Entertainment Computing 5 (2014) 313–322

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Application of Learning Analytics in educational videogames ☆

Ángel Serrano-Laguna *, Javier Torrente, Pablo Moreno-Ger, Baltasar Fernández-Manjón

Complutense University of Madrid, C/Profesor José García Santesmases s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 31 May 2013
Revised 19 January 2014
Accepted 6 February 2014
Available online 13 April 2014

Keywords:
Learning Analytics
Educational games
Educational data mining
Assessment
Reporting

ABSTRACT

Assessment of learning contents, learning progress and learning gain is essential in all learning experiences. New technologies promote the use of new types of contents like educational videogames. They are highly interactive compared to more traditional activities and they can be a powerful source of data for all forms of assessment. In this paper, we discuss how to apply Learning Analytics (LA) with assessment purposes, studying how students interact with games. One of the biggest barriers for this approach is the variety of videogames, with many genres and types. This makes it difficult to create a comprehensive LA model for educational games that can be generally applied. In order to maintain manageable costs, we propose a two-step approach to apply LA: we first identify simple generic traces and reports that could be applied to any kind of game, and then build game-specific assessment rules based on combinations of these generic traces. This process aims to achieve a balance between the complexity and reusability of the approach, resulting in more scalable LA models for game-based learning. We also test this approach in two preliminary case studies where we explore the use of these techniques to cover different forms of assessment.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Teachers of all educational levels and areas of knowledge are increasingly using new technologies to improve their teaching practices. Among others, they are starting to use Game-Based Learning (GBL) activities to explore new ways to educate their students. Studies like [1–3] support that videogame features such as high interactivity, engagement and challenge can impact the learning processes positively, both by increasing students' motivation and academic performance. The latest Horizon Reports from 2012 to 2013 [4,5] on emerging educational technologies include GBL as a technology that is almost ready for massive adoption.

In turn, assessment processes are central in educational contexts: instructional designers must assure the validity of their methods, instructors need to track the progress of the students in order to provide support and to measure the acquisition of knowledge or skills for formal grading, usually involving the use of some sort of test [6].

All these forms of assessment play a major role in current research on game-based learning. As an innovative form of content, games should undergo validation processes. Furthermore, it would be necessary to develop new forms of assessment especially devised for videogames, given that they present content in new ways and that much more student interaction data is produced with video games than with less interactive contents. In addition, the engaging interactive nature of videogames leads naturally to authentic learning tasks [7], which suggests that videogames may even be used as an assessment tool that can be better than traditional exercises [8]. However, assessment is not thoroughly contemplated in many GBL initiatives, leading to errors and lack of results. It also difficulties adoption because assessment is key in formal education, and having gaps in this area creates distrust in teachers and policy makers alike [9].

In order to increase the adoption of GBL approaches, it is necessary to create reliable assessment systems for videogames that are easy to use, that facilitate the different forms of assessment (e.g. formative, summative, etc.) and that leverage the interactive features of videogames in a cost-effective way.

Among the different perspectives from which this task can be approached, in this work we focus on the potential (and challenges) of applying the techniques typically used in Learning Analytics [10] and Game Analytics [11,12] in GBL scenarios. Learning Analytics addresses the processing and visualization of data collected from interaction and navigation through educational contents. In some cases, Learning Analytics techniques are used to predict future students' outcomes in different educational goals.

☆ This paper has been recommended for acceptance by Matthias Rauterberg.

* Corresponding author. Tel.: +34 913947623.

E-mail addresses: angel.serrano@fdi.ucm.es (Á. Serrano-Laguna), jtorrente@fdi.ucm.es (J. Torrente), pablom@fdi.ucm.es (P. Moreno-Ger), balta@fdi.ucm.es (B. Fernández-Manjón).

http://dx.doi.org/10.1016/j.entcom.2014.02.003
1875-9521/© 2014 Elsevier B.V. All rights reserved.

For that purpose, Learning Analytics is usually implemented as a combination of several technologies and techniques, like data mining, Web-Analytics [13] and Business Intelligence [14]. In turn, the discipline of Game Analytics spans a set of techniques widely used by the game development industry to better understand how users play their games, find errors and improve the game play experience. For the sake of simplicity, along this paper we will simply use the term "Learning Analytics", while we may be referring to either Learning Analytics, Game Analytics or a combination of both techniques.

However, transferring these techniques to the specific challenges of assessment in GBL scenarios poses unique challenges: the potentially available data for applying Learning Analytics techniques is much bigger because interaction is very intense during a gameplay session, and the existing constructs from game analytics techniques are often game-specific (which reduces scalability) and do not cover the special requirements of educational contexts.

In order to facilitate the convergence, in this work we present a two-step approach to define a scalable Learning Analytics System that can support different forms of assessment in GBL activities. The two steps are driven by the technical requirements of scalability, uniformity and reuse of efforts: we start by defining a small and easily treatable set of generic traces, and then build higher level assessment rules by combining those generic traces, resulting in game-specific traces that do not require a full redesign of the assessment system for each individual game.

The paper is structured as follows: first, we describe the general context of this work, the challenges and the requirements. Second, we define a set of basic universal game-agnostic traces that are easily applied, gathered and processed. Then, we describe the second step, in which game-specific assessment rules can be constructed by combining the basic traces. We then exemplify the use of these steps through two exploratory case studies, in which we studied how the system would support the different types of assessment required. Finally we summarize our conclusions about the advantages and disadvantages of our approach and outline some future lines of work.

2. Games, assessment and Learning Analytics

The term assessment is often used to describe different activities and therefore it is necessary to clarify its meaning along this paper. In the literature, it is common to make the distinction between assessment *of* learning (e.g. for certification) and assessment *for* learning (e.g. to provide support for students during the learning experience) [15]. A third form is typically contemplated, namely assessment *as* learning (e.g. self-evaluation and peer-evaluation), although in this work we have focused on the first two concepts (assessment of/for learning).

In addition, when using innovative forms of content, an additional previous step is required: the assessment of the content artifact itself, in order to find design and implementation issues that may hinder the learning experience. The next subsections briefly discuss the forms of assessment considered in this work, along with their relationship with Learning Analytics.

