Using Simva to evaluate serious games and collect game learning analytics data

Cristina Alonso-Fernández1, Iván J. Pérez-Colado1, Antonio Calvo-Morata1, Manuel Freire1, Iván Martínez-Ortiz1 and Baltasar Fernández-Manjón1

1 Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, C/ Profesor José García Santestmases, 9, 28040 Madrid, Spain

Abstract. The evaluation of serious games and the assessment of their players is commonly done with pre-post experiments: a questionnaire before the application and another questionnaire after the application to then compare their results. The tool Simva was designed to simplify the complexity of these experiments, reducing times in preparation and deployment and linking all the information gathered anonymously from each specific user. This information includes game learning analytics data which can provide further insight about players’ progress and results. On this paper, we present three experiences conducted in real settings using the different features of Simva to validate three serious games using pre-post experiments and collect game learning analytics data of players’ in-game interactions. We conclude by summarizing the lessons learned from these experiences that could be used for further research on serious games evaluation and assessment of students playing.

Keywords: Serious Games, Learning Analytics, Evaluation, Assessment.

1 Introduction

Serious games are applied for multiple purposes, including: increase knowledge, raise awareness, change attitude or behaviors. To ensure that these types of games fulfil their intended purposes, they first need to be formally evaluated. Their formal evaluation will ensure that games indeed are useful for their purposes and that players change because of their playing experience [1].

Once games have successfully undergone the evaluation process, it is commonly required to have some way to measure how much effect they have had on the players that have used them. That is, we want to be able to measure how much these serious games have increased players’ knowledge, awareness, etc.

Pre-post experiments are the common method to evaluate serious games and assess students who play them [2]. These experiments comprise three steps:

1. A pre-test: an initial questionnaire which assesses students’ characteristic (e.g. knowledge) before the intervention. It can be paper-based or computer-based and it
should be a valid measure of the characteristics the game aims to change. Therefore, the questionnaire must also be formally validated.

2. The intervention: the activity that is intended to change the students’ characteristics. In the case of serious games, it will be the gameplay itself, usually from beginning to end. There should be no time elapsed neither between the pre-test and the intervention, nor between the intervention and the post-test.

3. A post-test: a post-game questionnaire which assesses students’ characteristics (e.g. knowledge) after the intervention. This questionnaire will be handled and has the same requirements as the pre-test. Additionally, it may include optional questions about the experience etc., but at least should include the same questionnaire to measure the characteristics.

The change on the students’ characteristics is then measured by comparing the pre-test and the post-test results. As the only intervention between the pre-test and the post-test is the gameplay, if there is a significant change (usually we will be looking for an increase in knowledge, awareness) in the characteristic measured by these questionnaires, it is proven that the intervention successfully changes that characteristic. With this methodology, the serious game is formally evaluated. Once this process is completed, we can move to the real deployment of the game when it can be used in real settings as it is proven that it is useful for its intended goals.

On the deployment phase of games, the educators, managers or researchers applying games will typically want to know how much effect the game is having on their players. For this purpose, it is possible and common to make use of the pre-post experiments. These questionnaires will provide a measure of the characteristic before the intervention (pre-test) and after the intervention (post-test). By comparing those measures, we could say how much effect the game has on each user. For instance, for a game aim to make players learn something, the pre-test will tell us how much players know about the topic before playing, the post-test will tell how much they know about the topic after playing, and the comparison of both measures will say how much players have learned with the game.

Besides the external measures provided by the questionnaires, another option to effectively measure the changes on students’ characteristics or to obtain insight of students’ gameplays on serious games is by analyzing their in-game interactions. In the field of Game Analytics for entertainment games, information has been collected for players’ interactions transparently (in a process called tracking) for many years, primarily with rentability purposes [3]. For serious games, the combination of these Game Analytics techniques with the purposes of Learning Analytics (applied in all kind of learning environments) provides the so-called Game Learning Analytics (GLA) [4]. The GLA data collected from serious games can provide information from the in-game interactions both from an educational perspective and a gaming perspective. That is, it can provide information both to evaluate and improve the game itself, but also to get insight about students’ progress, results and even to assess them. The information captured from players’ interactions therefore can provide a rich insight for a wide set of stakeholders (teachers, managers, educational authorities, researchers, students) and for
a variety of purposes (validate game design, assess students, improve the game, display real-time feedback, provide overview metrics).

