Lessons learned applying Learning Analytics to assess Serious Games

Cristina Alonso-Fernández[a]*, Ana Rus Cano[a], Antonio Calvo-Morata[a], Manuel Freire[a], Iván Martínez-Ortiz[a] and Baltasar Fernández-Manjón[a]

[a]*Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, C/ Profesor José García Santesmases, 9. 28040 Madrid, Spain. E-mail: {calonsofernandez, anarcano, acmorata}@ucm.es, {manuel.freire, imartinez, balta}@fdi.ucm.es*

**Address for correspondence: Complutense University of Madrid, Computer Science Faculty, C/ Profesor José García Santesmases, 9. 28040 Madrid, Spain.*

Email: calonsofernandez@ucm.es.

Lessons learned applying Learning Analytics to assess Serious Games

Serious Games have already proved their advantages in different educational environments. Combining them with Game Learning Analytics can further improve the life-cycle of serious games, by informing decisions that shorten development time and reduce development iterations while improving their impact, therefore fostering their adoption. Game Learning Analytics is an evidence-based methodology based on in-game user interaction data, and can provide insight about the game-based educational experience promoting aspects such as a better assessment of the learning process. In this article, we review our experiences and results applying Game Learning Analytics for serious games in three different scenarios: (1) validating and deploying a game to raise awareness about cyberbullying, (2) validating the design of a game to improve independent living of users with intellectual disabilities and (3) improving the evaluation of a game on first aid techniques. These experiences show different uses of game learning analytics in the context of serious games to improve their design, evaluation and deployment processes. Building up from these experiences, we discuss the results obtained and provide lessons learnt from these different applications, to provide an approach that can be generalized to improve the design and application of a wide range of serious games in different educational settings.

Keywords: learning analytics; game analytics; serious games; game-based learning; evidence-based learning

Subject classification codes: Analysis and Evaluation Methods; Case Studies; Design Experiments; Surveys and Questionnaires; Learning Technologies and Tools

1. Introduction

The gaming industry has experienced a vast growth worldwide in recent years (Entertainment Software Association, 2017). The application of games with a non-entertainment primary purpose, so-called Serious Games (SGs) (Abt, 1970), can provide multiple benefits (Boyle et al., 2016) in environments where games were not traditionally present. Such is the case of education, where non-interactive contents still constitute the majority of learning materials, and there is little consensus about how to best include technology in the classrooms (Adams Becker et al., 2017). However, this slow adoption of learning games in a broad sense, contrasts with the use of games in specific educational domains (e.g. business, military (Kato & Klerk, 2017)) and with the major presence of games in the spare time of students (Pew Research Center, 2018).

New techniques such as Learning Analytics (LA), are trying to provide insight about the educational processes and improve the common educational scenarios benefiting from data-driven approaches. LA, as defined in (Long, Siemens, Gráinne, & Gašević, 2011), aims to measure, collect, analyze and report data from learning tools, such as LMSs (Learning Management Systems) or MOOCs (Massive Open Online Courses), to extract useful information about how students learn with the purpose of understanding and optimizing their learning processes and contexts (Sclater, 2017).

LA techniques can clearly be applied to game environments, where their interactive nature is adequate to the data-capturing process. Data from serious games can therefore be collected while students are playing – both providing information about the impact the game is making (e.g. in their learning, as will be the goal of LA) but also providing information about the appropriateness of the game design and its mechanics (aligning with the Game Analytics field for entertainment games (Seif El-Nasr, Drachen, & Canossa, 2013)). This combination of LA and GA techniques results in Game Learning Analytics (GLA) (Freire et al., 2016) for serious games. Consequently, in-game interactions can be analyzed with many different purposes, in particular: 1) to better understand how students learn using games and 2) to validate the actual educational and game designs. Several studies have been carried out for these two purposes for instance: for purpose 1), authors have been able to assess students based on in-game data (Kiili, Moeller, & Ninaus, 2018) or to predict learning results based on students' interactivity

(Hernández-Lara, Perera-Lluna, & Serradell-López, 2019); while for purpose 2), learning analytics data have been used to validate serious games (Tlili, Essalmi, Jemni, & Kinshuk, 2016) or to find possible improvements in game design (Hicks et al., 2016).

The application of analytics to serious games is not new, however, few studies have reported empirical evidence to inform about the learning process adequately (Chaudy, Connolly, & Hainey, 2014). Despite the increased interest in the application of LA techniques for assessment in serious games, authors have identified the need for more data-based research regarding this topic (Liu, Kang, Liu, Zou, & Hodson, 2017).

This paper contribution aims to fill the gap found in the literature available by providing data-based evidence of the possible applications of GLA data for serious games, giving an example of three specific experiences. Each of these three applications was performed with a different SG, was focused on a different domain and was used in real-world educational scenarios, instead of only relying on experiments in controlled environments. The SG *Conectado* is a tool to be used by the teachers in the classroom to address bullying and cyberbullying (Calvo-Morata, Rotaru, et al., 2018). GLA was used in *Conectado* to improve validation and deployment in schools. The SG *DownTown* is designed for promoting independence in users with intellectual disabilities (Cano, Fernández-Manjón, & García-Tejedor, 2018); and its use of GLA illustrates how to validate a game design in situations where information cannot be directly gathered from the users. Finally, the *First Aid Game* is designed to teach first-aid maneuvers to teenagers, and had already been formally validated (Marchiori et al., 2012). The use of GLA for the *First Aid Game* focuses on improving the evaluation and deployment of games by applying data mining models to predict students' knowledge after playing based on interaction data (Alonso-Fernández, Caballero Roldán, Freire, Martinez-ortiz, &

Fernández-Manjón, 2019), proving that games can help to accurately assess students' knowledge. This can greatly contribute to SG generalization and adoption in schools.

