Simplifying Serious Game Validation: An Integrated Tool for Experimental Design and Learning Analytics

Tharife Nabil Al-Khuffash
Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
https://orcid.org/0000-0002-3829-7028

Julio Santilario Berthilier
Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
https://orcid.org/0009-0003-2520-507X

Cristina Alonso-Fernández
Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
https://orcid.org/0000-0003-2965-3104

Antonio Calvo-Morata
Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
https://orcid.org/0000-0001-8701-7582

Baltasar Fernández-Manjón
Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
https://orcid.org/0000-0002-8200-6216

***Abstract*—This full research paper presents a tool (SIMVA) to automate the experimental design of Serious Games (SGs) validation. This tool supports diverse experimental designs (e.g., pre-post surveys, control-treatment comparisons, learning analytics based evaluations), helping researchers to manage different aspects such as participants, data, sessions, and study conditions. Validation studies usually involve collecting and managing data from multiple sources, which present several challenges in handling these data manually or with non-specialized tools. Among these challenges are the necessity of possessing technical skills and the consideration of data protection and privacy regulations. We have successfully used SIMVA to validate SGs developed by our research team, reducing efforts and costs. It has reduced the complexity of experimental setup and data collection. Also, it has allowed us to aggregate multimodal data and get visual insights. Moreover, it reduced the risks related to data coherence, persistence, and privacy. SIMVA has been primarily used and tested throughout its development by our research group. We consider that SIMVA can contribute to more scalable research on SGs by simplifying and automating tasks in SGs' evidence-based validation processes.**

Keywords—Experimental Research, Analytics, Educational Technology, Games

I. INTRODUCTION

In e-learning (learning methods that use electronic means in the learning process), vast amounts of data are being collected from learning activities [1] , and Learning Analytics (LA) is the process of capturing and analyzing all learners' static and dynamic data and their contexts [2] to improve the learning process [3] [4]. Different stakeholders seek LA [1] for evidence-based decisions: it enables teachers to make appropriate interventions [3], directors to adopt enhancements at the institutional level [2] and students to reflect on their learning [5]. In addition, LA contributes to different educational domains; among these is game-based learning (GBL) [6], which is one of the growing approaches in education [7] that utilizes games as a learning activity [3]. GBL provides better learning experiences by motivating and engaging learners [8], and it has been shown to improve learning outcomes compared to traditional learning methods [3]. In GBL, students learn better because they perform a task rather than just knowing how to perform it, which is referred to as the learning by doing theory [9], [10]. Moreover, GBL facilitates other learning theories, such as situated learning theory, which indicates that learning is more efficient when the learning context puts the learner in scenarios that are like real-life situations [10].

Recently, there has been growth in using digital games in classes, especially Serious Games (SGs) [11], which are digital games designed for different purposes other than entertainment [3], such as to increase knowledge or to raise awareness toward social issues. With SGs, students could be more motivated and excited, which improves the learning experience and promotes innovative thinking [8]. However, although SGs have the potential to bring all these advantages to an educational context, they require validation before deploying them in classes. Evidence on their effectiveness is needed so that the stakeholders can trust SGs. Usually, this process is complicated and presents different challenges. One major challenge is handling extensive data from different sources (e.g., Learning Analytics, validation, or usability questionnaires). Additionally, the use of multiple technologies when collecting the validation data (e.g., to capture data from the games, store it, and analyze it) [12] increases the process complexity [13]. Further difficulties lie in data privacy [14] and technical skills literacy [15]. Data protection and privacy should be ensured during the different phases of the validation process, from collecting to analyzing to storing data. Data privacy in LA is challenging; data should not be personally identified, especially in sensitive fields [16], such as in the medical domain, which forms 50% of the primary use of SGs, and in the educational domain, which forms 20% [17].To do experiments in these fields, it's very critical to determine how the data are used, who has access to them, and to get participants' consent to collect their data [16], [18].

On the other hand, the validation process includes the use of different tools for collecting, analyzing, and storing data, which demands technical expertise that not all researchers have. Thus, if the mentioned issues are not addressed, LA is considered costly, time-consuming [12], and skills-demanding.

In this paper, we propose a multimodal platform that simplifies integration and the analysis of heterogeneous learning data from educational scenarios that use Serious Games (SG). We refer to this platform as "SIMVA". The rest of this paper is structured as follows: Section 2 reviews related work on SGs validation; Section 3 presents our proposed tool (SIMVA) to simplify SGs validation, integrating LA and the configuration of experimental designs; and Section 4 discusses the results of the use of SIMVA for the validation of several SGs in real scenarios. Finally, Section 5 summarizes the conclusions and some future lines of work.

II. RELATED WORK

Evidence-based validation is vital in SGs [19] as it assists all interested parties to make decisions about the game [3]. Game designers and developers need validation to identify game improvements and ensure that their educational and game design meet their intended goals. Also, researchers need validation to evaluate the effectiveness of SGs as educational activities [2]. Without the evidence that this validation provides, it is highly unlikely that educational centers have enough trust in these tools to integrate them into the curricula.

Usually, researchers validate SGs within experiments that include different activities where participants play the game to be evaluated [16]. They collect multiple datasets: game interactions, players' perception of the game, experts' feedback, and the change in participants' knowledge or behavior. Data collection for validation may take place at multiple time points or at staggered times. These data are represented in multiple formats at their origins, and they need to be systemized to make the appropriate analysis, which may require special technical skills not all researchers have.

In the literature, researchers have used several techniques to collect data for the SGs' validation experiments. According to a systematic literature review [20], the most frequent technique was using questionnaires; the percentage lies between 55% and 90% [17], [20]. Other methods also applied were observations, discussions [20], and semi-structured interviews [17]. Some studies used a combination of these techniques, such as a study to measure the impact of a game on teaching software project management. They used online Google Forms to collect data about participants' demographics, perceived learning, game flow, intrinsic motivation, and usability. In addition, they observed participants' behavior (body language and facial expressions) while playing the game and interviewed them immediately after playing the game [21].