The rest of this paper is structured as follows: Section 2 provides an overview of the tool Simva to simplify scientific validation of serious games. Sections 3, 4 and 5 detail three real scenarios in which we have used Simva to carry out pre-post experiments to validate games, also capturing GLA data from players´ interactions. Section 6 discusses the lessons learned from these experiences. Finally, Section 7 summarizes the conclusions of our work.

2 Simva

Simva is a tool to simplify the scientific validation of serious games. It manages all the items required for conducting pre-post experiments: questionnaires, classes of students, and GLA interaction data. It additionally deals with other required issues such as privacy and anonymity. Simva was built on top of LimeSurvey, a software that manages questionnaires. The creation, edition and all management of questionnaires is dealt with this connection to LimeSurvey. Simva additionally links these questionnaires to be used in the experiments with the students that are going to complete them. In Simva, classes can be created as groups of students to undergo a serious games validation. For each of the students created, Simva provides anonymous 4-letter identifiers to be used as their username instead of any personal identifier that can go against privacy requirements.

![Image](image-url)

Fig. 1. Simva tool provides: classes with students’ questionnaires and interaction traces collected (top figure) and list to be printed of anonymous tokens for students (bottom figure). Pictures retrieved from [5].
Tokens are available to download to be printed and cut to be handed for students (see Fig. 1, bottom image). This pseudo-anonymization technique allows teachers to track each student’s learning process (relating them via the token) while ensuring privacy requirements and regulations (e.g. GDPR). This technique also allows to carry out recall experiments and longitudinal studies.

Serious games are then configured to access the specific questionnaires created in Simva. A configuration file provides the information about the questionnaires (pre-test and post-test) that the game should access. These questionnaires are linked in Simva to one or more classes that are to use those questionnaires. Each class will then have all the information about the students’ identifiers that belong to that class (and therefore can access and play the game) and the questionnaires that are to be completed by them, before and after the gameplay. All the information collected from each student is then linked together by their anonymous identifier, including: pre-test, post-test and game interaction data (see Fig. 1, top image). All the information is then available for researchers or managers who have access to Simva to download.

An overview of the Simva architecture can be seen in Fig. 2. As explained before, Simva has been designed to transparently manage external systems as LimeSurvey and the analytics framework. Externally, the teacher can create classes and surveys, and create assignments between them, allowing a class to participate in a survey.

Classes are designed to unify student management. In the moment of creation, the teacher sets the number of students, and that amount of four letters anonymous identifiers (4-letter random tokens such as FJCD or PWNB) are created for students to have

![Fig. 2. Overview of the Simva architecture managing surveys, classes, and the Analytics Framework. Teachers can access surveys and classes in Simva; players complete a pre-survey, a gameplay and a post-survey. All that information is sent to Simva.](image-url)
access both to surveys and the game that is going to be used. After the class is created, student users are replicated in the Analytics Framework and a group with all those users is also created. With all these steps, using the Analytics Framework, an activity can be created with this group, and students will be able to send authenticated data using their anonymous identifiers.

Surveys, on the other hand, are designed to measure learning. A survey object is built with up to 3 internal surveys: a pre-gameplay survey, post-gameplay survey and, if needed, an auxiliary survey. When surveys are created, the LimeSurvey schema files are uploaded to Simva and with those schemas the surveys are created in LimeSurvey using the API, saving their identifiers for later usage.

Last, to use everything in a lesson, a survey must be assigned to a class. When this happens, all the anonymous identifiers from the class are added as participants of the surveys in LimeSurvey. Along with that, and using LimeSurvey API, Simva can retrieve survey completion status and use it to allow or deny access to the game itself, preventing the students from playing without answering the survey with a simple status request. When the game ends, an endpoint is available for game-results upload, where logs, scores, or statistics can be appended for later analysis.

The following sections describe 3 real scenarios where we have used Simva to collect information from questionnaires and game learning analytics data when validating three different serious games. In each scenario, we have used the different specific features of Simva, described in more detail in [5], according to the requirements of each situation.

Section 3 describes the experience with Conectado a serious game to raise awareness about bullying and cyberbullying. In these experiments, Simva was used to conduct the pre-post experiments to evaluate the game while also collecting GLA data. Section 4 describes the experience with the 15 Object test, a visual task to train memory. In these experiments, Simva was used to evaluate and compare two different versions of the game and two formats (paper-based and computer-based). Finally, Section 5 describes the experience with First Aid Game, a game to teach first aid techniques, where Simva was used to collect pre-post questionnaires and GLA data over an original experiment and a posterior recall experiment.