These three games have been chosen to showcase the usefulness of GLA, as they are representatives of different goals: *Conectado* aims to raise awareness; *First Aid Game* aims to improve students' knowledge; and *DownTown* aims to train skills. These case studies also have different target users: *DownTown* focuses on adults with intellectual disabilities, like Down Syndrome or Autistic Spectrum Disorders; while the other two studies focus on young students, although while *First Aid Game* is designed focusing only on the players, *Conectado* is designed as a tool to be used in class supervised by teachers. Due to the different goals, target users and educational scenarios, GLA played a different role during lifecycle of the serious game development: from game validation (*DownTown*), to overcome deployment issues in actual scenarios (*Conectado*) and students' assessment (*First Aid Game*).

This paper summarizes our experiences using GLA to provide teachers and researchers with reliable, evidence-based insights on the accuracy of the game design to the expected learning outcomes, and to facilitate their deployment in actual real educational scenarios. Therefore, we consider that this GLA approach can be generalized to improve the design and application of a wide range of SGs in different educational settings. The paper is structured as follows: Section 2 describes the methodology followed while applying GLA techniques; the three case studies collected are explained in detail in Sections 3, 4 and 5; finally, Section 6 discusses the results and Section 7 summarizes the main conclusions of the different applications of GLA data for Serious Games.

2. Methodology

In this paper, we review three different case studies applying game learning analytics data with serious games. Each of the three games were developed based on specific educational designs, which established the different goals to be achieved with each game. To connect these goals to the actual Learning Analytics data to be analyzed, we followed the Learning Analytics Model (LAM) (Perez-Colado, Alonso-Fernández, Freire-Moran, Martinez-Ortiz, & Fernández-Manjón, 2018). LAMs describe: 1) how the educational design and learning goals are linked with specific game goals and game mechanics and 2) which interaction data should be gathered, how it will be collected, and how to analyze that data to be meaningfully presented to the different stakeholders. Traditionally, collected data is analyzed after the gameplay session is over as a report for the current gameplay session, to generate aggregated reports from several gameplays or to extract complex metrics and relationships (high computational cost). However, we consider SGs as an educational tool that can be used also during the class, so it is required to provide some feedback / insights while games are running using near real-time analysis. In both scenarios, the results can be used to fill dashboards that provide visual feedback (at near real-time or for later analysis) about the performance of the students for the involved stakeholders, such as teachers, students, or managers of educational institutions.

The three games reviewed in this paper had different learning goals, and their development was driven by their LAMs in order to extract the data of interest considering the specific goals and design characteristics of each game. Once GLA data is collected, it can be analyzed to validate or refute the appropriateness of the educational design of each game. For instance, checking that the game mechanics are properly designed for its target users in terms of complexity, duration, number of tasks assigned, etc. Results of the analysis can also be used to validate additional hypothesis established by educators or researchers regarding the expected learning outcomes and abilities of the players while interacting with the game.

The GLA data collected in all three scenarios, specified in their LAMs, followed the xAPI-SG Profile (Serrano-Laguna et al., 2017), a standard collection model for tracking interaction data from Serious Games. This interaction model is implemented in Experience API (xAPI) (ADL, 2012), a data format to track learning activities whose traces have three main fields: the "actor" who makes the action, a "verb" which is the action itself and an "object" which is the target of the action. This tracking model is similar to the model used to track user activity in social networks defining a common set of verbs, activity types and extensions.

All xAPI-SG traces collected (see examples depicted in the figures in Subsections 3.2., 4.2. and 5.2.) use a unique pseudo-anonymous token given to players as name; describe the interactions using a combination of in-game actions and targets, possibly including additional data in the "results" field, and contain a timestamp identifying the time in which the interactions occurred.

Of course, enjoying the advantages of collecting GLA data comes at a price, and in this case a GLA infrastructure is required to be able to collect and analyze the data. In our case, we have reused both the xAPI tracker and the GLA open code infrastructure developed in the H2020 projects RAGE and BEACONING that is available online[1].

Each of the following three sections begins by providing an overview of the goals, target users and design of the SGs used in the respective case study, including their experimental designs, the evaluation methods, the GLA data collected, and finally, the results obtained via analysis.

[1] https://github.com/e-ucm/rage-analytics

3. Case Study: *Conectado*

Conectado is a graphic adventure SG to increase bullying and cyberbullying awareness for students between 12 and 17 years old, which is considered a serious universal problem (Kowalski, Giumetti, Schroeder, & Lattanner, 2014). The game places the player in the role of a student suffering cyberbullying by schoolmates upon arriving in a new school. The aim of the game is to promote empathy with the victims making players experience the feelings that aggression victims usually experience. The game is designed to be played at schools with the supervision of teachers; the game is therefore intended to be completed in 30 to 40 minutes, leaving time for a 15-minute discussion with the teacher afterward, which would fit into a standard 55-minutes high school lecture in Spain.

The plot of the story occurs during 5 days in which the players move from home to school and back home, in the usual places where cyberbullying occurs (Figure 1). Empathy and other emotions are brought to players through mini-games presented as nightmares at the end of each day. Dialogues with other in-game characters also raise players' emotions, as the schoolmates slowly turn against the main character.

Figure 1. Screenshots of *Conectado* depicting the classroom scenario and the home use of the in-game mobile classmates social chat.

The player can choose among several options in different in-game situations. Choices taken alter the story, e.g. the ending is determined by the protagonist's relationship with classmates and parents, and by whether the character has asked the teacher for help or not. No matter the choices taken, a satisfactory ending cannot be reached until the end of the fifth day, so all players go through the complete game experience. Also, physically aggressive answers are excluded from dialogues.

