphy, P. & Greenberg, J., "Evaluating CAL at the Open University: 15 Years On", *Computers in Education*, 26 (1-3), 5-15. [http://dx.doi.org/10.1016/0360-1315\(95\)00064-X](http://dx.doi.org/10.1016/0360-1315(95)00064-X)
- [41] Squires, D. & McDougall, A., "Choosing and using educational software: A teachers' guide", London: The Falmer Press.
- [42] De Freitas, S. & Oliver, M., "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?", *Computers & Education*, 46, 3, 249-264. <http://dx.doi.org/10.1016/j.compedu.2005.11.007>
- [43] Amory, A., "Game object model version II: a theoretical framework for educational game development", *Educational Technology Research and Development*, 2006, Volume 55, Number 1, Page 51. <http://dx.doi.org/10.1007/s11423-006-9001-x>
- [44] Gunter, G. A., Kenny, R. F. & Vick, E. H., "Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games", *Educational Technology Research and Development* December 2008, Volume 56, Issue 5-6, pp 511-537. <http://dx.doi.org/10.1007/s11423-007-9073-2>
- [45] Wason, P. C., "Reasoning about a rule", *Quarterly Journal of Experimental Psychology*, vol. 20, pp. 273-281, 1968. <http://dx.doi.org/10.1080/14640746808400161>
- [46] Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P. & Berta, R., "Assessment in and of Serious Games: An Overview", *Advances in Human-Computer Interaction*, Volume 2013, DOI: <http://dx.doi.org/10.1155/2013/136864>. <http://dx.doi.org/10.1155/2013/136864>
- [47] Carrington, P., Scott, J. & Wasserman, S., "Models and Methods in Social Network Analysis", Cambridge University Press, Cambridge (MA), 2005. <http://dx.doi.org/10.1017/CBO9780511811395>