3 Evaluating a serious game to raise awareness: Conectado

Conectado is a video game of the graphic adventure genre that aims to raise awareness about bullying and cyberbullying. The game is designed as a tool for teachers to start a discussion or debriefing sessions about the topics covered in the game with their students after they all have shared the common experience of the gameplay. So far, the game has been validated through several experiments in high schools with more than 1000 students between 12 and 17 years old, as well as with more than 200 teachers and educational science students [6, 7].

The game validation consisted of pre-post experiments using a formal questionnaire which assesses the players’ awareness about bullying and cyberbullying. This question-
naire was used as pre-test and post-test in the experiments to compare players’ awareness before and after playing Conectado. Additionally, the most relevant interactions of the players with the game were collected to further analyze the players’ progress, their interactions with other game characters and their in-game choices and attitudes. All this information can provide further insight on how students have used the game and their behavior on a bullying and cyberbullying situation such as the one depicted on the game.

Before conducting the experiments, the main researcher who managed the experience prepared the pre-post questionnaires in Simva. This included the following steps: the surveys were registered in Simva, and the groups which would be using them were created, with 30 students per group included, so there will be enough space for all the students in each group. All the users created were identified by their tokens, unique sets of 4 random letters. The list of tokens provided by Simva was printed in advance and carried out to the schools. The interaction data captured by the game and sent to Simva, was also sent to the Analytics Server for its analysis. Therefore, the users created with Simva were linked to the ones created in the Analytics System used to collect and analyze user interaction data during the experiments.

During the different sessions of the experiments, the main researcher only had to distribute the different printed tokens, one per student. Students then used their tokens to access the game, which has a welcome screen to introduce the user identifier with which data will be sent to the analytics system. At this stage, the game checks in the configuration file which are the questionnaires to be used. Then, the game access Simva and checks that the assigned pre-test questionnaire is available for the introduced token. If so, it automatically opens the browser with the initial survey that players must fill in. Simva checks that the surveys are correctly configured for the user, given by the unique identifier used to enter the game. If Simva indicates that the survey does not exist or that it is not available for the indicated user, the game will not continue. When the pre-test is completed, the results are sent to Simva and users can access the game. After the gameplay is finished, the interaction data is sent to Simva. For the post-test, the same checks and process for the pre-test are repeated. If everything is correctly configured, the post-test survey is opened and, when completed, the results are sent to Simva.

Once the experiments were finished, the main researcher could download the answers to both questionnaires as well as the interaction data from the corresponding Simva screen. The different data sources captured from each user (pre-test, post-test and game interactions) were linked together by the unique identifier of each player, facilitating the next analysis step.

With the information gathered in these experiments using Simva, the evaluation of the game Conectado could be performed, analyzing that the game indeed increases the awareness about bullying and cyberbullying, as measured by the pre-post questionnaires. Additionally, analysis of the interaction data captured allowed to extract further information such as times taken to complete the game, progress, or different in-game choices and interactions with game characters.
4 Comparing two versions of a serious game for active aging: 15-Objects test

The 15-Objects Test (15-OT) is a visual task that presents 15 overlapping objects users need to identify as fast as possible. The aim of this test of visual discrimination is to evaluate the slowing of cognitive processing in Parkinson's disease [8]. The test is carried out with two figures of superimposed images of 15 objects, traditionally provided to participants on paper.

For these experiments, in addition to the traditional paper-based version of the test, we developed a new computer-based version of the 15-Objects Test with the same structure and characteristics. This new version was early tested with 18 adults [9]. For this test, two different configurations of the 15-OT were created (A and B), each with a different configuration of the 15 superimposed objects. To further compare the paper and computerized scores of each participant, as well as the two versions of the game (A and B), participants were randomly assigned to four experimental conditions, balanced by age and sex (see Fig. 3). These experiments were a proof-of-concept to test whether the computerized version of this traditional test could be used to further investigate active aging.

For this experiment, the required groups of participants were created in Simva. Both questionnaires (pre-test and post-test) were created and managed using Simva. The questionnaires were then linked to the groups of participants that were going to use them. All participants needed to complete both questionnaires in different moments according to the conditions shown on Fig. 3. Paper-based versions of the test were prepared in advance and handled to participants either at the beginning of the experiment (for participants assigned to experimental conditions II and IV) or at the end (participants assigned to conditions I and III). All participants were provided with their anonymous 4-letter identifiers at the beginning of the experiment. They were asked to write

![Fig. 3. Four experimental conditions in the 15-Object test experiments. Participants were randomly assigned to one of these four conditions. Picture adapted from [9].](image-url)
down their unique identifiers on both the pre-post questionnaires, and in the paper-based 15-OT tests as well as to introduce them on their computer-based 15-OT tests.