3.1. Experimental Design

Conectado provides a linear flow with some available choices to take in-game that arrive in one of the three possible endings. The game comprises the most common situations, scenarios and roles involved in bullying and cyberbullying, as identified in literature (El Asam & Samara, 2016; Larrañaga, Yubero, Ovejero, & Navarro, 2016; Patchin & Hinduja, 2006). Evaluation of the game is done with pre- and post-tests which assess the level of cyberbullying awareness. The questionnaire used in the pre- and post- tests was adapted based on previous (cyber)bullying measurement questionnaires (Álvarez-García, Núñez Pérez, & Dobarro González, 2013; Ortega-Ruiz, Del Rey, & Casas, 2016). Reliability of questionnaire was verified statistically (Cronbach's Alpha = 0.95). Full questionnaire can be found (in Spanish) as an Appendix in (Calvo Morata, 2017).

The game was validated in a single group pre- and post-test experiment (reasons for this choosing are fully explained in the experiment publication) (Calvo Morata, 2017) with N=257 high-school students aged between 12 and 17 years old from three schools in Spain, in June 2017. The pre- and post-test shared 18 questions to measure bullying and cyberbullying awareness.

3.2. GLA data

GLA data collected included: options taken in some relevant dialogues (e.g. if players have shared their personal password or not, if they have decided to tell parents or teachers about the situation, if they have chosen the most confronting option when meeting the aggressor), changes in the patterns of friendship with the classmates and parents and the general risk value, the specific ending reached (out of the 3 possible), interactions with other classmates and parents, interactions with other game elements (e.g. mobile phone, computer, school bathroom), times in completing each game day and the full game, and whether players have completed the game or not.

Figure 2 depicts an example xAPI-SG trace collected for *Conectado* showing that the player with name “XXXX” has interacted (xAPI-SG verb, depicted in red) with the game object (xAPI-SG activity type, in blue) with identifier “Computer” (orange); extensions (green) represent the day and hour in-game and that the protagonist has mobile messages.


```
{
  "actor" : {
     "name" : "XXXX"
  },
  "verb" : {
     "id" : "http://adlnet.gov/expapi/verbs/interacted"
  },
  "object" : {
     "id" : "http://a2:3000/api/proxy/gleaner/games/<game-id>/<version-id>/Computer",
     "definition" : {
        "type" : "https://w3id.org/xapi/seriousgames/activity-types/game-object",
    },
  "result" : {
      "extensions" : {
        "GameDay" : 1.0,
        "GameHour" : "21:30",
        "MobileMessages" : "True"
    }
  },
  "timestamp" : "2018-05-17T12:04:56.835Z"
}
```


Figure 2. Example collected xAPI-SG trace, describing a player’s interaction with the in-game computer in *Conectado*.

3.3. Results

The increase in cyberbullying awareness was measured with the pre-test and post-test, each containing eighteen 7-point Likert items to validate the game. The average score in the pre-test was 5.72 (SD=1.26), compared to 6.38 (SD=1.11) in the post-test, a statistically-significant effect (paired Wilcoxon test yields p-value < 0.001).

In the optional questions of the post-test, out of the players who answered (73.3%), most of them (85.9%) considered that they had learned something new about bullying or cyberbullying while playing, and 80% of the opinions given about the game were positive. Most students (88%) did not feel represented by any in-game character but some admitted feeling identified with the classmates (9%) or the victim (2%).

The analysis of GLA data showed that the younger the players, the longer the time required to complete the game (for 16 years old, mean completion time was 28 minutes, which increased up to 38 minutes for 12 years old). Also, women took longer to finish the game than men (36 minutes on average for female players, compared to 31 minutes for males). Regarding the three possible endings, data shows that most players who finished arrived at the best ending (74.4%).

3.4. GLA application: Lessons learned

GLA data was used at near real-time to allow teachers to monitor what students were doing while they were playing the game (Calvo-Morata, Alonso-Fernández, Freire, Martínez-Ortiz, & Fernández-Manjón, 2018). This information was provided using a dashboard that comprised several visualizations including the ones depicted in Figure 3: gauge chart showing the average friendship level with in-game characters (a); bar chart showing number of players that were in each day in-game (b); and pie chart showing the players in each possible ending (c).

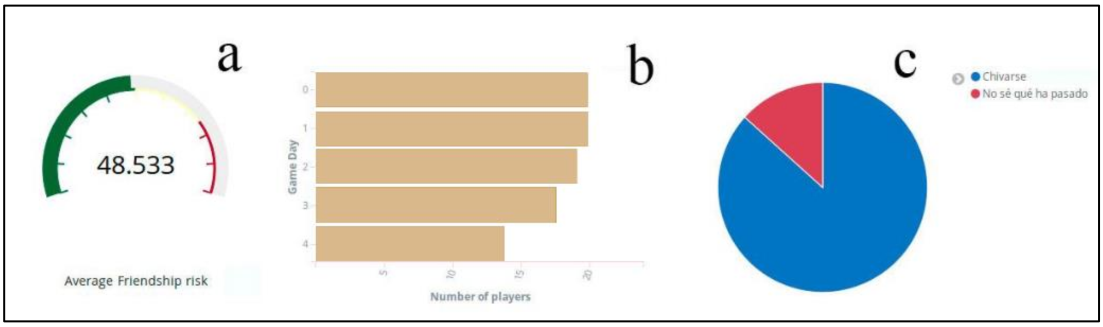

Figure 3. Some of the visualizations shown in the dashboard for teachers while students were playing *Conectado*: average friendship level of players with in-game characters; day in-game players are at; and ending reached by players.