2.1. Assessment of the game artifact

When new instructional materials are created, the assessment of their appropriateness is important, given that the extra effort required for their adoption must be justified. Indeed, most instructional design approaches are based on the generic ADDIE model [16], where the E stands for Evaluation.

Such evaluations may vary, ranging from basic "student acceptance" evaluations to complex multi-stage evaluation procedures, but as a rule of thumb the more expensive the content is, the more rigorous the evaluation should be. One of the first steps is evaluating the overall suitability of the game for the target audience and context. In this sense, de Freitas and Oliver proposed four distinct dimensions to evaluate the suitability of a specific game design, focusing respectively on *pedagogic considerations*, the *learner specification*, the *context* and the *mode of representation* [17]. However, when the proposed game is a new development, basic software validation techniques are typically used, usually in the form of formative evaluations. In some scenarios with strict formal requirements, further evaluations are typically conducted to validate the appropriateness of the game formally (e.g. by comparing their effect in a randomized trial).

When the instructional material is as complex as an educational game, testing and validating the games before their application is a significant challenge. First, making sure that a game is engaging and fun is an elusive process, since these abstract constructs are difficult to measure, and usually require applying invasive techniques. In addition, the usability of games in general and serious games in particular is complex and time-consuming due to the specific traits of this family of applications [18]. Sim, MacFarlane and Read also explored the challenges associated with understanding when an educational game is successfully performing its function [19], and Eladhari and Olilla have explored the complexity of the problem of evaluating interactive games in general [20].

From this situation, it seems that there is a lack of universal evaluation/validation procedures that can be applied to any game to detect issues and identify solutions. In this context, we believe that it may be possible to leverage the use of Learning Analytics to support the first part of the process (detecting spots where the game misbehaves or where the users get lost) so that game designers can try new solutions to avoid those issues.

2.2. Assessment of learning

Assessment with games has been traditionally dealt with by conducting debriefing sessions or regular written tests after the activity. In this sense, the information generated within the game activity is not always considered for assessment. Sometimes the reason is that the videogame lacks the necessary tools to facilitate collecting and extracting game play data. However, some games include features that can be useful for assessment, like Questions and Answers (Q&A), where the game calculates a final score based on students' answers. More rarely the games implement an ad-hoc assessment system and the results are shown as feedback (during game play or at the end) to the student.

Q&A is the natural translation of traditional exams into a game, yet this mechanism is not fully compatible with the nature of the games. Prompting the student to answer questions very frequently can break the game flow, putting engagement and motivation at risk [21]. This has resulted in a growing search for stealth assessment approaches [22,23] that evaluate the performance of the student transparently.

We expect that the Learning Analytics approach, when applied to serious games, can represent a powerful mechanism to conduct stealth assessment procedures.

2.3. Assessment for learning

Assessment for learning typically focuses on conducting early assessment not with the purpose of grading, but with the purpose of providing adequate instructional scaffolding to the students [24]. Indeed, early formative assessment is a keystone in any educational process that aims to support the students as they learn, and some authors argue that scaffolding and formative assessment are essentially the same thing [25].

In the specific context of serious games, providing this type of scaffolding requires the instructor to understand how students are interacting with the game, in order to identify those students that are either struggling or working incorrect assumptions about the content. When the number of students and the format of the gameplay sessions is adequate, merely observing the students play may be enough for the instructor, who may detect these situations and provide support when required. However, such methods are difficult to apply, and it is more common to resort to posterior debriefing sessions [26] or to using games that offer reports about how each student played the game [27].

The typical use of Web Analytics in general and Learning Analytics in particular is often aligned with the goal of identifying users that are encountering obstacles in their navigation, sometimes even in real time. When we apply the Learning Analytics approach to the field of serious games, we expect to be able to identify gameplay patterns that pinpoint students who are getting stuck within the game, or displaying low performance records. Depending on the context, this identification and the provision of additional support for those students may be performed by inspecting the gameplay traces after the session. And, if the context allows it, real-time identification of students finding difficulties would provide instructors with a powerful tool to improve those gameplay sessions by helping struggling students right when they need it.

3. A "universal" set of traces

The application of Learning Analytics techniques to serious games poses significant technical challenges. There is a great variety of educational videogames including a wide range of game genres and types;in factsome videogames can even be considered as unique pieces that blend art and software, requiring individual treatment. This variety hinders the creation of reusable assessment systems that can be applied to different videogames. However, in all videogames the player interaction and the internal game logic produce simple events that are meaningful in the development of the game flow. These events represent everything that happens in the game and can be potentially used for assessment purposes. In the following subsections we present a set of meaningful interactions, called traces, which can capture common events present in all games (see Fig. 1). These traces will be the basis to build up the assessment system

3.1. Game traces

A first step to assess the educational videogame activity in general and each student in particular, is to identify which students actually played the game and if they finished it. This is especially important if the game is designed for unsupervised use (e.g. to play after class). To cover this aspect we propose three types of traces:

- *Game Start*: This trace is generated when a student begins to play a game. We use this trace to identify the user through credentials given in the tracking system (e.g. a user and a password). If students can play the game several times, the tracking system should generate a session number for the player each time. This trace includes basic information (e.g. initial timestamp) and could also contain data about the execution context. For example, technical data as the operating system or the browser where the game runs, or any other valuable data as the system's localization and language settings.
- *Game End*: This trace is generated when the student finishes the game. This trace could contain additional information about the end reached (e.g., if the game had several endings).
- *Game Quit*: This trace is generated when the student quits the game. This trace can also contain some context that identifies the point where the student was when he quitted the game (e.g. the phase or level, a part of the game state, etc.) and about the completion status. Due to technical restrictions, sometimes games cannot generate this trace (e.g., if the game is running in a browser and the user closes the window, the game cannot send a "game quit" signal to the server). To address this problem (at least partially) another trace containing contextual information could be issued in a fixed timely ratio basis (e.g. every 60 s). The last trace received could be considered as the *game quit* trace.

3.2. Phase changes

Most educational games include some kind of internal narrative structure organized in chapters, missions or phases. These phases usually set out sub-goals that must be accomplished in the game (e.g. collect items in some phase, defeat a final boss). In some games players must fulfill all the secondary goals to accomplish the main goal. Sometimes players must complete these phases in a specific order (like levels), while in others, they can explore them

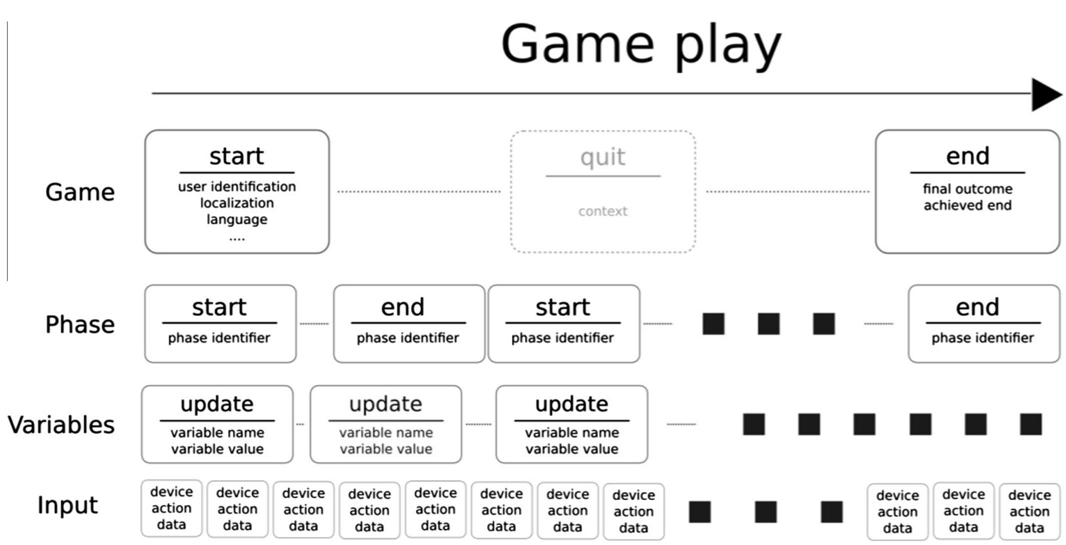

Fig. 1. Diagram with all the traces of our "universal" set, with the data associated to every of them, ordered by frequency during the gameplay.

freely. We propose logging how phases are explored with two types of traces:

- *Phase Start*: This trace is generated when the phase starts. This trace should contain an id identifying the phase.
- *Phase End*: This trace is generated when a player ends a phase. This trace could contain extra information about the completion status, like a "success" or "fail".

3.3. Meaningful variables

All games use variables to keep some sort of game state. Among these variables we can find scores or attempts used to complete the game. Posterior analysis can use final values for relevant variables to define a heuristic for player performance in the game. Variable updates over time represent the player's progress. By monitoring the evolution of a set of relevant variables over time, we can observe intermediate states the player goes through.

This type of trace can play as a wildcard to record various types of traces, as almost everything that happens in a game is registered somehow in a variable.

3.4. Input traces

The previous traces track in-game situations. They contain data related to the game logic and they are "unintentionally" generated by the game engine. But while students play, they produce a lot of data derived from their direct interaction with input devices used to control the game (e.g. mouse, keyboard, controller, joystick, etc.).

Ideally, trackers should be able to collect every single click, key or button pressed by the player. This information can be transmitted (assuming that the game engine is capable of reporting it) to a back-end and stored for latter processing. Grouping all those traces, it would even be possible to reproduce the entire play session afterwards.

However, storing traces for every single interaction can be resource consuming, demanding significant bandwidth and storage space. In most cases, the number of traces tracked can be reduced to a small subset of interactions and that is powerful enough to feed the Learning Analytics System. Therefore, while the proposed framework is designed with the capability of storing every single interaction, we explicitly contemplate the possibility of filtering either on the client side (culling irrelevant information so that it is never transmitted) or on the server side before committing the data to the database.

In any case, all input traces should contain information about the device that generated it (e.g. keyboard, mouse, controller, joystick, etc.), the action performed (e.g. button pressed, moved, clicked, etc.) and the input data associated (e.g. key press code, mouse position, controller button, etc.).

4. Analytics and reporting

The open issue is how to turn simple game and user interaction data into assessment information that could be relevant and useful for the teacher. Using general Web-Analytics techniques on the traces defined in Section 2, analysis tools can create generic reports containing game-agnostic information. For example, using *game traces* it is possible to know how many students played the game, how many finished it, and how many of them quit before finishing. As all traces can contain a timestamp, we can calculate best, average and worst completion time for a group of students. With the *quit game* trace context, we can identify in which points of the game the users stopped playing. This information can help to infer the reason why they stopped playing. When students recurrently quit the game at the same point, the authors may be pointed to a game design problem or a technical error. It could also denote a flaw in the educational content used or poorly balanced difficulty. By identifying these flaws the game can be refined iteratively.

Using *phase traces* it is possible to generate visual reports (e.g. using charts) representing how players access the game structure (e.g., represented in a directed graph), and using timestamps, how the overall time spent is distributed between all the phases. These reports also allow the identification of game design flaws by displaying excessively time-consuming phases, or which educational topics are more cumbersome for the students. If players can explore phases freely, it is also possible to show reports with the completion status of all the phases.

Using *meaningful variables* traces, we can display the final values for some interesting variables. For example, if the game has a score variable, or a variable tracking the final state of a secondary goal, their final values for every player can be shown.

Using *input traces*, we can partially (or completely, depending on the amount of input recorded) recreate the complete gameplay session. In turn, combining these traces with the information recorded by *phase traces,* it is possible to generate reports of how interaction is distributed in every phase. More specifically, we have explored the generation of heat-maps for each scene: this is achieved by aggregating the coordinates where each click of the mouse happens, and visualizing their locations on top of the scene's background. This allows the identification of those points in the screen that receive more attention from the users.

However, although this information provides insight into how students interact with the game and can provide some relevant feedback, it is not enough to support a full assessment process. More specific reporting is required to gain insight into how students are interacting with (and learning) the material.

5. Assessment conditions and goals

In the previous section we described the basic reporting approach based on simple analyses inspired in traditional web-analytics reporting. Those reports can be generated automatically for all games without a significant effort. In turn, educational assessment is typically based on goals achieved by students in certain areas. A goal is considered fulfilled when certain conditions are met, and these conditions are heavily dependent on each specific game. However, the design and collection of specific information traces for each individual game is not desirable, as the results will often be expensive.

In summary, tracing simple data is not enough by itself, and tracing advanced data is game-dependent and therefore more costly. Alternatively, we advocate for building game-specific assessment rules on top of the generic information traces described in the previous section. Our assessment process is based on this idea. As a starting point, using the "universal" set of traces, we define the following types of conditions:

- *Time from A to B is less/more than T*: Since all traces contain a timestamp, it is possible to define conditions were time passed between times A and B is less/more than a value set by the teacher. For example, time spent to complete the game must be less than 10 min; time spent to complete "phase two" must be less than 30 s.
- *Variable X has as final value equals to/less than/greater than Y*: The condition checks the value of some variables at the end of the game. X can be obtained from operating two or more variables, and Y could be a configurable value or other variable. For example, the "number of deaths" value must be less than five; the variable "score" must be greater than one thousand; variable "goal 1 completed" must be true.

- *Trace Z is present*: A specific trace was generated during game play. For example, "game finish".

More complex conditions could be defined by combining these ones with logical operators (or, and, not). Obviously these reporting rules cannot be made generic: the specific locations (*A*, *B*), or values (*T*, *Y*) are closely dependent on each specific game. Moreover, they must be carefully adjusted to match the educational objectives of the games. However, the information sources for these assessment rules are events broadly covered by the "universal" set of traces, and therefore it would be possible to create template rules that can be applied to different games, with only minor adjustments.

Combining these simple rules, it is possible to define complex goals that define the situations that are considered to reflect that the student has achieved the expected educational gain. The educational goals can be individual, or set for the group.

Including some Business Intelligence mechanisms, the teacher could set alarms if some goals are not accomplished after a certain date. Or, if some goals are generally poorly performed, teachers can make interventions to improve the results (e.g. debriefing the students about their insufficient performance, or trying to understand the reasons using the information collected).

6. Technical considerations

The generic traces have been designed with the intention of simplifying their collection from different games and platforms. However, it is obvious that the game platform used for running the games must allow collecting or generating traces as described in this paper.

In order to apply this approach in different games, it would be necessary to extend open-source game engines or to use game engines that are already prepared to track and report this information. In addition, since we expect to collect meaningful variables, the game platform must include an explicit model to represent the definition of a game. This model should be kept in separate files from the code that runs the games in a format that is easy to process by a machine.

Finally, the recorded data logged needs to be stored and collected separately. This will typically require a remote server that receives and collects the traces, storing them in a database, which increases the complexity of deploying the games in many settings (for example, primary school computers are usually behind a firewall/proxy that prevents direct access to external servers.

7. Case studies

In this section, we put into practice all the concepts presented in the previous sections using two different games developed by the e-UCM group, both implementing a tracking system similar to the one described in [28]. In each of the case studies we put into practice different aspects of this work: Case study 1 focused on assessment of the learning artifact and used reports based on both generic traces and very simple aggregated assessment rules. In turn, Case 2 focused primarily on assessment *of* learning (although we also explored assessment *for* learning), using assessment rules built by combining generic traces as described above (see Fig. 2).

7.1. The Big Party

"The Big Party" is an adventure game developed with the eAdventure platform [29]. The main goal of the game is to teach persons with psychical disabilities certain habits and common skills important for every person in daily life, like taking care of one's personal hygiene, clothing, social behavior in professional environments and how to address other people (Fig. 3). The game starts setting the players in their bathroom. They must shower and get dressed to go to a party with people from the office where they work. Once there, they must interact with their co-workers.

7.1.1. Study setup

"The Big Party" was the first game where we applied our Learning Analytics System. The game had been previously developed and tested with a group of users as part of a project for empowerment of disabled persons. However, its specific characteristics made it especially interesting for a deeper study focused on *assessment of the game artifact*. Its target audience represents a challenge for game designers, who may not anticipate which elements of the game may end up being a distraction, or causing users to get lost inside the game.

The tracking system was directly attached to the game, and we used exclusively the universal traces to gather information about how the users interacted with the product. We therefore had two objectives in this case study:

(1) Checking whether it was technically feasible to add the tracking mechanisms to a game that was developed separately.
(2) Testing whether a simple analysis of low-level interactions could be sufficient to identify game design issues and points in which the users were getting lost.

We invited 19 users from the target group, but who had not had any previous experience with the game, and allowed them to play without further directions, while the games generated the complete set of traces and sent it to the server. Then, we performed different simple analyses of the data, and tested possible visualizations that may help the designers learn about in-game issues, as well as to understand how much time users employed to complete the game. The analyses focused solely on assessing the design of the game rather than the actual acquisition of knowledge.

7.1.2. Results

Fig. 4 shows one of the reports generated for this game. To generate the *heatmap*, we collected all the *input traces* containing clicks in that concrete *phase* (i.e. those that happened before the *phase start* trace and the *phase end* trace). Each click is represented with a pale blue dot, and when an area receives more clicks the dots are clustered and we use warmer colors to identify the most clicked areas (with the color red representing the maximum number of clicks).

This report was relatively easy to create, and only used basic traces (input events and the markers for when the phase started and ended). However, it made it very easy to detect that users were mostly trying to interact with the actual active objects in the scene (toothbrush, toothpaste, comb, deodorant and shaving razor).

Interestingly, it also helped to identify minor pitfalls. For example, in the picture there is a mirror. The mirror is not an active part of the scene, and had never received any attention during the development process. However, there is a reflection in the mirror, and the analysis showed that some students actively tried to click on objects reflected on the mirror.

We also generated other reports using the available data. For example, we measured the time elapsed between *game start* and *game end* traces for every player. With this report, we could identify maximum and minimum completion times, as well as which players were faster and which ones were slower. This type of report is extremely simple, but it is still useful for detecting outlier cases that required abnormal amounts of time to complete the game.

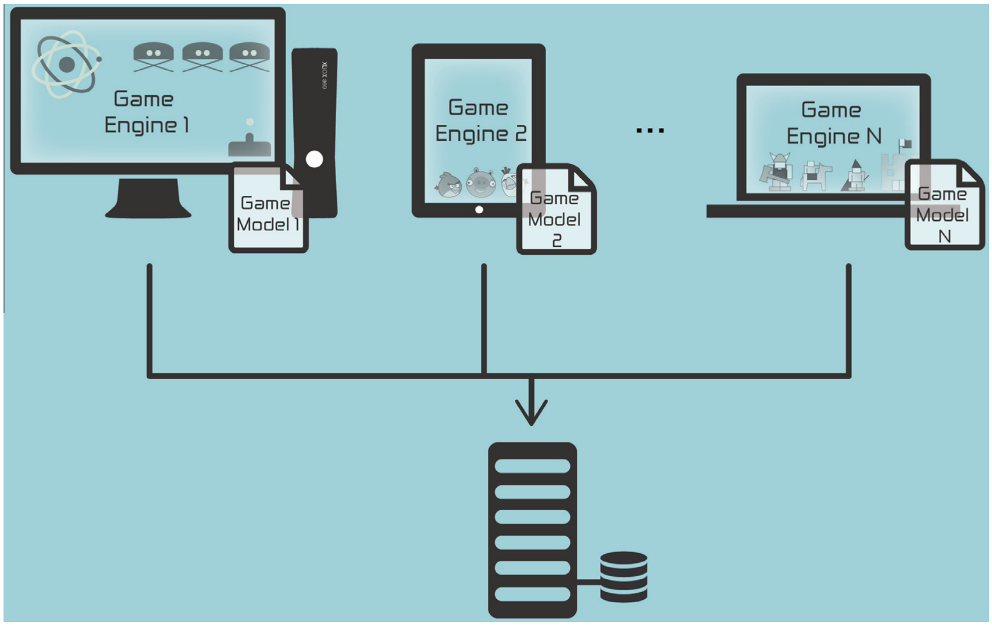

Fig. 2. General architecture.

Fig. 3. Screenshot from The Big Party game.

Fig. 4. A heatmap representing all the clicks performed by players in one scene.

In turn, we also used all the *phase start* and *phase end* traces to analyze how each student explored the game. Using the timestamps from these traces, it was possible to reproduce the path followed by the players. With this report, we could identify how many students were accessing "help phases" (special phases containing clues to complete the game) or getting caught in undesired loops.

Fig. 5 shows a screen capture of this animated report. It recreates how each player moved through all the phases of the game in real time (or as a later replay that can be accelerated). Each colored circle represents a student and his localization inside the game.

7.1.3. Discussion and lessons learned

The main objective of the case study was to test whether simple combinations of the universal traces would be able to provide valuable feedback about game design. The analysis based on the heatmap was especially useful as it provided an intuitive visualization. In fact, it allowed the designers to see how users were trying to interact with the mirror. The developers reacted by blurring a bit the image in the mirror so that it would receive less attention.

On the other hand, the time measurements were useful for identifying outlier figures. Two players took abnormally long to finish the game. While this does not provide directly useful information about the design of the game or the actual stumbling points, it allows the identification of important cases, prompting the designers to take a close look at the full pathway took by those specific students.

For this closer inspection, the directed graph diagram (Fig. 4) that shows the pathway of the students can be especially helpful for identification of undesired cycles or excessive amounts of time spent in a specific phase. As a last resort, having the detailed input traces it would be theoretically possible to replay the entire play session, although this would also require specific modifications of the game engine that we have not explored yet.

7.2. Lost in Space <XML>

Lost in Space <XML> is a puzzle game designed to learn basic XML and DTD concepts (Fig. 6). The game was divided in

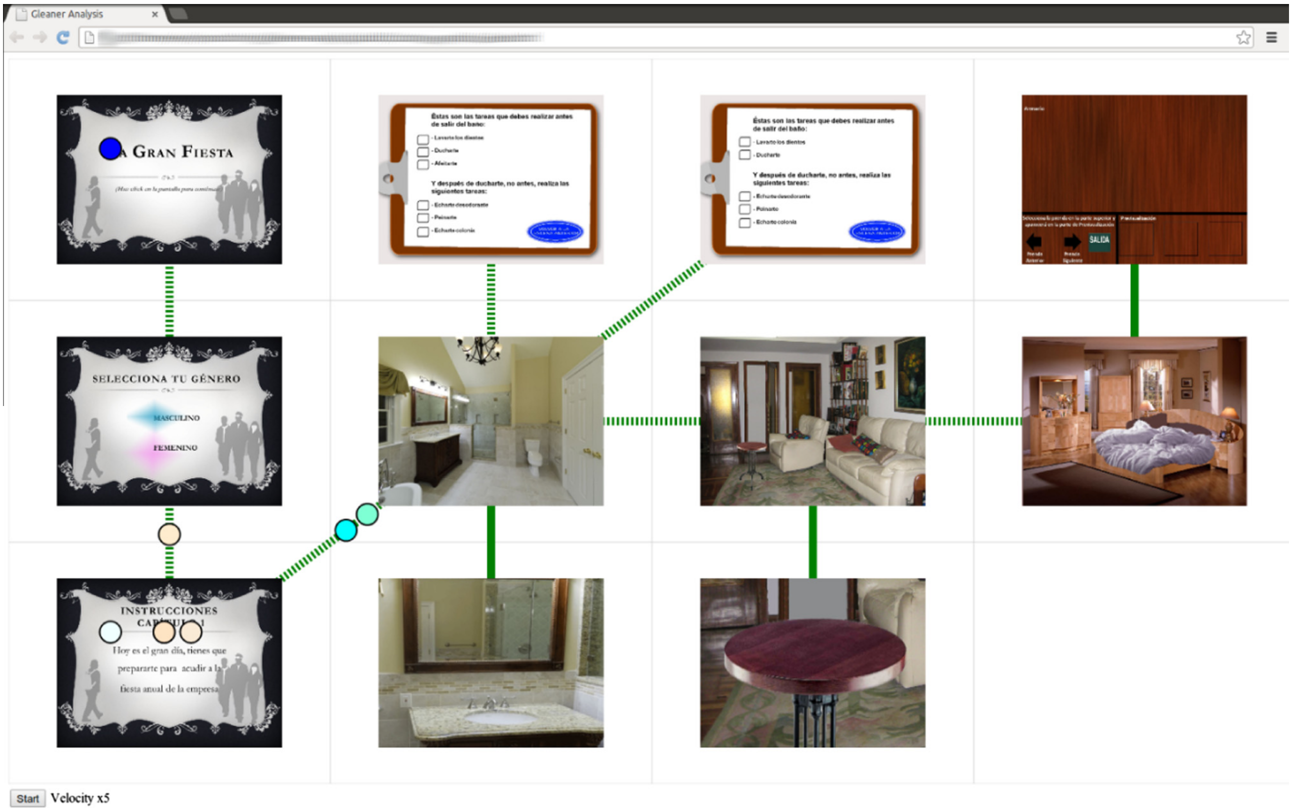

Fig. 5. Graph representing how players explore all the different phases contained by the game. Each circle represents a student and her localization in the game. Circles on edges represent students that are transitioning from one scene to another.

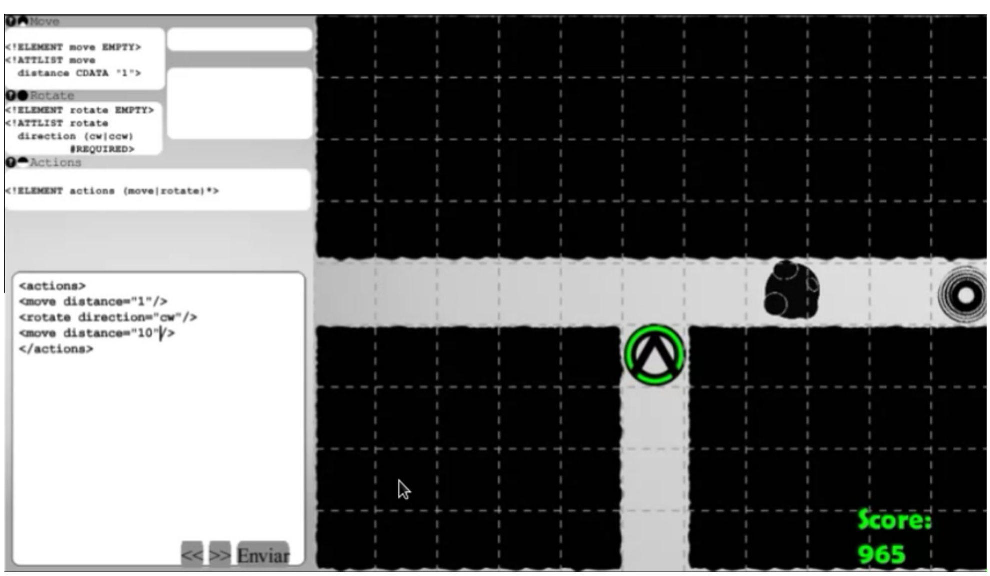

Fig. 6. Lost in Space XML screen capture. Students must enter XML documents in the interpreter (bottom left area) following a DTD (top left area) defining movements necessary to lead the spaceship to the wormhole.

phases of incremental difficulty. In each phase, students must lead a spaceship on a grid board from an origin point to a wormhole. They must type actual XML snippets (complying with a DTD that defines spaceship actions) that are then sent to an interpreter that translates the snippets into in-game actions that guide the spaceship. Some of these XML interactions include advancing one space, rotating, waiting for a few seconds or shooting.

User: 1988 / Session: 1

Goals	Definition	Success
Score	score >= 1000	✓
Valid XML	valid_xml / (valid_xml failed_xml) >= 0.75	✓
Completion time	game_end.timestamp - game_start.timestamp <= 30 min	✗

Fig. 7. Lost in Space <XML> assessment report, showing the goals out comes for one particular player's session.

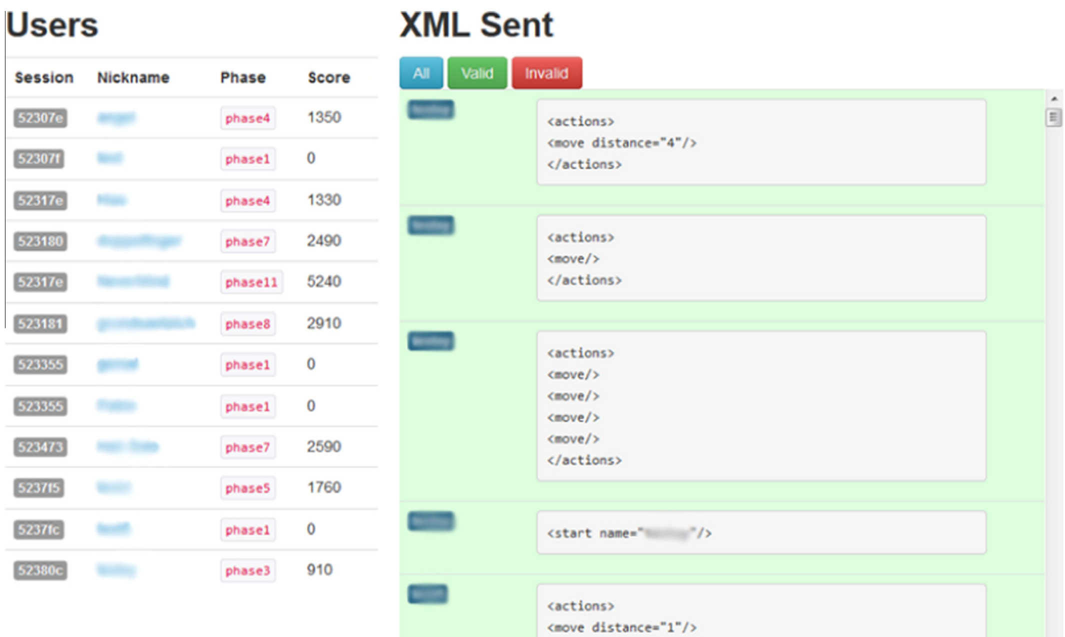

Fig. 8. Real-time dashboard for the Lost in Space <XML> assessment report.

7.2.1. Study setup

While the previous game focused on the assessment of the game artifact, this second case study focused on the other two forms of assessment (assessment of learning and assessment for learning). We wanted students to have to apply frequently the knowledge recently acquired as they played by writing actual XML documents that were the main source for assessment data.

We designed the case study to explore real-time data visualization and to generate complex scores, therefore covering the two forms of assessment mentioned above.

We set three assessment conditions, following the procedures presented in Section 5:

– *Score goal*: This focused on the universal trace based on meaningful variables. In particular, the game reported variations in the variable *score*, and the assessment condition was considered successful if it achieved a value greater than 1000 (maximum score 3000).
– *Valid XML documents*: For each session, the game also exposed a variable with the percentage of correct XML snippets. The assessment rule was activated if the percentage of correct XML documents ended up being higher than 75%.
– *Completion time*: This rule measured the time employed to complete the full game, as time from the first *phase start* to the *phase end* of the last level. The result should be less than 30 min.

Combining these traces, it is possible to create easy to read reports where the instructor can quickly verify individual progress in the game (see Fig. 7).

In addition, in order to test how these traces may be used in real time to assess the progress of each student (and perform corrections if needed), we tested a real-time visualization tool that also reported the current score and latest level successfully completed by each student.

7.2.2. Results

We tested the game with a group of 37 students from a web programming course, who played the game in a computer laboratory while the instructors watched, helped and kept track of the progress through the reporting screen.

The session was an open test of the feasibility of the approach and the usefulness of the reporting tools, and for this reason it was not structured as a formal evaluation. However 94% of the students completed the game successfully (all but two) and 81% of students achieved scores over 1000. In turn, only 24% of the students completed the game in less than 30 min.

However, the real goal of the session was not to obtain assessment of the students, but to try the approach in a real setting. The results in this sense were interesting, with the real-time visualization proving to be a very valuable tool for the instructors conducting the live session (Fig. 8).

7.2.3. Discussion and lessons learned

Through this case study, we tested the possibility of creating more complex assessment tools as combinations of the basic sets of traces. Regarding assessment *of* learning, we established different assessment conditions and goals that could use the universal traces as data sources. The result successfully allowed the instructors to get basic feedback about how well each student had performed during the play session. It should be noted, however, that these scores were exploratory, and they were not considered in computing the official grade of the students.

Regarding assessment *for* learning, the real-time tracking screen was useful for the instructors, displaying the progress and allowing the identification of those students that were having issues. The remediation (scaffolding) action was to approach the students and explain why they were failing to formulate the correct solution.

8. Conclusions and future work

In this paper we have proposed a two-step approach to apply Learning Analytics techniques to support the assessment of educational videogames. The main concern that we try to address is that, while LA techniques seem promising in general and in games in particular, their application to serious games is hard to generalize, expensive to scale and difficult to organize. While typical LA approaches focus on measuring the interactions with a web server (which are often carried out through well-established protocols), game analytics require more complex processes to gather and analyze information [12], potentially resulting in very good results, but also with a high complexity and cost. We are trying to lay the foundations for further work in applying LA in such a way that the technology would not be an issue, by offering a common ground for different vendors to provide and use games with LA subsystems.

The key idea is to propose a first layer consisting of a simple set of generic traces applicable to most (if not all) serious games. This set of traces is intentionally very simple, in order to maximize its potential spread. In turn, an additional assessment layer is used to generate more elaborated reports by tapping into those generic traces to gain insight into how the students are playing and learning. These reports serve three purposes: first, the study of the traces allows the developers to balance the game design, spotting weak points in the game deserving authors' close attention. Secondly, setting assessment conditions and rules allows the definition of success indicators that could eventually be used as a grading mechanism. Finally, the provision of real-time visualizations offers the teachers insight into how each student is progressing, therefore providing an opportunity to intervene.

We have tested this approach and its technical validity through two basic case studies created with different game engines. The first case study (*The Big Party*) focused on game validation and minor adjustments, using a game previously created with the eAdventure game platform. The second case study (*Lost in Space <XML>*) focused on testing the application in terms of assessment *of* learning and assessment *for* learning, using a browser game created with HTML5 tools. While the case studies were not formal validation experiments, they showed the feasibility of the approach. In particular, we aimed to use very simple and direct analyses, demonstrating that even small-scale tracing could yield important insights into the gameplay experience.

Regarding the assessment of the game artifact, we managed to identify design issues with these simple analyses. However, the approach is mostly valid for identifying points in which the players get confused, which does not necessarily give the designers insight on how to solve the issues. However, the identification of the stumbling point is an important and necessary step that facilitates the designers' work.

In turn, regarding the assessment of learning, the traces can be used to create reasonably complex assessment conditions that can be combined in an assessment report. We created such reports in our case study, but our main intention was to test the viability of the approach, not to actually grade the students. Further work on using such games as formal tests would require the validation of the consistency of these scores, for example by correlating them with a previously validated formal test.

Regarding its application as a tracking tool to provide scaffolding, the real-time visualizations offer a promising tool for an instructor overseeing a gameplay session. During the case study, the instructors were actually the authors of this work, and therefore this study is not a validation of whether instructors in general would find it useful for their specific teaching environments. Further work should focus on generalizing the real-time dashboard so that it can be used in different games and validated by external instructors.

Another relevant lesson learned is the difference between the kind of reporting required for fine-tuning a game and identifying pitfalls and the kind of reporting required for assessing the educational gain. Interestingly, fine-tuning can often be performed using the universal set of variables, traces and processing rules, which are independent of the specific game being studied. Therefore, the addition of these evaluation features has a reduced cost and can be easily added to different games. In contrast, assessing the educational game does require specific traces and processing rules, tailored for each specific game, given its specific design and specific learning objectives. Fortunately, we have observed that it is relatively straight-forward to define game-specific rules based on game-agnostic information traces.

We think that this is a first step towards a new model of student assessment based on educational games that can complement other methods. While these case studies are not a formal validation of the approach, our preliminary results suggest a pathway towards affordable assessment systems based on applying Learning Analytics techniques that rely on standard and easily generalizable information traces.

Acknowledgements

The following sponsors have partially supported this work: the Spanish Ministry of Science and Innovation (Grant No. TIN2010-21735-C02-02 and FPU12/04310); the European Commission, through the Lifelong Learning Programme (projects "SEGAN Network of Excellence in Serious Games" – 519332-LLP-1-2011-1-PT-KA3-KA3NW and "CHERMUG" – 519023-LLP-1-2011-1-UK-KA3-KA3MP) and the 7th Framework Programme (project "GALA – Network of Excellence in Serious Games" – FP7-ICT-2009-5-258169); and the Regional Government of Madrid (eMadrid Network – S2009/TIC-1650).

References

[1] H. Tuzun, M. Yilmazsoylu, T. Karakus, Y. Inal, G. Kizilkaya, The effects of computer games on primary school students' achievement and motivation in geography learning, Comput. Educ. 52 (1) (2009) 68–77.
[2] L.A. Annetta, J. Minogue, S.Y. Holmes, M.-T. Cheng, Investigating the impact of video games on high school students' engagement and learning about genetics, Comput. Educ. 53 (1) (2009) 74–85.
[3] H.A. Spires, J.P. Rowe, B.W. Mott, J.C. Lester, Problem solving and game-based learning: effects of middle grade students' hypothesis testing strategies on learning outcomes, J. Educ. Comput. Res. 44 (4) (2011) 453–472.
[4] L. Johnson, S. Adams, M. Cummins, NMC Horizon Report 2012 Higher Education Edition, 2012.
[5] L. Johnson, S. Adams Becker, M. Cummins, V. Estrada, A. Freeman, H. Ludgate, NMC Horizon Report: 2013 Higher Education Edition, Austin, Texas, USA, 2013.
[6] L. Communiqué, Towards the European Higher Education Area: Responding to Challenges in a Globalised World, 2007.
[7] H. Jenkins, E. Klopfer, K. Squire, P. Tan, Entering the education arcade, Comput. Entertain. 1 (1) (2003) 17.
[8] I. Garcia, A. Duran, M. Castro, Comparing the effectiveness of evaluating practical capabilities through hands-on on-line exercises versus conventional methods, in: 2008 38th Annual Frontiers in Education Conference, 2008, pp. F4H–18–F4H–22.
[9] R.T. Hays, The Effectiveness of Instructional Games: A Literature Review and Discussion, Naval Air Warfare Center, Orlando, FL, 2005.
[10] R. Ferguson, Learning analytics: drivers, developments and challenges, Int. J. Technol. Enhanc. Learn. 4 (5/6) (2012) 304.
[11] K. Hullett, N. Nagappan, E. Schuh, J. Hopson, Data analytics for game development, in: Proceeding of the 33rd International Conference on Software Engineering – ICSE '11, 2011, p. 940.
[12] M. Seif El-Nasr, A. Drachen, A. Canossa (Eds.), Game Analytics – Maximizing the Value of Player Data, Springer-Verlag, London, 2013.
[13] J. Burby, S. Atchison, Actionable Web Analytics: Using Data to Make Smart Business Decisions, Wiley, 2007.
[14] S. Williams, N. Williams, The Business Value of Business Intelligence, Intelligence 8 (310) (2003) 30–39.
[15] R. Berry, Assessment for Learning (Google eBook), Hong Kong University Press, 2008. 220.
[16] M. Molenda, In search of the elusive ADDIE model, Perform. Improv. 42 (5) (2003) 34–37.
[17] S. de Freitas, M. Oliver, How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?, Comput Educ. 46 (3) (2006) 249–264.
[18] P. Moreno-Ger, J. Torrente, Y.G. Hsieh, W.T. Lester, Usability testing for serious games: making informed design decisions with user data, Adv. Human Comput. Interact. 2012 (2012) 1–13 (article ID 369637).
[19] G. Sim, S. MacFarlane, J. Read, All work and no play: measuring fun, usability, and learning in software for children, Comput. Educ. 46 (3) (2006) 235–248.
[20] M.P. Eladhari, E.M.I. Ollila, Design for research results: experimental prototyping and play testing, Simul. Gaming 43 (3) (Apr. 2012) 391–412.
[21] J. Chen, Flow in games (and everything else), Commun. ACM 50 (4) (2007) 31.
[22] V.J. Shute, J.M. Spector, SCORM 2.0 White Paper: Stealth Assessment in Virtual Worlds, 2008.

[23] V.J. Shute, Stealth assessment in computer-based games to support learning, in: S. Tobias, J.D. Fletcher (Eds.), Computer Games and Instruction, Information Age Publishers, 2011, pp. 503–523.
[24] R. Sims, G. Dobbs, T. Hand, Enhancing quality in online learning: scaffolding planning and design through proactive evaluation, Distance Educ. 23 (2) (2002) 135–148.
[25] L.A. Shepard, Linking formative assessment to scaffolding, Educ. Leadersh. 63 (3) (2005) 66–70.
[26] V. Peters, G. Vissers, A simple classification model for debriefing simulation games, Simul. Gaming 35 (1) (2004) 70–84.
[27] P. Moreno-Ger, D. Burgos, J.L. Sierra, B. Fernández-Manjón, Educational game design for online education, Comput. Human Behav. 24 (6) (2008) 2530–2540.
[28] A. Serrano, E.J. Marchiori, A. del Blanco, J. Torrente, B. Fernandez-Manjon, A framework to improve evaluation in educational games, in: Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON), 2012, pp. 1–8.
[29] J. Torrente, Á. Del Blanco, E.J. Marchiori, P. Moreno-Ger, B. Fernández-Manjón, e-Adventure: introducing educational games in the learning process, in: IEEE Education Engineering (EDUCON) 2010 Conference, 2010, pp. 1121–1126.