The particular characteristic of this experiment was that it presented two different game versions (A and B). As seen on Fig. 3, some participants (the ones assigned to conditions II and III) completed the version A of the 15-OT test on paper and the version B on the computer, while the rest of participants (assigned to conditions I and IV) completed version B on paper and version A on the computer. Therefore, it was required to know which version of the game each participant was performing on the computer (and therefore which one they were completing on paper). To have that information linked to participants’ answers to questionnaires was, if not required, at least very recommended, to then simplify the analysis step. To link this information to participants’ questionnaires, we used the metadata feature available in Simva. This feature allows to add information for each participant. Therefore, it was possible to directly add in Simva which version of the game each participant was completing on the computer. For each player, it was directly stored in Simva which version of the game each student had played on the computer (A or B). This information is displayed in Simva as shown in Fig. 4.

After the experiment, all participants’ responses to both questionnaires were stored in Simva, linked together with the information of the game version used in each condition by the unique identifier provided for students. Researchers could then analyze the questionnaire responses together with the game version to compare both the paper-based and the computer-based versions of the test, as well as to study the equivalence between the two versions of the game used.

The comparison of both game versions could be easily carried out with the information gathered in these experiments using Simva, additionally comparing the paper-based and the computer-based versions of the test. Results yielded no significant differences between both game versions, proving their equivalence. No significant differences were either found on the results between the paper-based and the computer-based version of the test, showing that the computerized version of the test is a valid and equivalent alternative to the traditional paper-based version.
5 Collecting GLA data and conducting a recall experiment:
First Aid Game

*First Aid Game* is a videogame to teach maneuvers in case of different emergency situations for young players. The game had already gone through a traditional paper-based pre-post formal validation in a previous experiment. In that validation, the game was even compared with a control group that assisted to a theoretical-practical demonstration of the same topics covered in the game [10]. That game validation experience was a clear example of the problems that these types of experiences can rise: after completing the experiments, researchers had to deal with a large number of questionnaires on paper, to read and process them and copy the results to a computer for their analysis.

A new set of experiments was carried out using *Simva*, where data were collected for more than 300 students from 12 to 17 years old [11, 12]. Students completed the pre-post questionnaires assessing their knowledge about first aid techniques, adapted from the questionnaires used on the original validation experiment [10]. For this experience, the game had been adapted to a new technology so these experiments were used to validate that the updated version of the game was still effective at increasing players’ knowledge. Additionally, the tracking of in-game interaction data was incorporated to this new version of the game. The interaction data collected while students played the game included game scores, in-game choices and responses, and their interactions with the different game elements.

In these experiments, the pre-post questionnaires and the groups of students were handled using *Simva*. At the beginning of each session, teachers provided students the tokens that they had previously downloaded and printed from *Simva*. During the session, teachers wrote down the name of each student next to their token in the printed copies. In this way, teachers are the only stakeholder who have the correspondence between anonymous 4-letter identifiers and the students they correspond to. This way

![Fig. 5. Experimental setting for First Aid Game original and recall experiments. Simva simplified the linked between the information from both experiments and the different data sources on each experiment.](image)

each token could be reused by the same student in the future. No personal information is entered in the game. Teachers were encouraged to keep the token printouts for possible future activities.

A few weeks after completing the training with the game in the school, researchers returned to perform an additional experiment to measure recall of knowledge learned with the game. For this experiment, teachers provided the same token to each student from the paper files they kept (where they had manually written the name of each student next to their assigned token). Simva tokens allowed all information from students to be grouped by student while preserving anonymity (at least for researchers), both for the original experiment and the subsequent recall experiment. This simplified the process of analyzing whether students recalled what they had learned, as all the information from their questionnaires and in-game interactions from both set of experiments could be linked together with the anonymous token. Fig. 5 depicts the experimental setting of these two consecutive experiments using the First Aid Game.

The combination of both set of experiments helped to measure how much students have learned while playing but also how much they could remember a few weeks after the original validation experiment. From their initial knowledge (measured in the pre-test of the original experiment) to their final knowledge (measured in the post-test of the recall experiment), we could determine how much their knowledge had improved with the experience and in the time in-between (where they could have had other interventions related to the topics covered in the game). In a more fine-grained analysis, recall from the first experience to the second one could be measured by comparing their final knowledge (post-test) on the original experiment with the knowledge they have a few weeks later, before any other intervention (pre-test in the recall experiment). This shows not only that players learn while playing, but also that they are able to recall the things they have learned with the game.