Results of GLA helped to identify some design problems, for instance, the first version of the game (tested in an initial formal evaluation with N=64 students) took too long to complete and left no additional time for discussion with the teacher; this was amended in the second version of the game. All the results reported correspond to this second version of the game. Also, as data was captured at near real-time, some available visualizations allowed to have feedback to control the intervention (e.g. check if data was being received correctly, at which stage of the game players were at every moment). The game was validated as data proved that all target users (despite their different age, gender, school, previous intervention conditions) indeed increased their cyberbullying awareness with the intervention and most of them found the game enjoyable.

This experiment has so far been deployed in 8 schools where some of the interventions where performed by instructions with little support by the authors. Hence, this experiment leverages the advantages of GLA to facilitate the deployment of SGs in classroom settings in a systematic way.

4. Case Study: *DownTown*

DownTown, a Subway Adventure is an espionage-themed game for players from 18 to 45

Figure 4. Screenshots of *DownTown* depicting the user avatar in a Metro wagon and in a hall before using the ticket.

years old with Intellectual Disabilities (ID) such as Down Syndrome, mild cognitive disability or certain types of Autism Spectrum Disorder (ASD). The game aims to train them in using the public subway transportation system of Madrid (Spain), following the positive results of other game-learning experiences with ID users (Kwon & Lee, 2016).

The game was developed in a 3D realistic perspective, so players can identify the in-game scenarios with reality when they travel alone, as depicted in Figure 4. Users can navigate in the game as they do it in real life, so they are trained in choosing the right route travelling from one station to another. Routes are randomly assigned by the game based on the selected difficulty and also can be manually configured. Although specific cognitive skills may vary among individuals, the game design considers the most common cognitive features and barriers of the users. The game includes different levels, so players can progress from easier tasks to more complex ones. Default routes are available, but they can also be manually configured, so players can play the routes that they travel the most in their daily routine.

Downtown also includes puzzles and quests designed to train basic daily skills (e.g. independence, long- and short-term memory, spatial vision) and social aspects that are complex for the target users (e.g. interacting with subway operators) to promote their independent life. Further details about the game design can be found in (Cano, Fernández-Manjón, & García-Tejedor, 2016).

4.1. Experimental Design

DownTown provides different missions in four available difficulty levels. The design considers the cognitive needs of the target users which were especially relevant when it comes to develop the game mechanics and procedures. As feedback could not be directly gathered from target users (e.g. questionnaires), the evaluation was fully based on GLA data, meaning that traced interactions provided all the data used to validate the game design.

The game was tested with N=51 adults with intellectual disabilities (Down Syndrome, mild cognitive Disability or certain types of Autism Spectrum Disorder), aged between 19 and 41 years old, in the Fundación Síndrome de Down in Madrid (Spain), in May-June 2017. Students played a total of 3 hours spread over 3 sessions.

4.2. GLA data

GLA data collected included: avatar configuration and accessibility preferences, attempts to complete each minigame, number of correct and incorrect stations in each route, number of clicks in some interface elements (e.g. accessibility menu, “help” items), progress per timestamp, time spent in each route, time in completing each session or minigame, total game time, inactivity times, and time and number of attempts completing each task after asking for help.

Figure 5 depicts an xAPI-SG trace collected showing that the player with name “XXXX” has progressed (xAPI-SG verb, red) 0.33 (extension, green) in the quest (xAPI-SG activity type, blue) with identifier “Mission_1_Ex_ElMaletinMarron” (orange).

```
{
  "actor" : {
     "name" : "XXXX"
  },
  "verb" : {
     "id" : "http://adlnet.gov/expapi/verbs/progressed"
  },
  "object" : {
     "id" : "http://a2:3000/api/proxy/gleaner/games/<game-id>/<version-id>/Mission_1_Ex_ElMaletinMarron",
     "definition" : {
        "type" : "https://w3id.org/xapi/seriousgames/activity-types/quest",
    }
  },
  "result" : {
     "extensions" : {
        "https://w3id.org/xapi/seriousgames/extensions/progress" : 0.3333333
    }
  },
  "timestamp" : "2018-01-08T18:28:36.211Z"
}
```


Figure 5. Example collected xAPI-SG trace, describing a player's progress in a quest in *DownTown.*

4.3. Results

GLA data showed that most students (85.8%) were able to reach their destination following the right path. Half of the mistakes (50.8%) occurred during the first 30 minutes of playing (once students completed a few routes to understand the mechanics). Beyond making less errors, users also improved their performance in the videogame as they played more sessions reducing their inactivity time (by an average of 58%).

Additional hypotheses were contrasted with data: no significant differences were found between players who customized their avatars (71%) and players who did not (29%) (as suggested by majority of literature (Griebel, 2006; Klimmt, Hefner, Vorderer, Roth, & Blake, 2010; Newman, 2002), although it may be significative that the majority of the users that changed the avatar were Down); players with previous transportation training completed routes faster (taking 6% less time); and regular videogame players completed the tasks quicker (taking 12% less time), showing less inactivity time and making less mistakes that non-players.

4.4. GLA application: Lessons learned

The analysis of the GLA data also helped to validate the game design. From the results, it was drawn that the complexity of the tasks in the different levels was adequately adjusted to the users' intellectual abilities, as there were no peaks in resolution times; that the number of tasks in each level was balanced with its difficulty, as each game level included one more task on average than the level with difficulty immediately below; and that there was a failure in the game design as there was not a correspondence between the number of stations that users navigate in each level and its difficulty (users transited more stations –eleven– on average in the "medium" level than in the "hard" level –ten– , which was not intended in the game design).

Therefore, results of the analysis helped to validate the game design decisions regarding the cognitive skills of the users as well as to identify some problems in the development (that were not found in the beta-testing of the game). Hypotheses that educators had regarding the abilities of the users based on their different IDs and previous experience were contrasted with the data and it has been proven that, after some initial time to become familiar with the game environment, the majority of the users were able to achieve the learning goals of the game (i.e. successfully reach a destination). Next phase of the research will be to compare the behavior of the users that played *DownTown* in the subway versus the ones that were not previously trained with the videogame.

5. Case Study: *First Aid Game*

First Aid Game aims to instruct in cardiopulmonary resuscitation (CPR) maneuvers for students between 12 and 17 years old, following the guidelines defined by the European Resuscitation Council (ERC) (European Resuscitation Council Guidelines Writing Group, 2015). The game-like simulation presents three different scenarios with an in-game character, as depicted in Figure 6, suffering from different emergencies (chest pain, choking and/or unconsciousness). The player then faces a linear situation showing the adequate procedure. The game is designed to have a maximum gameplay duration of 30 minutes and players can repeat each level as many times as they want to.

The specific knowledge to be learned in each of the three situations is assessed with different types of multiple-choice questions that players can retry when choosing a wrong option (Figure 6). The more mistakes the user makes, the less score is given for that situation. After choosing the right option, additional in-game videos show the complete procedures. The player can also interact with some game assets: the main character suffering from the specific situation, a mobile phone to call the emergency services or a semi-automatic external defibrillator.

The game was validated in a usual pre-post experiment with a control group in 2011 with more than 300 students in four secondary schools of Aragon (Spain), as described in (Marchiori et al., 2012). Results showed that, although slightly lower than in the control group (that took part in a theoretical and practical demonstration of the maneuvers by an accredited instructor), the increase in the results from the learning experience was significant in the experimental group which played the game.

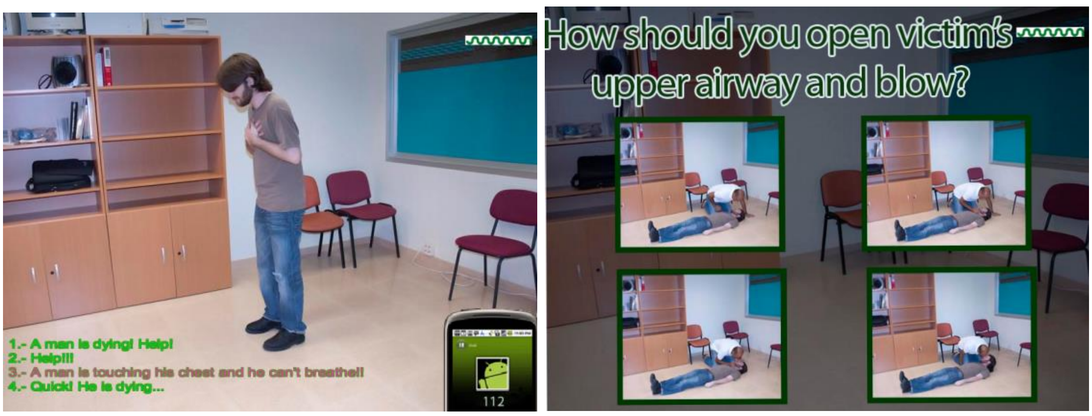

Figure 6. Screenshots of *First Aid Game* depicting a conversation with the user and an in-game selection using pictures.

5.1. Experimental Design

First Aid Game had an evaluation model based on scores given for each of the three in-game situations. The scores decrease based on the errors made in each situation and their relevance. As the game was already validated in a pre-post experiment with a control group, now it is possible to construct prediction models for the post-test results based on GLA interaction data.

The game was tested with N=227 high-school students aged between 12 and 17 years old from one school in Madrid (Spain) in January-February 2017. The pre- and post-test shared a common part with 15 questions to assess students' knowledge on the basic life-support maneuvers covered in the game.

5.2. GLA data

GLA data collected included: whether players have completed the game or not, total score obtained, first and maximum scores obtained in each of the three levels, interactions with game elements, correct and incorrect answers in questions and how many times each level was repeated.

Figure 7 depicts an example xAPI-SG trace collected for *First Aid Game* showing that the player with name "XXXX" has selected (verb, in red) the response "112" (result, green) in question (activity-type, blue) with identifier "NumeroEmergencias" (orange) and, as the result "success" is set to "true" (green), the option selected is the correct one.

5.3. Results

We first checked that learning was still significant in this experiment (i.e. reproducibility of previous results): from pre-test, with mean of 8.06 (SD=2.05), to post-test, with mean of 9.83 (SD=2.38), the paired sample Wilcoxon Signed-Rank test showed a statistically significant increase ($p < .05$). Then, prediction models were carried out to predict post-

```
{
  "actor" : {
     "name" : "XXXX"
  },
  "verb" : {
     "id" : "https://w3id.org/xapi/adb/verbs/selected"
  },
  "object" : {
     "id" : "http://a2:3000/api/proxy/gleaner/games/<game-id>/<version-id>/NumeroEmergencias",
     "definition" : {
        "type" : "http://adlnet.gov/expapi/activities/question",
    }
  },
  "result" : {
      "success" : true,
      "response" : "112"
  },
  "timestamp" : "2017-01-27T03:20:25.571Z"
}
```


Figure 7. Example collected xAPI-SG trace, describing a player's selection in a question in *First Aid Game.*

test scores with two different targets: exact score in scale [0-15] and pass/fail classification (stablishing "pass" in 8 correct answers out of the 15 questions in the post-test). For both targets, some models were developed taking the pre-test and the GLA data as inputs, while other models took only the GLA data, to further avoid the pre-test. Models were compared using 10-fold cross validation.

5.4. GLA application: Lessons learned

For pass/fail predictions, decision trees, logistic regression and Naïve Bayes Classifier were tested. The best model with all previous information was a logistic regression which obtained 89% precision, 98% recall and 10% misclassification rate. Without pre-test information, the best model was another logistic regression with slightly worse results (87% precision, 98% recall and 13% misclassification rate).

For score prediction in scale [0-15], regression trees, linear regression and SVR with non-linear kernels (polynomial, radial basis and sigmoid) were tested. The best prediction model of scores in range [0-15] taking as input the pre-test and the interaction data was a SVR whose mean error was 1.5 (SD=1.3). Without pre-test information, the best model was again a SVR whose mean error increased to 1.6 (SD=1.4).

Additionally, we have determined the features derived from GLA data that are most relevant in the predictions; these include the total number of interactions with the game character and the score in one game level, providing a baseline of in-game actions that can be traced from any SG for assessment purposes. These specific GLA data were related both with the game mechanics and the educational game design of the *First Aid Game*, highlighting that both must be considered to determine which data should be captured from serious games for assessment.

With these preliminary results, we could predict with high accuracy the students' results in the post-test, therefore avoiding the post-test itself. As expected, results are better when simply predicting pass/fail categories than when predicting the exact score, but still good results are obtained for score predictions. Moreover, in some scenarios, it may suffice for teachers to know whether students have acquired enough knowledge to pass or fail the topic. As expected, models taking as inputs both the pre-test and the xAPI-SG GLA data provide better results than models without pre-test information; however, results in models without pre-test are only slightly worse. Therefore, we could also avoid conducting the pre-test and let students simply play the game and, from their interactions, predict their knowledge after playing. Therefore, game deployment in schools is greatly simplified as there is no need of explicit questionnaires to know if students have learned from the game.

6. Discussion

We have summarized three experiences, including their design, analysis and results, with each focusing on a different application of Game Learning Analytics to the corresponding serious game:

(1) Simplify the validation and deployment in schools of a SG that increases bullying and cyberbullying awareness for students (*Conectado*).

(2) Validate the design of a SG that trains students with intellectual disabilities in using the subway without needing explicit feedback (*DownTown*).

(3) Improve the evaluation and deployment of a SG to teach first aid techniques by predicting knowledge after playing to avoid carrying out the post-test (*First Aid Game*).

From the first experience, with *Conectado,* we have been able to validate the game proving that it indeed increases cyberbullying awareness and that it is a helpful tool for students. Data also showed multiple insights into how users played the game, including how they interacted with other characters and the ending reached. Once the game has been validated for its target users using questionnaires and GLA data, its consequent deployment is greatly simplified as it can be used in a larger number of scenarios without the presence of a researcher, without having to further conduct the pre- and post- tests, or even outside of a classroom setting.

From the second experience, with *DownTown,* we have validated the game design in a scenario where obtaining explicit feedback from the users is not always accurate. After users understand how the game works, results show that they effectively improve in completing routes in the game. The game also has a positive effect on players, increasing their motivation and engagement in the learning process. Data also helped to validate the game design and its mechanics and tested whether the specific goals and mechanics are adequately adapted to the users' intellectual characteristics.

From the third experience, with the *First Aid Game*, we have introduced data mining techniques looking for an improvement in the evaluation of players of SGs using the potential of the GLA data collected from in-game interactions. The highly-accurate results obtained suggest that this approach could indeed be applied to avoid the costly pre-post-tests experiments. This may simplify the deployment of SGs and the evaluation of students, as games could simply be played without the additional questionnaires, allowing longer gameplay times, and without requiring the presence of researchers to conduct the tests, as interaction data could be remotely tracked.

In this work, we have revised three experiences we have carried out applying serious games in real settings and applying game learning analytics techniques with different purposes. We consider that our experiences can provide some guidelines for future research on this topic and authors could benefit from some of the lessons learnt:

- **Standardize GLA data collection**: in our three experiences we have used the standard format to collect data from serious games, the xAPI-SG Profile. The use of this standard has simplified collection as we could easily define and match the interactions to be captured for each game with the specific verbs and activity types of the xAPI-SG Profile. The use of a standard has also simplified integration with larger systems - in our case, we have been able to provide real-time information by connecting the data collected from games with the Analytics System. In addition, this standardization would allow real-time analysis easily comparing the interactions of different games (e.g. times of use, learning, completion...). Authors had previously identified the need of data collection standards, to compare studies' outcomes and reuse in-game data (Liu et al., 2017; Smith, Blackmore, & Nesbitt, 2015). We highly encourage researchers to use some standard when collecting game learning analytics data from serious games as it can simplify integration with larger systems and even reusability of data when openly shared for research purposes.

- **Purposes of GLA data**: in our experiences, we have showcased how game learning analytics data can be effectively used for different purposes at different stages of the serious games' lifecycle, and specifically: to validate the game design (*DownTown* case study), to validate and simplify deployment of a game (*Conectado* case study), and to simplify assessment of learners with games (*First Aid Game* case study). Additionally, we have shown examples on how game learning analytics data can provide further information about how students played games or how games promote motivation and engagement. Previous research had been carried out for purposes related to these, for instance: one of the focuses of serious games analytics is to improve game design (Loh, Sheng, & Ifenthaler, 2015); while other works focus on *stealth assessment* (Shute, Ke, & Wang, 2017). The collection of purposes described and exemplified on this work can be used as a baseline for further research.
- **Stakeholders to benefit from GLA use**: It is also important to notice that these purposes cover the interest of different stakeholders: for game designers and developers, to simplify validation of their designs; for teachers and educators, to simplify the application of games in their classes, to obtain real-time information while games are in play, and even to assess their students; for students/learners, to be more effectively assessed based on their in-game interactions and to know their progress and statistics themselves. Previous studies had identified that stakeholders, beyond students and teachers, should be considered for issues related with serious games application for education, as each stakeholder will have their interests and requirements (Jaccard, Hulaas, & Dumont, 2017).

7. Conclusions

On each of the three experiences described on this paper, we have gathered different uses of GLA data for Serious Games: to validate the game design, verifying that design choices were adequate for its goals (e.g. gameplay time, difficulty of levels); to prove that all the game target users can reach the expected outcomes; to test additional hypotheses expected by educators or researchers; to provide visual feedback while games are in play to follow the intervention; or to predict learning results, traditionally measured with questionnaires, simplifying deployment and assessment.

However, these are not the only possible uses of GLA data for Serious Games. At early stages of the design process, interaction data collected with some target users could help to quickly iterate and find problems in early versions of the game; feedback from users at any stage could be collected remotely simplifying early testing or large deployment; and improvements for subsequent versions of an already deployed game could be extracted from players' interaction data.

From our experience, we have also identified that it is essential that games have an underlying learning design that allows for evidence-based assessment. For this purpose, it is convenient that the design of games follows a LAM to clearly stablish its goals, and how they are to be measured with interaction data adequately collected, but this is not enough; from the very beginning, games need to be designed so that data can be extracted from them and provide the information required to validate the games and assess students using them. In that sense, games need to be designed bearing in mind the key data that is to be collected so the desired evaluation has the required input data to be adequately performed. Additionally, it is highly recommended that the collected GLA data follows some standard format.

The positive results obtained in the three experiences reviewed showcase the importance of game learning analytics in different contexts, simplifying serious games' validation and deployment, as well as players' assessment, and even being used as the sole means to obtain players feedback. All these applications intend to promote the use of serious games with different goals in real contexts. We consider that the use of game learning analytics techniques can and should be generalized to improve the design and application of serious games in different educational settings.

Acknowledgments

We would like to thank the anonymous reviewers for their detailed comments and recommendations that have greatly helped us to improve this paper.

This work has been partially funded by Regional Government of Madrid (eMadrid P2018/TCS4307), by the Ministry of Education (TIN2017-89238-R) and by the European Commission (RAGE H2020-ICT-2014-1-644187, BEACONING H2020-ICT-2015-687676, Erasmus+ IMPRESS 2017-1-NL01-KA203-035259) and by the Telefónica-Complutense Chair on Digital Education and Serious Games.

References

Abt, C. C. (1970). *Serious Games*. Viking Press.

Adams Becker, S., Cummins, M., Davis, A., Freeman, A., Hall Giesinger, C., & Ananthanarayanan, V. (2017). *NMC Horizon Report: 2017 Higher Education Edition*. Austin, Texas: The New Media Consortium.

ADL. (2012). Experience API. Retrieved March 20, 2016, from https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/

Alonso-Fernández, C., Caballero Roldán, R., Freire, M., Martinez-ortiz, I., & Fernández-Manjón, B. (2019). Predicting students' knowledge after playing a serious game based on learning analytics data (under review). *IEEE Access*.

Álvarez-García, D., Núñez Pérez, J. C., & Dobarro González, A. (2013). Cuestionarios para evaluar la violencia escolar en Educación Primaria y en Educación Secundaria: CUVE3-EP y CUVE3-ESO. *Apuntes de Psicología*, *31*(2), 191–202.

Retrieved from http://www.apuntesdepsicologia.es/index.php/revista/article/view/322/296

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., … Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2018). Making Understandable Game Learning Analytics for Teachers. In *17th International Conference on Web-based Learning (ICWL 2018)* (pp. 112–121). Springer. https://doi.org/10.1007/978-3-319-96565-9_11

Calvo-Morata, A., Rotaru, D. C., Alonso-Fernandez, C., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2018). Validation of a Cyberbullying Serious Game Using Game Analytics. *IEEE Transactions on Learning Technologies*, 1–1. https://doi.org/10.1109/TLT.2018.2879354

Calvo Morata, A. (2017). *Videojuegos Como Herramienta Educativa En La Escuela: Concienciando Sobre El Ciberbullying (Master Thesis)*. Complutense University of Madrid.

Cano, A. R., Fernández-Manjón, B., & García-Tejedor, Á. J. (2016). Downtown , A Subway Adventure : Using Learning Analytics to Improve the Development of a Learning Game for People with Intelectual Disabilities. In *ICALT 2016 - 16th IEEE International Conference on Advanced Learning Technologies*. https://doi.org/10.1109/ICALT.2016.46

Cano, A. R., Fernández-Manjón, B., & García-Tejedor, Á. J. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology*, *49*(4), 659–672. https://doi.org/10.1111/bjet.12632

Chaudy, Y., Connolly, T., & Hainey, T. (2014). Learning Analytics in Serious Games : a Review of the Literature. *Ecaet 2014*, (March 2016).

El Asam, A., & Samara, M. (2016). Cyberbullying and the law: A review of psychological and legal challenges. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2016.08.012

Entertainment Software Association. (2017). 2017 Essential Facts About the Computer and Video Game Industry. *Entertainment Software Assotiation*, *4*(1), 1–20. Retrieved from http://www.theesa.com/wpcontent/uploads/2017/09/EF2017_Design_FinalDigital.pdf%0Ahttp://www.theesa.com/facts/pdfs/ESA_EF_2008.pdf

European Resuscitation Council Guidelines Writing Group. (2015). European Resuscitation Council Guidelines for resuscitation. Retrieved March 6, 2017, from http://ercguidelines.elsevierresource.com/european-resuscitation-council-guidelines-resuscitation-2015-section-1-executive-summary/fulltext

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game Learning Analytics: Learning Analytics for Serious Games. In *Learning, Design, and Technology* (pp. 1–29). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_21-1

Griebel, T. (2006). Self-portrayal in a simulated life: Projecting personality and values in the sims 2. *Game Studies*, *6*(1).

Hernández-Lara, A. B., Perera-Lluna, A., & Serradell-López, E. (2019). Applying learning analytics to students' interaction in business simulation games. The usefulness of learning analytics to know what students really learn. *Computers in Human Behavior*, *92*, 600–612. https://doi.org/10.1016/j.chb.2018.03.001

Hicks, D., Eagle, M., Rowe, E., Asbell-Clarke, J., Edwards, T., & Barnes, T. (2016). Using game analytics to evaluate puzzle design and level progression in a serious game. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16* (pp. 440–448). New York, New York, USA: ACM Press. https://doi.org/10.1145/2883851.2883953

Jaccard, D., Hulaas, J., & Dumont, A. (2017). *Using Comparative Behavior Analysis to Improve the Impact of Serious Games on Students' Learning Experience*. (J. Dias, P. A. Santos, & R. C. Veltkamp, Eds.) (Vol. 10653). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-71940-5

Kato, P. M., & Klerk, S. De. (2017). Serious Games for Assessment: Welcome to the Jungle. *Journal of Applied Testing Technology*, *18*, 1–6.

Kiili, K., Moeller, K., & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training - In-game metrics as learning indicators. *Computers & Education*, *120*, 13–28. https://doi.org/10.1016/j.compedu.2018.01.012

Klimmt, C., Hefner, D., Vorderer, P., Roth, C., & Blake, C. (2010). Identification with video game characters as automatic shift of self-perceptions. *Media Psychology*, *13*(4), 323–338. https://doi.org/10.1080/15213269.2010.524911

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014).

Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073–1137. https://doi.org/10.1037/a0035618

Kwon, J., & Lee, Y. (2016). Serious games for the job training of persons with developmental disabilities. *Computers & Education*, *95*, 328–339. https://doi.org/10.1016/j.compedu.2016.02.001

Larrañaga, E., Yubero, S., Ovejero, A., & Navarro, R. (2016). Loneliness, parent-child communication and cyberbullying victimization among Spanish youths. *Computers in Human Behavior*, *65*, 1–8. https://doi.org/10.1016/j.chb.2016.08.015

Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning Analytics as an Assessment Tool in Serious Games: A Review of Literature. In *Serious Games and Edutainment Applications* (pp. 537–563). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-51645-5_24

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious Games Analytics: Theoretical Framework. In *Serious Games Analytics* (pp. 3–29). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4_1

Long, P., Siemens, G., Gráinne, C., & Gašević, D. (2011). LAK ’11 : proceedings of the 1st International Conference on Learning Analytics and Knowledge, February 27 - March 1, 2011, Banff, Alberta, Canada. In *1st International Conference on Learning Analytics and Knowledge* (p. 195). Retrieved from https://dl.acm.org/citation.cfm?id=2090116

Marchiori, E. J., Ferrer, G., Fernandez-Manjon, B., Povar-Marco, J., Suberviola, J. F., & Gimenez-Valverde, A. (2012). Video-game instruction in basic life support maneuvers. *Emergencias*, *24*(6), 433–437.

Newman, J. (2002). The myth of the ergodic videogame. *New Media & Society*, *4*(3), 405–422. Retrieved from http://www.gamestudies.org/0102/newman/

Ortega-Ruiz, R., Del Rey, R., & Casas, J. A. (2016). Evaluar el bullying y el cyberbullying validación española del EBIP-Q y del ECIP-Q. *Psicologia Educativa*, *22*(1), 71–79. https://doi.org/10.1016/j.pse.2016.01.004

Patchin, J. W., & Hinduja, S. (2006). Bullies Move Beyond the Schoolyard: A Preliminary Look at Cyberbullying. *Youth Violence and Juvenile Justice*, *4*(2), 148–169. https://doi.org/10.1177/1541204006286288

Perez-Colado, I. J., Alonso-Fernández, C., Freire-Moran, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2018). Game Learning Analytics is not informagic! In

IEEE Global Engineering Education Conference (EDUCON).

Pew Research Center. (2018). *Teens, Social Media & Technology 2018*.

Sclater, N. (2017). *Learning analytics explained*. (Routledge, Ed.). New York and London: Taylor & Francis Group.

Seif El-Nasr, M., Drachen, A., & Canossa, A. (2013). *Game Analytics*. (M. Seif El-Nasr, A. Drachen, & A. Canossa, Eds.). London: Springer London. https://doi.org/10.1007/978-1-4471-4769-5

Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, *50*, 116–123. https://doi.org/10.1016/j.csi.2016.09.014

Shute, V., Ke, F., & Wang, L. (2017). Assessment and Adaptation in Games. In *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 59–78). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_4

Smith, S. P., Blackmore, K., & Nesbitt, K. (2015). A Meta-Analysis of Data Collection in Serious Games Research. In *Serious Games Analytics* (pp. 31–55). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4_2

Tlili, A., Essalmi, F., Jemni, M., & Kinshuk. (2016). An educational game for teaching computer architecture: Evaluation using learning analytics. *2015 5th International Conference on Information and Communication Technology and Accessibility, ICTA 2015*. https://doi.org/10.1109/ICTA.2015.7426881