The data collected through the questionnaires also varied. Some studies used self-efficacy surveys before and after playing the game to address the change in the level of knowledge or awareness. For example, a research team at Queen's University Belfast evaluated an SG on academic integrity using pre/post tests. In this study, students answered Likert questions on Microsoft Forms about their knowledge of academic integrity before and after playing the game. Later, the researchers used SPSS (a statistical analysis software) to analyze the collected data [22]. Other studies collected data about users' feedback on the SG, like the case in the UnderWater game, a game on teaching ultrasound technology. The researchers collected 42 participants' responses on paper-based questionnaires and analyzed the results with SPSS software. The pre-test collected information about participants' background on the topic, and three post-tests: about their prior experience in playing video games, and two to validate the game itself as an educational tool, its design and content [23].

These external measures are not the only possible data collected in the SG validation process. Game Analytics (GA) are the extracted game-interaction data in games in general, which includes rich and extensive data that may also be collected from Serious Games (SG) [24]. These data can be categorized into technical and player data. Technical data include data about the game itself, like memory usage and number of errors, which are important to game developers and designers to improve the game's performance and design. Player data includes players' interactions with the game, such as time spent in it, choices made, number of replays, and scores obtained [25]. GA traces are usually sent in real-time [24] and continuously (instant by instant), so that even a quick play can generate massive datasets [25].

When using SGs in an educational context, more contextualized data associated with the learning objectives and competencies expected to be learned/trained in the game are also essential for the SG validation. In other words, unstructured, more complex, and extensive GA data can be combined with Learning Analytics (LA) data, relevant data about the learners' process, which is usually collected from traditional e-learning methods (e.g., Learning Management Systems). In 2016, Freire and colleagues formulated an early definition for Game Learning Analytics (GLA), which is the process of collecting/analyzing players' learning data and game interaction data from serious games [25].

However, although the types of data collected during the experiments varied in the literature, very few studies have used game analytics data. For example, in Newton's Playground SG, researchers had embedded stealth assessments for the learning competencies inside the game. This way, they could capture students' progress on set goals from the game logs (e.g., positions students struggled with most, average time spent), which were converted into assessment indicators. In addition to game logs, they used pre/post tests to address the change in the conceptual understanding of physics after playing the game [26]. In [27], authors evaluated the Mission HydroSci (MHS) game, a game that teaches middle-school students water science with a proposed approach of a logging system, designed and implemented in association with the game design and implementation. The idea was to identify in-game features that correspond to learning performance.

To conclude, the most frequent technique to collect data in SGs validation experiments is questionnaires, where researchers need to design the surveys, whether in paper as in the case of the validation of UnderWater game [23] or using available digital tools. Moreover, these experiments are often conducted in several sessions, requiring efficient data analysis and control approaches throughout the whole life cycle of the validation process. However, capturing GLA has the potential to improve the game learning design and the validation process, but the challenge is in aggregating and relating surveys' data with game interaction data, which is challenging due to their diverse representation at their origin.

On the other hand, data analysis is a fundamental stage, and the deficiency of specialized tools to analyze the collected data to validate SGs adds further intricacy to the process. Having to create and collect surveys' responses in a tool and using another tool to analyze the data requires additional effort, especially to avoid the loss of the data. For instance, in the study at Queen's University Belfast, researchers used Microsoft Forms to collect data and SPSS to analyze them [22].

Another concern to consider when validating SGs is the data ownership and data privacy regulations [14]. Since data are collected from diverse learning resources and multiple stakeholders may have access to them, this raises the question of data ownership. Researchers should comply with the data privacy regulations (e.g., GDPR) and not collect participants' personal information [16].

We have searched in the most common databases like Google Scholar and ACM, and we have not found tools that simplify the entire validation process and offer the capability to unify the format of different data sources (questionnaires, games) or anonymize data at the source to comply with GDPR. Although studies mention using questionnaires or interaction data for game validation, they do not mention any tool that supports all these capabilities. Neither mentioned tools that simplify all the steps needed to validate an SG: to manage different experimental designs, to collect and track both interaction data (GLA) and responses to surveys, and to integrate these diverse data to extract meaningful insights. For example, in "StoryPlay" (a mobile multimodal platform for SG analytics), researchers evaluated SGs by capturing extensive data on players' behaviors from gameplay, mobile sensors, and human observations. This platform supports pre-post tests as part of the observer workflow but not as a full experimental design setup [28]. We aim to fill this gap with our proposed tool, SIMVA, an integrated tool to support SG validation.

III. SIMVA

SIMVA is a free and open-source tool designed to streamline the validation process of Serious Games (SGs). Researchers can use it through all the stages of the validation process, from designing experimental studies to accessing real-time insights [29]. Using SIMVA, researchers can set up diverse experimental studies like A/B testing, pre-post experiments, and studies with control groups [30] and manage the participants. Also, they can create study questionnaires (e.g., pre/post-tests), collect participants' answers to these questionnaires, and collect game interaction data. SIMVA standardizes the collected data and automates the collection/analysis process, and provides visual insights, decreasing effort and limiting the probability of faults caused by manual intervention.

In the following subsections, we discuss SIMVA's systematic approach to collecting data while conforming to data privacy standards. Also, we explain how it simplifies and automates the SG validation process.

A. *Game Learning Analytics: From Data to Visual Insights*

Across the lifecycle of the SG validation process, data are received in different formats from heterogeneous sources, which complicates the later data analysis. SIMVA orchestrates the data flow from multiple origins, identifies participants across the different stages with anonymous identifiers, and works with external tools efficiently. SIMVA constitutes a credible data ecosystem for evidence-based research.

The validation surveys are created on a third-party platform called LimeSurvey. SIMVA's authentication is accomplished with the Single Sign-On (SSO) service provided by an open-source identity and access management platform called Keycloak. With this service, users (whether the researchers to create the surveys or the participants to answer them) are redirected from the platform to LimeSurvey using unified credentials. Also, Keycloak generates random tokens that the researcher can export to provide each participant with his/her token (in person). Participants will use this token as their unique identifier to sign in to SIMVA.

This approach enables the mapping of individual participant's data across different phases of the experimental study, including game interactions and surveys' data. This way, SIMVA makes a direct connection of all individual participants' data from the different activities included in the study while enforcing privacy and security controls [16], [30] since no personal information is collected or stored.

The unification of the diverse data in SIMVA is based on the IEEE xAPI-SG standard. We adopted the IEEE Experience API for SG profile, an extension of a learning experience standard, IEEE Experience API (IEEE xAPI), that represents data in a JSON format [30], [31]. The xAPI-SG standard is built on xAPI to record Game Learning Analytics (GLA), i.e., data captured from educational contexts that use Serious Games as learning resources. In general, the xAPI standard is developed to track universal learning activities (online and offline) from diverse learning environments (Learning Management System (LMS), mobile apps, real-world activities, etc.). The standard tracks learning activities in detailed and structured statements. Each statement contains three essential fields: actor, verb, and object. The actor field represents the performer of the action, the verb field represents the action itself, and the object field represents the entity with which the actor interacted. The key advantage of xAPI is that its structure is extensible and more fields (i.e., data to be captured) can be added. Typical fields include context, timestamps, and results of the activity. In the context of SGs, there is a need to define common vocabulary and patterns. The xAPI-SG profile (outlined in Table I) addresses this issue by defining SG's frequent interactions with flexibility to add game-dependent vocabulary. The definition consists of three main interaction categories: completable, accessible, and alternative. Also, it refers to the corresponding game objects. Each interaction is identified in the xAPI traces using defined verbs. In addition, the profile defines the game's entities that fall under each category. For instance, the game itself is an entity of the category completable; therefore, we can collect the progress in it by using the defined verbs initialized, progressed (with multiple degrees of progression), and completed.

TABLE I. XAPI-SG PROFILE

Interaction category	Verbs used in xAPI-SG traces	Game's entity on which the interaction is carried out
completable	initialized progressed completed	game, session, level, quest, stage, combat, storynode, race
accessible	accessed skipped	screen, area, zone, cutscene
alternative	selected unlocked	question, menu, dialog, path, arena
gameobject	interacted used	enemy, npc, item

Having a structured profile for SG allows capturing data from any game that speaks this language. This way, consistent tracking across different games and analytical tools is achieved.

Fig. 1 illustrates an example of an xAPI-SG statement representing the game interaction "John Smith progressed on Level 1". As we can see in the additional field of result, his achieved progress is 0.5.

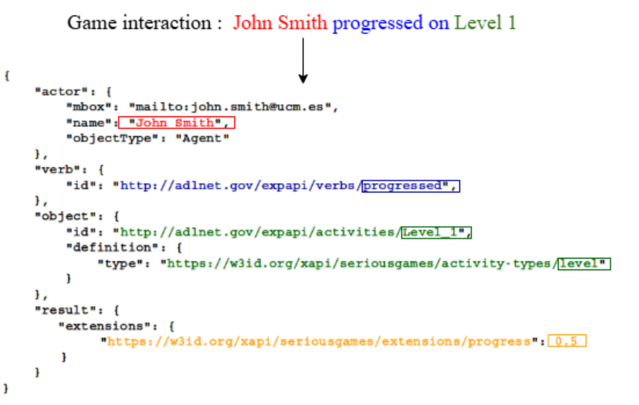

Fig. 1. Example on xAPI-SG trace

Based on this standard, we can capture all relevant information for an SG validation process in SIMVA:

1) Game interaction data: SGs communicate with the platform using SIMVA′s API. This connection allows the games to send xAPI-SG interaction data traces through any of the available game trackers (e.g., JSTracker in the case of web games or through C#Tracker in the case of desktop Unity games). SIMVA′s API checks whether the traces are in valid JSON format and structured according to the xAPI standard, containing the main fields: actor, verb, and object. The traces are sent in real-time and stored in our Learning Record Store (LRS). Traces are identified with the pseudonymous token sent by Keycloak.

2) Survey Data: LimeSurvey sends survey responses to SIMVA′s API via the LimeSurvey xAPI tracker, which encodes responses from its proprietary format into a consistent and structured model such that survey data are captured as xAPI traces. We used the xAPI SCORM (Sharable Content Object Reference Model) profile with a proven vocabulary to represent the life cycle of a survey throughout its different stages [31]. Each xAPI trace should include a verb data field with one of the following values: initialized, progressed, responded, or terminated. Table II explains the significance of each verb. Responses to surveys are sent in real-time and identified using Keycloak tokens. If the validation experimental study has more than one survey (e.g., pre/post), each survey is stored in a different endpoint in the LRS. Following this approach, it is also possible to automate data integration and transformation from other survey platforms, which ensures scalability.

TABLE II. XAPI-SCORM PROFILE

Verb used in xAPI-SCORM traces	Description
initialized	The participant started answering the survey
progressed	The participant finished a part of the survey given by a float value within the interval [0-1]
responded	The participant answered a survey question. This trace also contains "interactionType" field to distinguish different types of survey questions (date, ranking, text, etc)
terminated	The participant answered all the required questions in the survey

3) Visual Insights: While participants are answering the surveys or playing the game, SIMVA provides information about each participant's progress in the different activities of the validation process (e.g., playing the game, answering a questionnaire) in real-time on a dashboard. As seen in Fig. 2, the dashboard of the survey activity consists of two main parts: progress bars and a table. The progress bars indicate the extent of achievement at the group level for the activity while the table contains the list of participants, showing for each participant: the interaction status with the activity ("Completed" column in Fig. 2), a completion percentage in the range [0-100] ("Progress" column in Fig. 2), and a button to download the received traces in a file (a blue arrow in "Result" column in Fig. 2). In gameplay activity, it is possible to change a participant′s status manually. This may be needed for certain cases of technical issues on the participant's side, when it is necessary to allow participants to be able to continue with the activity (e.g., access the post-game survey).

Completed: 20% [2/10]

Results: 20 (50) % [2 (5) /10]

Progress:20 (50) % [2 (5) /10]

User	Completed	Progress	Result
xztr	true	100%	Completed
gznc	false	50%	Started
vrhv	false	0%	No Results
wdvo	false	0%	No Results
cqpd	false	0%	No Results
qoex	false	0%	No Results
vmxm	false	0%	No Results
cdnz	false	25%	Started
eheq	false	75%	Started
evba	true	100%	Completed

Fig. 2. SIMVA's real-time dashboard for post-survey activity

In addition to the real-time dashboard, we developed a trace monitor module (T-Mon). With this module, it is possible to analyze and visualize game-interaction traces using the xAPI-SG. It transforms raw game interaction data into informative insights, which are more understandable by non-experts. By adopting xAPI-SG, T-Mon can be used with other games that speak the same language: the xAPI-SG format. Currently, it is under improvement to include the analysis of the survey responses, now collected in xAPI format, which will enhance the analysis process of all data sources and ensure data integration for different resources. Fig. 3 displays some example visualizations generated in T-Mon, based on the xAPI-SG fields: a pie chart of games started and completed, and progress in the game for players.

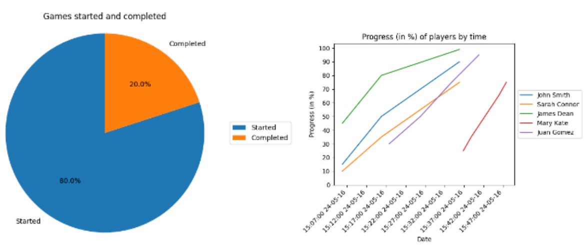

Fig. 3. Sample of reports generated by T-Mon

To conclude, SIMVA unifies and standardizes the representation of the data collected from multiple sources. All the stored traces align with the same standard (xAPI) and utilize the anonymous tokens provided by SIMVA as identifiers. Data are aggregated to present comprehensive insights across all the activities during the SG validation process (e.g., gameplay data, pre-post tests responses) using these tokens. With these anonymous identifiers, SIMVA links the data of the participant throughout all the study phases and activities, while no personal data are stored in the system. This way, we simplify the validation process, ensuring compliance with data protection and privacy regulations (e.g., EU GDPR).

B. Experimental Design Automation

SIMVA provides the researcher with full control over the SG validation process. SIMVA solves the most challenging issue in scientific experiments, which is data collection and handling.

Using this tool, the researcher interacts with a high-level abstracted platform that operates with external technologies to provide several functionalities. With SIMVA, researchers can capture data from different sources (e.g., game via telemetry libraries -trackers-, survey), create surveys and collect their data, manage participants, collect game interactions, track progress in real-time, and access visual insights on the analyzed data.

Fig. 4 displays an overview of the ecosystem of SIMVA. SIMVA-frontend allows researchers to configure participants, studies, and activities (middle part, in orange). They sign in to the platform via Keycloak (middle-left in red), which provides an SSO service such that if survey creation is needed, the researcher is redirected from SIMVA-frontend to an external survey manager platform (LimeSurvey) with the same credentials. The LimeSurvey xAPI tracker (middle-right in blue), transforms survey responses into an xAPI-supported format (same format used in gameplay data). While playing the game, the trackers (top-left in blue) capture game-interaction data as xAPI traces. Currently, we have two trackers: JSTracker to collect interactions from web games and C#Tracker to collect interactions from Unity games. With these trackers, survey responses and gameplay data formats are unified and stored in the database consistently and linked to their corresponding participants using pseudonymous identifiers. In the meantime, Kafka (top-right part in Fig. 4, in blue), which is an external open-source distributed event streaming platform, acts as a trace processor receiving the traces from SIMVA and processing them to ensure no data are lost. Then, it distributes them into multiple storage repositories. The tool enables the researcher to filter the records and decide which data to use in the analysis, depending on the experimental design and the type of research. With SIMVA, the researcher can exclude data with a negligible impact. For example, the data from a participant who answered the pre-post surveys but did not play the game or played it for a brief period.

For data persistence, SIMVA provides an option to enable the storage of game interaction traces, which are large amounts of data. When this option is activated, gameplay traces are stored in an external open-source software called MinIO (top-right part in Fig. 4) database. Researchers can later access these data and download them, which ensures the scalability of SIMVA for long-term audit and validation.

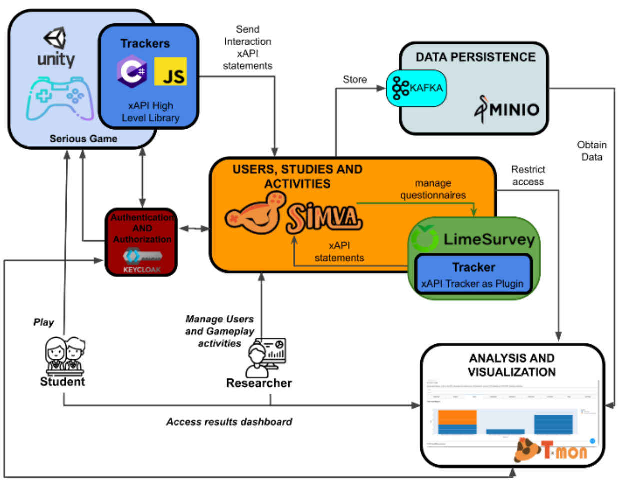

Fig. 4. Overview of SIMVA's ecosystem

The main items in SIMVA's interface are called SIMLETs (SIMVA APPLETs) and Session. A SIMLET is a container in the SIMVA-front-end web application to encapsulate the data of different sessions that serve an equivalent purpose. For example, to compare the effectiveness of a particular SG with other educational resources or to compare the effect of a SG on different groups of students. A session is a container in the SIMVA-front-end web application to encapsulate the data of several activities (e.g., pre/post-test, gameplay, traditional learning activity) to be undertaken by a group of participants.

Fig. 5 shows a visual illustration of SIMVA's representation of two experimental studies with two SIMLETs and different sessions. The left side of Fig. 5 represents a validation of an SG called "Articoding", a game to teach computational thinking. Consider that the researcher wants to measure the effectiveness of this game by carrying out a control-treatment comparison study, using pre-post questionnaires that assess participants' knowledge of programming before and after playing the game. In SIMVA, the researcher creates a SIMLET with the goal of validating the SG to learn programming. For that SIMLET, the researcher creates two different sessions: (1) a session with the experimental group, who would go through a pre-post session playing the game; and (2) a session with the control group, who would go through a pre-post session but learning with a traditional method. Therefore, the first session has three activities: pre-test, gameplay, and post-test; while the second session has three other activities: pre-test, traditional learning activity, and post-test. In the right part of Fig. 5, we display another scenario: the validation of another SG called "Conectado". In this case, the same three activities (pre-test, gameplay, and post-test) are compared among three different groups of participants, one group in each session.

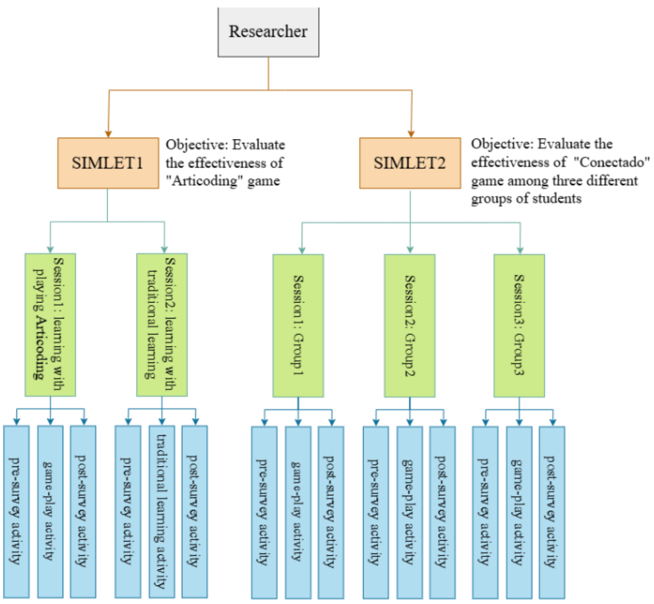

Fig. 5. SIMVA hierarchy to represent the validation process for two SGs.

The current version of the interface is available in two languages: Spanish and English. It enables the use of customized colored tags that help in categorizing and filtering the different SIMLETs. In a session, the researcher can create tags by project, by topic, by school, etc. As seen in Fig. 6, a SIMLET shows all the tags associated with its sessions. For example, the researcher can create a tag to indicate whether a particular session is a test or a real one, or a tag to indicate which SG is being played in the session.

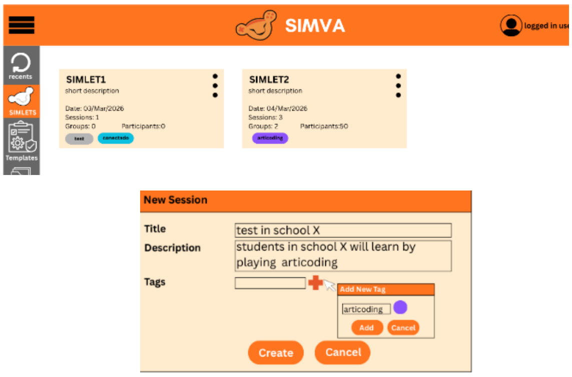

Fig. 6. SIMLETS with some tags (top) and adding a tag feature when creating a session (bottom).

Overall, in SIMVA, the researcher can design a full experimental study with several activities and determine the study flow, for example: pre-survey, then gameplay, then post-survey. When needed, the flow can be modified. Once the study is launched at the participant client, the different activities (questionnaires, gameplay) will be automatically executed in the same order the researcher determined in the setup without interruptions. The study design setting and configuration are also stored with the ability to modify them or reuse them.

IV. RESULTS

A. *Evaluation*

SIMVA was developed through an incremental, iterative process and was primarily used and tested by our research group throughout development.

Moreover, an initial evaluation was conducted with eight researchers specializing in educational technologies. The evaluation focused on assessing the functionality provided by SIMVA for conducting experiments with digital educational activities (such as serious games or web applications) and questionnaires. The researchers were asked to set up a SIMLET and a session from scratch. They were asked about the tool's ability to simplify experimental design, its usefulness, ease of use, and their interest in using the tool in the future. Also, they were asked for their opinions on the tool's strengths and weaknesses. Table III summarizes the findings of this evaluation, where responses were rated on a 5-point Likert scale from strongly disagree (1) to strongly agree (5).

TABLE III. SIMVA'S INITIAL EVALUATION RESULTS

Survey Item	Mean	Median	STD
I understand the SIMVA system and the options it provides for simplifying experimental design.	3.72	4	0.83
I find the SIMVA system useful.	3.81	4	0.99
I consider myself capable of using the SIMVA system.	3.53	4	1.13
I would be interested in using the SIMVA system.	3.12	3.5	1.2

Responses to the survey's open questions show that researchers' feedback highlights the benefit of having a consistent format for all data and the ease of linking data for a single user by using a one-time code. On the other hand, they point out the need to further simplify the interface and improve the visualizations.

B. Deployment

In our earlier SG validations, the validation process was complicated, and it was challenging to handle questionnaires on paper and manually link data from the same participants coming from multiple sources. Moreover, this approach often resulted in ad-hoc modifications in the game and server to send and capture analytics data. With the deployment of SIMVA, the process was streamlined, all tasks were automated, and it offered real-time control throughout the process. The analysis relied on more reliable data with fewer mistakes while protecting participants' privacy.

We have used SIMVA to validate several SGs developed by our research team. With SIMVA, we could adopt control vs experiment, A/B testing, and pre/post studies, and all the steps were automated, and no manual intervention was needed. At the time of drafting this paper, we have used it to validate 6 SGs (fully designed and developed within the research group) across Spain and Mexico, in a total of 16 educational centers, where we had full control over the experiments, and data handling was simplified. In these experiments, data were collected from more than 2300 students, teachers, and volunteers. The platform has also been used to support the validation of other SGs within the context of two European and two national (Spanish) projects, as well as in a Serious Games university course to test several of the games developed in the class.

For example, we used SIMVA to validate "Conectado", which is a free and open-source web game developed by our research team. It is a game on cyberbullying awareness among students aged in the range of 12-17 years. The experiments were conducted in 11 schools where +1400 users (students and teachers) participated. The validation aimed to evaluate the game's effectiveness in changing students' empathy toward cyberbullying victims. This was measured by collecting participants' perceptions of cyberbullying before and after playing the game. In SIMVA, we created the pre-post tests and gathered the students' responses to them. Data gathered from the surveys, along with game traces, were aggregated in the tool. The tool allowed us to compare the results of the pre-post surveys for each participant, and we found that their empathy toward victims of cyberbullying had increased after playing the game [15], [32].

Moreover, we used SIMVA to validate the different versions of other SG, which is "Articoding" (a game to promote programming learning) across several schools [33], [34]. The tool aggregated data from over 500 participants (students and lecturers). Diverse data were collected, including game interactions, pre-post questionnaires, and surveys about the game's usability, the students' prior programming knowledge, and participants' feedback. Also, SIMVA optimized the data-integrity verification. It allowed us to identify incomplete responses (e.g., records of the participants who either failed to submit the post survey or did not play for a significant amount of time) and to exclude them [33].

When the first version was validated, the visual insights provided by SIMVA facilitated the extraction of the findings. The analyzed game traces (e.g., time spent on each level, number of replays) [33] allowed us to detect that the players had problems with the concept of variables and their use. This issue was addressed in the next version of the game. In the later stages, we validated the second version of this SG with another group of students in different schools. The main finding was that the students progressed faster than in the previous version of the game with the parts of variables and their use, which means that the modification in the game had been useful [33].

Finally, all these experiments conducted by our team were transparent and preserved data privacy regulations (e.g., EU GDPR). They have been formally approved as authorized educational activities by each school's review committee. Also, participants were informed about the experiment's goals and terms, and that their data would be treated confidentially for scientific research purposes, and they had the right to accept/reject the contribution. To accomplish that, we provided a notification message at the beginning of the sessions with a mandatory checkbox to accept/reject the terms. Participants who did not check this checkbox could not access the session's activities and were excluded from the study. In addition, we printed the list of random tokens from SIMVA, and each participant was given his/her token. This token was used to access the game and the surveys. The dashboards that showed real-time progress displayed the aggregated data using these tokens as identifiers. Teachers in charge of the sessions were the only users given this list, so they could manually write down the student who corresponded to each anonymous identifier (if they wanted to do so, to then be able to relate the information from the dashboard back to the corresponding student). Researchers participating in the sessions did not have access to this list, and the matching between names and identifiers (if done) was never input into the system.

Overall, the experiences we carried out using SIMVA have been positive and have allowed us to continuously test and improve the tool prior to its future validation outside the context of our research group.

V. CONCLUSIONS AND FUTURE WORK

To increase the application and deployment of Serious Games (SGs) in real scenarios, a more formal and complete validation that ensures that those SGs meet their intended goals is required, so educators and institutions have trust that their use will bring clear educational benefits. Researchers typically collect data from questionnaires to validate SGs [14], [16], while other studies have started to incorporate GLA in the validation process for better analysis [21]. So far, there is no standardized model for these heterogeneous data. Moreover, this validation process includes significant efforts to create the experimental design and prepare for it (e.g., creating questionnaires), to control the validation sessions (e.g., participants' management, multi-day sessions), and to gather/integrate/analyze data from diverse sources (e.g., game interactions and survey responses) [16], [18]. Therefore, adopting conventional validation methods is costly, protracted, arduous, and with risks of inconsistency. In addition, it typically requires a range of skills (experimental design, data analysis, privacy regulations) that are not commonly found in all researchers. To address these challenges, we developed a standardized and modular tool

(SIMVA) that grants full control over the different stages of the SG validation process, preserves full ownership of data, limits the technical barriers, and unifies the representation of heterogeneous data sources.

These features are achieved by interoperating with multiple external components and by adopting the xAPI standard for data collection. SIMVA converts all the captured data into an xAPI-supported format, integrates the data from the different stages of the validation process, and links them to the corresponding participants using anonymous identifiers. SIMVA user interface presents coherent and relevant data on its real-time dashboard, which supports evidence-based validation. Thus, validating SGs using SIMVA streamlines the validation and reduces manual intervention, which makes it a time-efficient and cost-effective process.

With all these features, SIMVA has been used in the validation process of several serious games within our research group, enabling researchers to carry out the whole process, reducing the effort required to carry out experiments. SIMVA simplifies setting up different types of experimental designs (pre-post comparison studies, experiments with control groups, etc.), managing participants and classes, including privacy and ethical aspects, and simplifying the later analysis of the collected data from different data sources (the games themselves via connected trackers and the different pre-post questionnaires used in the studies). This way, SIMVA has contributed to making serious games validation more feasible by avoiding different experimental designs and game learning analytics mistakes.

SIMVA has been through several upgrades to reach its current state. A continuous improvement to its interface is being implemented to optimize user experience. Also, its interoperability opens doors to integrating additional components. For example, currently we are working on developing an xAPI tracker to support games developed in Godot, a free and open-source game engine. On the other hand, since SIMVA can be embedded into Learning Management Systems (LMS) that support the xAPI standard, we are working to include more learning activities to be validated with this tool (not only digital games). In addition, we are in the migration process for T-Mon from Jupyter Notebook (where it was originally developed) to a web-based Python application to enhance its performance and to simplify data analysis from larger amounts of data. Finally, we plan to involve a more diverse group of experts in the future evaluations of SIMVA.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

Acknowledgment

This work was partially funded by the Ministry of Education (PID2023-149341OB-I00), and by the Telefonica-Complutense Honorary Chair on Digital Education and Serious Games.

References

[1] M. Marolt and D. Jakšić, "Overview of Research on Higher Education Teachers' Involvement in Learning Analytics," *Organizacija*, vol. 59, no. 1, pp. 95–109, Feb. 2026, doi: 10.2478/orga-2026-0005.

[2] D. Ifenthaler, D. Gibson, D. Prasse, A. Shimada, and M. Yamada, "Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners," *Educational Technology Research and Development*, vol. 69, no. 4, pp. 2131–2150, Aug. 2021, doi: 10.1007/s11423-020-09909-8.

[3] M. Boothe, M. Gopalakrishnan, M. Huynh, Y. Wang, and X. Ochoa, "Game-Based Learning Analytics: Insights from an Integrated Design Process," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , vol. 15259 LNCS, pp. 108–123, 2025, doi: 10.1007/978-3-031-74138-8_9/FIGURES/3.

[4] E. İnan and M. Ebner, "Learning Analytics and MOOCs," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12205 LNCS, pp. 241–254, 2020, doi: 10.1007/978-3-030-50513-4_18.

[5] A. A. ElSayed, A. A. ElSayed, M. Caeiro-Rodríguez, F. A. MikicFonte, and M. Llamas-Nistal, "Research in Learning Analytics and Educational Data Mining to Measure Self...," *World Conference on Mobile and Contextual Learning*, vol. 2019, no. 1, pp. 46–53, Sep. 2019, Accessed: Mar. 23, 2026. [Online]. Available: https://www.learntechlib.org/p/210600/

[6] Y. J. Kim, J. A. R. Valiente, D. Ifenthaler, E. Harpstead, and E. Rowe, "Analytics for Game-Based Learning," *Journal of Learning Analytics*, vol. 9, no. 3, pp. 8–10, Dec. 2022, doi: 10.18608/jla.2022.7929.

[7] F. Horn and S. Göbel, "AI as a Co-creator: A Survey on AI Support for Educational Game Authoring Tools," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , vol. 15259 LNCS, pp. 3–18, 2025, doi: 10.1007/978-3-031-74138-8_1.

[8] T. Anastasiadis, G. Lampropoulos, K. Siakas, T. Staff, and M. Student, "Digital Game-based Learning and Serious Games in Education," *International Journal of Advances in Scientific Research and Engineering (IJASRE), ISSN:2454-8006, DOI: 10.31695/IJASRE*, vol. 4, no. 12, pp. 139–144, Dec. 2018, doi: 10.31695/ijasre.2018.33016.

[9] Y. Anzai and H. A. Simon, "The theory of learning by doing," *Psychol. Rev.*, vol. 86, no. 2, pp. 124–140, Mar. 1979, doi: 10.1037/0033-295X.86.2.124.

[10] S. K. Banihashem, H. Dehghanzadeh, D. Clark, O. Noroozi, and H. J. A. Biemans, "Learning analytics for online game-Based learning: a systematic literature review," *Behaviour & Information Technology*, Sep. 2024, doi: 10.1080/0144929X.2023.2255301.

[11] C. K. W. Tan and H. Nurul-Asna, "Serious games for environmental education," *Integrative Conservation*, vol. 2, no. 1, pp. 19–42, Mar. 2023, doi: 10.1002/inc3.18.

[12] L. K. Lee, S. K. S. Cheung, and L. F. Kwok, "Learning analytics: current trends and innovative practices," *Journal of Computers in Education 2020 7:1*, vol. 7, no. 1, pp. 1–6, Feb. 2020, doi: 10.1007/s40692-020-00155-8.

[13] I. J. Perez-Colado, A. Calvo-Morata, C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Simva: Simplifying the scientific validation of serious games," *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019*, pp. 113–115, Jul. 2019, doi: 10.1109/ICALT.2019.00033.

[14] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review," *Online Learning*, vol. 20, no. 2, Jan. 2016, doi: 10.24059/olj.v20i2.790.

[15] A. Calvo-Morata, M. Freire-Morán, I. Martínez-Ortiz, and B. Fernández-Manjón, "Applicability of a Cyberbullying Videogame as a Teacher Tool: Comparing Teachers and Educational Sciences Students," *IEEE Access*, vol. 7, pp. 55841–55850, 2019, doi: 10.1109/ACCESS.2019.2913573.

[16] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Simplifying the validation and application of games with Simva," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11984 LNCS, pp. 337–346, 2020, doi: 10.1007/978-3-030-38778-5_37.

[17] E. Pacheco-Velazquez, L. Rabago-Mayer, A. Bester, and V. Rodes-Paragarino, "What do we evaluate in serious games? A systematic review," *European Conference on Games Based Learning*, vol. 17, no. 1, pp. 482–489, Sep. 2023, doi: 10.34190/ECGBL.17.1.1627.

[18] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Evidence-based evaluation of a

serious game to increase bullying awareness," *Interactive Learning Environments*, vol. 31, no. 2, pp. 644–654, 2023, doi: 10.1080/10494820.2020.1799031.

[19] C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Improving evidence-based assessment of players using serious games," *Telematics and Informatics*, vol. 60, p. 101583, Jul. 2021, doi: 10.1016/J.TELE.2021.101583.

[20] A. Calderón and M. Ruiz, "A systematic literature review on serious games evaluation: An application to software project management," *Comput. Educ.*, vol. 87, pp. 396–422, Sep. 2015, doi: 10.1016/j.compedu.2015.07.011.

[21] A. I. Wang, V. A. Knutsen, and E. Askestad, "Balancing enjoyment and learning in teaching software project management with game-based learning," *Computers and Education Open*, vol. 7, no. 3, p. 100226, Dec. 2024, doi: 10.1016/j.caeo.2024.100226.

[22] L. Creighton *et al.*, "Promoting Self-Efficacy of Nursing Students in Academic Integrity Through a Digital Serious Game: A Pre/Post-Test Study," *Nursing Reports 2025, Vol. 15,* vol. 15, no. 2, Jan. 2025, doi: 10.3390/nursrep15020045.

[23] T. J. Olgers, J. M. van Os, H. R. Bouma, and J. C. ter Maaten, "The validation of a serious game for teaching ultrasound skills," *The Ultrasound Journal 2022 14:1*, vol. 14, no. 1, pp. 29-, Jul. 2022, doi: 10.1186/s13089-022-00280-8.

[24] A. Nguyen, L. A. Gardner, and D. Sheridan, "A framework for applying learning analytics in serious games for people with intellectual disabilities," *British Journal of Educational Technology*, vol. 49, no. 4, pp. 673–689, Jul. 2018, doi: 10.1111/BJET.12625.

[25] M. Freire, Á. Serrano-Laguna, B. M. Iglesias, I. Martínez-Ortiz, P. Moreno-Ger, and B. Fernández-Manjón, "Game Learning Analytics: Learning Analytics for Serious Games," *Learning, Design, and Technology*, pp. 1–29, 2016, doi: 10.1007/978-3-319-17727-4_21-1.

[26] V. Shute, M. Ventura, by Les Todd, D. University Photography, and S. Assessment Shute, "Stealth Assessment: Measuring and Supporting Learning in Video Games," Mar. 2013, doi: 10.7551/mitpress/9589.001.0001.

[27] W. Lu, J. Griffin, T. D. Sadler, J. Laffey, and S. P. Goggins, "Serious Game Analytics by Design: Feature Generation and Selection Using Game Telemetry and Game Metrics--Toward Predictive Model Construction.," *Journal of Learning Analytics*, vol. 10, no. 1, pp. 168–188, 2023, doi: 10.18608/jla.2023.7681.

[28] L. Shoukry and S. Göbel, "Realizing a Mobile Multimodal Platform for Serious Games Analytics," *International Journal of Serious Games*, vol. 6, no. 4, pp. 19–48, Nov. 2019, doi: 10.17083/IJSG.V6I4.323.

[29] P. C. I. Jose, P. C. V. Manuel, M. O. Ivan, F. M. Manuel, and F. M. Baltasar, "Simplifiying serious games authoring and validation with uadventure and SIMVA," *Proceedings - IEEE 20th International Conference on Advanced Learning Technologies, ICALT 2020*, pp. 106–108, Jul. 2020, doi: 10.1109/ICALT49669.2020.00039.

[30] J. Santilario-Berthilier, A. Calvo-Morata, I. Martinez-Ortiz, M. Spaniol, and B. Fernandez-Manjon, "SIMVA: Serious Games Learning Analytics Based on Standards and Open Code," in *CEUR Workshop Proceedings*, 2025, pp. 241 – 250. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011072964&partnerID=40&md5=8167acfc74159ba24c8dba1e5d618f68

[31] J. Santilario Berthilier, A. Calvo-Morata, I. Martínez-Ortiz, and B. Fernández-Manjón, "Leveraging xAPI for Multimodal Learning Analytics Data Standardization in Serious Game Validation," pp. 137–155, 2026, doi: 10.1007/978-3-032-30781-1_9.

[32] A. Calvo-Morata, C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Creating awareness on bullying and cyberbullying among young people: Validating the effectiveness and design of the serious game Conectado," *Telematics and Informatics*, vol. 60, p. 101568, Jul. 2021, doi: 10.1016/J.TELE.2021.101568.

[33] A. Calvo-Morata, C. Alonso-Fernández, I. Martínez-Ortiz, and B. Fernández-Manjón, "Methodology for designing games: Articoding, a serious game to develop computational thinking through programming," *Revista de Educacion a Distancia*, vol. 26, no. 83, pp. 1–33, Jan. 2026, doi: 10.6018/RED.662331.

[34] A. Calvo-Morata, C. Alonso-Fernández, J. Santilario-Berthilier, I. Martínez-Ortiz, and B. Fernández-Manjón, "Learning Analytics to Guide Serious Game Development: A Case Study Using Articoding," *Computers*, vol. 14, no. 4, Apr. 2025, doi: 10.3390/COMPUTERS14040122.