6 Discussion

The three experiences described on this paper exemplify how Simva has helped to simplify the validation of serious games as well as the assessment of the students playing them. The evaluation of games has been performed by combining both traditional pre-post experiments with the information collected from the in-game interaction GLA data. Additionally, some features included on Simva have also simplified the execution of experiments with specific requirements, such as comparing two game versions or conducting a recall experiment.

On the detailed experiments, all the information gathered from the different sources (questionnaires and game interactions) have been kept together in Simva and linked for each user by their unique anonymous identifier. These identifiers are provided to players by the managers of the activity, who obtained them from the lists of tokens given by Simva when creating the required classes of students. For each participant, researchers have then been able to extract all the information of the experiment: pre-test, post-test, GLA interaction data, and in the specific cases, version of the game played, or pre-test, post-test, and GLA interaction data from the following recall experiment.
From these experiences, we have found some issues that we consider are key when conducting experiments in real settings validating serious games or deploying serious games to assess students. As we consider that bearing in mind these facts could help other researchers in this or similar fields, we have summarized these issues as the following lessons learned from our work:

- **Ensuring users privacy**: to adequately conduct the pre-post experiments, it has been essential that the tool we were using to manage students and questionnaires, in this case *Simva*, automatically deals with and ensures privacy. No personal information should be input into the system collecting information from the experiments so privacy can be effectively ensured. In our experiences, neither the questionnaires tool *Simva* nor the Analytics System, where interaction data was also being sent to, collected any personal information. Despite ensuring privacy, it is still required that all the information collected from each student (e.g. pre-test, post-test, game interactions, any additional metadata) is linked together for the later analysis. For this purpose, pseudo-anonymization, via the 4-random-letters identifiers, automatically provided by *Simva* when classes of students are created, has been an effective solution, as privacy is ensured while maintaining all students’ information linked together. For other researchers in similar scenarios, we encourage to use a simple anonymization system like the one used in *Simva*, that effectively links all the information gathered from each user, simplifying the later analysis, while ensuring privacy as the user identifier does not provide any personal information and is the only identification input into the system.

- **Collecting different data sources**: the online collection of questionnaires done with *Simva* has greatly reduced the times and costs of carrying out pre-post experiments, as well as the use of paper. An additional option was also available during the experiments to collect the information offline. With this option, all data was stored in the computers were students were playing to be later collected by researchers in case of network connection problems. All the interaction data was also stored and linked with the questionnaires online and offline. The option to include additional metadata in *Simva* as well as the possibility to link the information from several experiments has also been useful as it simplifies the later analysis of all the different data collected from each user. We encourage researchers to consider options to link together all different data sources on their experiments as it simplifies the later steps of analysis.

- **Moving from pre-post experiments to GLA**: although in the three cases presented we have used the traditional pre-post experiments to evaluate serious games’ efficacy and assess students playing the game, we consider that the information extracted from in-game interactions is also essential and research should move towards always including this type of information. In our case, we have used the xAPI-SG Profile [13] as the data collection standard for the in-game interactions. We also recommend researchers to use this or other standard when collecting interaction data as it simplifies data reuse and integration in larger systems and simplifies the collection process.

We consider that the combination of these lessons learned from our work, collecting different data sources from traditional pre-post experiments to more informative GLA
data while ensuring privacy, can benefit the execution of experiments to evaluate serious games and assess the students who play them. The tool Simva has been effective in the described experiments as it has fulfilled the requirements of the three experiences and has simplified the complexity of the different steps of the process.

7 Conclusions

Despite their drawbacks, pre-post experiments are still one of the most common evaluation methods for serious games and for the assessment of students who play them. Making these complex experiments more user-friendly and reducing their costs both in time and effort can greatly improve the application of games in real settings, simplifying their evaluation and deployment, and increasing their application including assessment of players.

The tool Simva, that we have used on the three experiences described, has shown a great potential towards these simplifications. Simva manages both questionnaires and groups of players, deals with privacy issues, allows the collection of information from different data sources (both questionnaires and in-game interactions), and includes additional features for different possible requirements (e.g. adding metadata information and simplifying recall experiments) in future experiments.

On this paper, we have revised three specific applications of Simva in real settings for different goals related to the evaluation of serious games and the assessment of students playing. We have shown some lessons learned from our experiences to contribute to further research on this area.

References

7. Calvo-Morata, A., Freire-Moran, M., Martinez-Ortiz, I., Fernandez-Manjon, B.:


