This original article has been published in Telematics and Informatics with
DOI https://doi.org/10.1016/j.tele.2021.101583

Improving evidence-based assessment of players using serious games

Cristina Alonso-Fernández[a,*], Manuel Freire[a], Iván Martínez-Ortiz[a], Baltasar Fernández-Manjón[a]

[a]*Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid, Spain*

Abstract

Serious games are highly interactive systems which can therefore capture large amounts of player interaction data. This data can be analyzed to provide a deep insight into the effect of the game on its players. However, traditional techniques to assess players of serious games make little use of interaction data, relying instead on costly external questionnaires. We propose an evidence-based process to improve the assessment of players by using their interaction data. The process first combines player interaction data and traditional questionnaires to derive and refine game learning analytics variables, which can then be used to predict the effects of the game on its players. Once the game is validated, and suitable prediction models have been built, the prediction models can be used in large-scale deployments to assess players solely based on their interactions, without the need for external questionnaires. We briefly describe two case studies where this combination of traditional questionnaires and data mining techniques has been successfully applied. The evidence-based assessment process proposed radically simplifies the deployment and application of serious games in real class settings.

Keywords: Data science applications in education; Evaluation methodologies; Games; Teaching/learning strategies.

1. Introduction

Serious Games (SGs) are games that "do not have entertainment, enjoyment or fun as their primary purpose" (Michael & Chen, 2005). Digital serious games provide an engaging, highly interactive environment with many possibilities for causing an effect on players (Dörner, Göbel, Effelsberg, & Wiemeyer, 2016). SGs also present an opportunity for proactive learning by involving players/learners in an immersive learning experience where they can apply their knowledge, learn from experience, and test complex or risky scenarios in a safe environment. Due to these features, SGs have been successfully applied in varied domains such as medicine, the military, or complex processes training, among others; example success-cases include dealing

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

with phobias (Donker, Van Esveld, Fischer, & Van Straten, 2018), medical training (Boada, Rodriguez-Benitez, Garcia-Gonzalez, Thió-Henestrosa, & Sbert, 2016), and in-company training (Michael & Chen, 2005).

In SGs, a wide and diverse range of interactions can be tracked and analyzed to gain insight into their players' behaviors. Collection of interaction data is widespread in the games industry. It forms the basis of the field of Game analytics (GA), defined as the application of analytics for "game development and game research" aiming to provide "support for decision-making at all levels (…) from design to art, programming to marketing, management to user research" (El-Nasr, Drachen, & Canossa, 2013). Data collection has also been applied in education via Learning Analytics (LA), "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long, Siemens, Gráinne, & Gašević, 2011), mainly focusing on predicting students' success and providing feedback (Gašević, Dawson, & Siemens, 2015). The combination of GA and LA techniques in the context of serious games is called Game Learning Analytics (GLA) (Freire et al., 2016). GLA can help better understand the player game experience and characteristics, allowing the game design to be adapted and improved based on this player interaction evidence.

Assessment of players in serious games is usually performed with formal external measures, typically gathered via questionnaires, while Game Learning Analytics (GLA) data is rarely included in the assessment process (Alonso-Fernández, Calvo-Morata, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019). From our experiences evaluating SGs and assessing players using them, we propose a new approach to assess players using serious games based on GLA data. Once a game is validated using traditional external questionnaires, players' interactions in the game are analyzed to obtain representative GLA variables and prediction models that can be used to predict serious games' effect on players, as measured by the external questionnaires. For further deployments, the chosen prediction models can automatically assess players in a non-intrusive way, solely based on their interactions. This type of assessment becomes particularly useful in large-scale game applications in real domains, as it avoids potentially costly external assessment of players. The information gathered can also be used to improve and adapt the game to players' characteristics and provide targeted and personalized feedback.

We propose an assessment of players using serious games based on their game interactions: the evidences collected from their gameplays provide the game learning analytics data that can be used to reliably predict the effect of the game on players.

The rest of this paper is structured as follows: Section 2 reviews related work on serious games evaluation and the assessment of their players; Section 3 describes our evidence-based process for assessment of players using serious games based on the collection and analysis of game learning analytics data from their interactions to predict the game's effects; Section 4 briefly presents two case studies, with different serious games, where we have tested the previous process; Section 5 discusses our process and its limitations; finally, Section 6 summarizes the conclusions of our work.

2. Related work

Serious games are commonly validated with questionnaires (Calderón & Ruiz, 2015) where players are asked to complete formal external questionnaires both before (pre) and after (post) the gameplay. The results from both questionnaires are then compared, and the game is considered to be effective if the difference between their results is statistically significant. This widely accepted methodology has some drawbacks: first, the questionnaires must be previously validated to ensure that they provide a reliable measure of the

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

characteristics the game aims to change. Second, use of questionnaires significantly increases the effort and time in preparation, administration, and analysis for teachers or researchers who apply them. As a final drawback, the assessment is performed outside the game environment, breaking player immersion and requiring additional mechanisms to deliver the questionnaire, collect player responses, and link them back to their authors.

A data-based, or at least data-informed, evaluation approach, taking advantage of the potential of game learning analytics data collected from players' interactions, could provide more authentic and precise evaluation metrics. These metrics could be analyzed for deeper insight, both while the game is being played (in real-time), or after the gameplay is complete. Analyzing the data, patterns can be discovered, both from single-player interactions and by combining data from multiple users (Shoukry, Göbel, & Steinmetz, 2014). For instance, player profiles could be created based on their game preferences to help SG designers tailor their games or model players' exploration to understand their learning pathways better. Visual analytics can also help to gain insight into players' behavior (Wallner & Kriglstein, 2015) and provide teachers with real-time information via learning analytics dashboards, for instance, to assist players as they play (Charleer, Vande Moere, Klerkx, Verbert, & De Laet, 2017).

Evidence-based approaches are also being used to assess players, as different studies have started to investigate how their performance can be assessed directly from user-generated data, instead of relying on external questionnaires (Loh & Sheng, 2015). The field of *stealth assessment* (V. Shute, Ke, & Wang, 2017) aims to embed the assessment in a non-intrusive manner into the gaming environment. Assessment is then based on what players do in the game, as opposed to the use of immersion-breaking external questionnaires. This approach is based on the collection of specific data from players' gameplays, which are stored in log files. These collected high-level metrics are then analyzed and correlated with players' knowledge (Valerie J. Shute, Ventura, & Kim, 2013). The use of games for assessment has consequently drawn the attention of many recent research studies. In fact, the most common application of data science to game learning analytics data is that of assessing students, either to measure learning or to predict performance (Alonso-Fernández, Calvo-Morata, et al., 2019). The work of (Halverson & Owen, 2014) presents a model for game-based assessment that collects data from keystrokes and clicks, and identifies 15 moments in a game as evidences of learning to correlate with learning gains. The model is exemplified with a science game, for which results showed that the type of mistakes made were the best learning predictors, even more so than the number of times that players played, or the number of successes or failures experienced. Other studies propose the use of higher-level metrics obtained from in-game interactions such as learning observables (Serrano-Laguna, Manero, Freire, & Fernández-Manjón, 2017) or variables with aggregated learning analytics data (Alonso-Fernández, Cano, et al., 2019). Research has also been conducted on game-based assessment of specific 21st century skills, such as persistence (Dicerbo, 2013). From log data, researchers can identify players with specific goals and then create measures of persistence toward those goals, including information such as progress and time spent completing difficult tasks. Game-based assessments are being incorporated in other related fields. For instance, gamified applications for language learning are including machine learning to create computer-adaptive assessments (Settles & Laflair, 2020).

Despite these studies, research conducted on games as tools for assessment is still limited (Homer, Ober, & Plass, 2018). While few serious games for learning have been primarily built for assessment (Sliney & Murphy, 2011), several authors consider it important to include assessment as part of the design phase (Ifenthaler, Eseryel, & Ge, 2012). Another critical issue is that, so far, studies have focused on game-based assessments on a case-by-case basis. This results from each serious game having different goals, structure, and contents, therefore providing different opportunities for assessment. For each game, or type of game,

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

different kinds of interaction data will be available for collection, at different levels of granularity, and offering different evidences for assessment. Therefore, research is needed to shed light on how game features and categories contribute or detract to their validity from the point of view of assessment (Kato & Klerk, 2017). These features and characteristics that could relate to players' assessment will be tightly related to the game design and learning design. In the literature review of (Liu, Kang, Liu, Zou, & Hodson, 2017), authors found out that serious game features and metrics were primarily used for learner performance and game design strategies, and highlighted the need for more data-based research studies on this topic. Authors have carried out several studies trying to leverage the costly process of creating game-based assessments. The work of (Kim, Ruipérez-Valiente, Tan, Rosenheck, & Klopfer, 2019) proposes a process for game designers and developers to create games for educational assessment, balancing assessment needs and the gameplay experience through the phases of design, development and evaluation. This process comprises the collection of evidences from the game, the analysis of such evidences to create relevant variables, and the evaluation of the model.

The generalization of evidence-based players' assessments playing serious games is key towards their widespread use and future impact. For this, better data collection and sharing are vital parts of a continuous process to improve teaching and learning (V.J. Shute & Rahimi, 2017). For this process to be generalized, it seems essential to combine data collection and analysis with standard and systematic processes. Simultaneously, serious game designers, developers, and researchers need access to tools that can capture educationally relevant data from their games, and even more importantly, tools that can analyze collected data to yield insights into the progress and actions of players in those games. This information can then also be used for players' assessment.

We propose an evidence-based process to assess players using serious games, described in detail in the following section. Taking advantage of well-known data science techniques, our approach uses data collected during game validation to create models that can predict the effect of the game on its players. Use of this approach greatly simplifies game-based assessments, which are currently limited and conducted on a case-by-case basis, while retaining the advantages of evidence-based assessment. We have only tested this approach with narrative adventure videogames, but we consider the steps to be generalizable to other similar genres, such as geolocalized videogames with a narrative component.

3. Evidence-based assessment of players using serious games

Our proposed evidence-based assessment process combines in-game player interaction data with the traditional external data collected from questionnaires during formal game validations. Fig. 1 provides an overview of our proposed evidence-based assessment process. It comprises the following phases and tasks:

1. Game validation phase.
    a. Collect common player interactions using a standard and validated format (Section 3.1).
    b. Analyze the collected traces to choose an initial set of variables containing GLA information based on the Learning Analytics Model, and/or game designers' guidance, and refined through exploratory analysis and visualizations in our data science environment called *T-Mon* (see *feature extraction* process, Section 3.2).
    a. Use the selected variables as input of the prediction models to measure the impact of games on their players, and the pre-post questionnaires as target output variable to be predicted by

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

the models. Predictions are used to validate the models, and the predictive relevance of the variables used provides feedback to iterate the process, if needed (Section 3.3).

2. Game deployment phase.
    a. After the validation, large-scale deployments can be conducted where players are assessed based solely on their game interactions (Section 3.4).

3.1. Collecting players data: pre-post questionnaires and in-game interaction data

In our process, the first step is to collect the data to assess players with. For this purpose, we need to collect both pre-post questionnaires (or any other validated measure to be used as the target value for the predictions) as well as interaction data. Questionnaires should be formally validated by experts in the domain, to ensure that they provide a reliable measure of the characteristic that the game seeks to affect, such as awareness or knowledge.

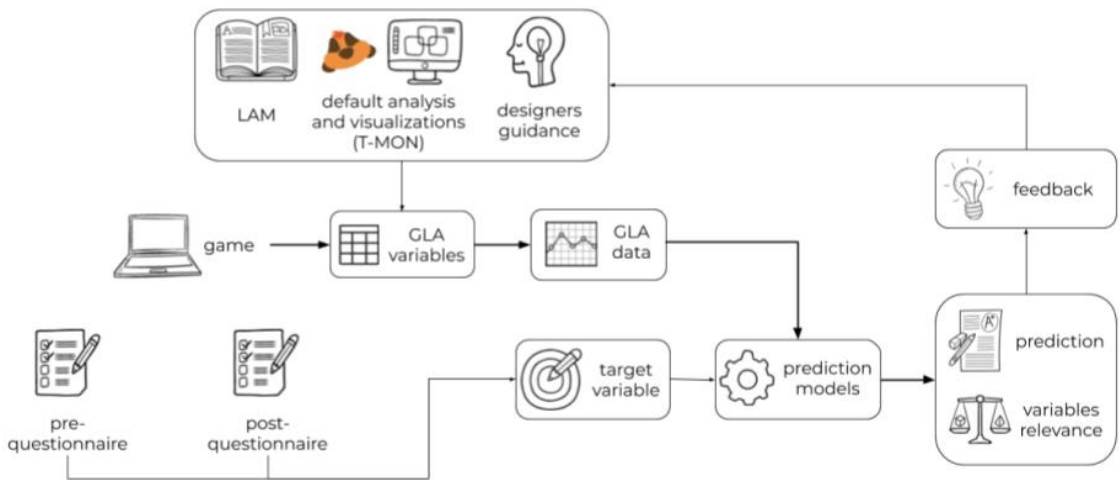

Fig. 1. Full process for evidence-based assessment of serious game players (tasks during game validation phase): the game interaction traces collected fill the pre-defined set of GLA variables to be used as input for the prediction models. The target variable used for prediction is based on pre-post results.

Although the type of data that can be collected from a serious game will depend on its content, structure and features, there are some common interactions in game analytics (GA) and learning analytics (LA) that can be extrapolated to serious games. GA data relates to game design and structure: number of clicks, avatar location in the game environment and characteristics, movements and changes of scenes or levels, items used, total time spent in the game, interactions with interface elements and non-player characters, points scored, in-game selections, and quest completions. These data can also be combined to derive "game metrics". As an example, by combining a "total time" and "points scored" for a given player, we can derive a "points per minute" metric [4]. GLA data relates to the learning design of the games, reflecting information about the learning progress and process of players/learners. Previous work on LA has identified multiple variables with strong predictive power, such as the interactions with elements of the learning environment (e.g. videos or activities) (Brinton, Buccapatnam, Chiang, & Poor, 2016; Ruipérez-Valiente, Cobos, Muñoz-Merino,

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

Andujar, & Delgado Kloos, 2017). When it comes to games, learning analytics data focuses on in-game player actions that affect learning.

To systematize the specific data to be collected from serious games, we propose the use of the validated and standardized *Experience API for Serious Games Profile* (xAPI-SG) (Serrano-Laguna, Martínez-Ortiz, et al., 2017). The Experience API describes an Application Programming Interface that allows e-learning content, such as serious games, to interact with a Learning Record Store (LRS) that stores the interaction data generated by the learning content. The xAPI-SG profile defines a common shared vocabulary and semantics for Serious Games. Software tools that support it need not be tied to specific SGs, thus allowing their users to reason on and compare data from multiple SG. Some of the commonly used interactions included in the xAPI-SG Profile are: verbs *initialized*, *progressed* and *completed* to measure completion and progress in so-called *completables* of types *serious-game*, *level* or *quest*; verbs *accessed* and *skipped* to collect changes in scenes of types *screen* or *area*; verbs *selected* and *unlocked* to track user-choices when confronted with a *question*, *menu* or *dialog-tree*; and *interacted* and *used* verbs to track interactions with *items* or *non-player-characters*.

The use of a standard data format such as xAPI-SG is a clear benefit when systematizing the collection of traces and their analysis to derive relevant information from user gameplays. Such formats facilitate the integration of tools from different providers and help to comply with personal data-protection laws: art. 20 of the EU GDPR requires data controllers to use a "structured, commonly used and machine-readable format" when users request access to their data, or transfer to other data controllers (European Commission, 2018). Additionally, while standard data collection formats are not commonly reported on the literature (Alonso-Fernández, Calvo-Morata, et al., 2019), their uptake would greatly assist in result replication and data sharing. Having a common interchange format also fosters the creation of a tool ecosystem created by different actors.

3.2. Extracting GLA variables from interaction data

Once the raw traces with user interaction data are collected, they can be analyzed to extract higher-level meaningful information about the actions of players within the game. Our process synthesizes the information available in the data traces (collected in xAPI-SG format) into a smaller set of GLA variables. Ideally, the definition of such variables should be described in the game's Learning Analytics Model (LAM) (Perez-Colado, Alonso-Fernandez, Freire, Martinez-Ortiz, & Fernandez-Manjon, 2018), cooperatively created by both educational experts and game designers. LAMs build on the game's learning design and game design, which define the educational goals of the game and how these are reflected on the specific game design choices taken depending on their educational goals. Based on both designs, a LAM determines the data to be collected from the game and how these data are to be analyzed into GLA variables and interpreted to provide meaningful information about the actions of a player in the game. It also may define any posterior visualization, feedback or reporting to do with the analysis results.

If such a LAM is not available, the game designers may suggest what information to obtain from the game and analyze it into GLA variables. Additionally, analysis and visualizations of collected xAPI-SG traces can provide important insights on the data collected and guide the choice of some GLA variables. For this purpose, we have created our data science environment called *T-Mon* (a trace monitor in xAPI-SG format). *T-Mon* contains a set of Python Jupyter Notebooks, available as open source at a GitHub repository[*]. *T-Mon* notebooks provide a default set of analyses and visualizations that can be applied to any given JSON file containing xAPI-SG traces: overall game progress; choices in alternatives, and if applicable, the fraction of

[*] https://github.com/e-ucm/t-mon

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

those considered correct and incorrect; progress, scores and times per game activity or subsection; content seen and skipped; and interactions with game items and areas and over time. The interactive interface allows to filter the data and configure the visualizations to gain a more in-depth insight into the data. *T-Mon* is intended both to provide quick overviews of collected data and to allow in-depth exploratory analysis to refine the choice of GLA variables that will be used in subsequent steps: the Jupyter Notebooks (Project Jupyter, 2020) *T-Mon* builds upon are a commonly used tool in data science to perform such analyses and provide access to an extensive and actively maintained collection of utilities to manipulate and explore data (Jupyter Team, 2020).

Table 1. Correspondence of xAPI-SG traces (object type, verb and other fields) to derive GLA variables.

xAPI-SG fields			**GLA variables**	
Object type	**Verb**	**Other fields**	**Name**	**Description**
Accessible: area, cutscene, screen, zone	Accessed	Object id	*Accessed_id*	Number of times the accessible *id* has been accessed
	Skipped	Object id (cutscene)	*Skipped_id*	Number of times the cutscene *id* has been skipped
Completable: serious-game, level, quest	Initialized	Object id, timestamp	*Duration_id*	Duration of completable *id* (calculated in combination with *completed* trace of same *id*)
	Progressed	Object id, result progress, timestamp	*Progress_id_time*	Progress in completable *id* per timestamp *time*
	Completed	Object id	*Completed_id*	True if completable *id* has been completed
		Object id, timestamp	*Duration_id*	Duration of completable *id* (calculated in combination with *initialized* trace of same *id*)
		Object id, result score	*Score_id*	Score obtained in completable *id*
Alternative: question, dialog-tree, menu	Selected	Object id (question), result success	*Correct_id*	True if question *id* has been successfully answered
		Object id (dialog), result response	*Response_id*	Response selected in dialog *id*
		Object id (menu), result response	*Selection_id*	Option selected in menu *id*
Target: non-player character, enemy, item	Interacted	Object id	*Interactions_id*	Number of interactions with target *id*
	Used	Object id (item)	*Uses_id*	Number of uses of item *id*

A further advantage of using xAPI-SG to collect data is that, since xAPI-SG is designed to model and report on essential structures and concepts found in serious games, those structures and concepts are likely to yield a right choice of initial GLA variables. Table 1 proposes a non-exhaustive set of pre-defined GLA variables for each player that can be easily derived from any set of traces that follow the xAPI-SG Profile. Such variables include the number of interactions with each in-game object and character (count of *interacted* traces per object), or the duration of each level/game (difference in timestamp of *completed* and *initialized* traces per object of type serious-game or level). We are currently working on extending the xAPI-SG Profile to include more precise definitions of the required fields in each trace (using *statement templates*) and the required sequence of traces (using *patterns*) to clarify the expected traces to extract such GLA variables. While game-specific variables as specified in a LAM or suggested by expert designer knowledge are of course preferable, a set of ready-to-use generic variables can be highly useful to complement game-specific

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

variables, and allows the use of our process even when no LAM or designers are available. They also constitute a good starting point for refinement using *T-Mon*.

Once the GLA variables have been chosen, they can be used for player modelling, as they can provide a rich insight into players' actions. Once enough data have been collected, richer information may be obtained via a wide variety of algorithms and techniques, for purposes such as assessment or adaptation. Supervised, unsupervised, or reinforcement learning techniques can be applied to the derived variables. In our case, we use supervised techniques (Devin Soni, 2007) to predict the serious game's effect based on the information found in GLA variables derived from user interactions.

3.3. Creating the prediction models with GLA evidences

The next step is to create the prediction models to accurately measure the effect of the game on its players. By default, we define such effect, which will be the target variable for our models, as the improvement between the scores of the pre- and post- questionnaires, caused by playing the serious game. If we were only interested in measuring the final effect on players after playing, the post questionnaire score alone could be used as the target variable. The prediction models use the previously defined GLA evidences, filled with the data captured during the game validation, as input data.

To consider a serious game effective in educational scenarios, it first needs to be validated, ideally using a formal validation process. We use the formal validation step to create the prediction models that will be used in the deployment phase. The formal validation of the serious game is commonly performed with pre-post questionnaires. The comparison of the results between both questionnaires should, ideally, show a statistically significant difference between pre-questionnaire and post- questionnaire. If such difference is significant, we consider that the game is experimentally validated. During these experiments, we also collect relevant game learning analytics data from players' in-game interactions. With both questionnaires and GLA data, we can create the prediction models that will be used for game effect assessment during the deployment phase. The prediction models take as input the GLA data from players' interactions and predict the improvement (difference between pre- and post- questionnaire results). This process is experimental and can be iterated until accurate-enough models are created, by changing and refining the GLA variables according to their relevance as reported by the results of the prediction models. In our experiments, we have found accuracies above 90% to be achievable, and suggest this figure as a workable goal. Once an accurate-enough level is reached, the final prediction model is retained for the next step of deployment, where it will be used for automatic non-intrusive assessment of players.

For the specific prediction models to be tested, an increasingly broad and varied range of options is available. At least in the first iterations, we recommend using interpretable models (Adadi & Berrada, 2018) that provide information about the relevance of the input variables towards the predictions. This will provide feedback about the importance of specific GLA variables (and, therefore, about users' interactions), allowing us to improve the process before moving to the deployment phase. Linear and tree-based prediction models are a simple baseline to start from. More complex models may improve the results: for instance, ensemble methods based on trees, such as random forest or gradient boosting. These complex models could provide more precise results while still giving feedback about how relevant the input variables are towards the prediction results. The models may then be reused and adapted for different contexts. Traces can be re-examined to generate additional GLA variables or change existing ones based on variable relevance as reported by such models.

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

The creation of prediction models relies on both questionnaires and interaction data. By collecting player interaction data while the serious game is formally validated, prediction models can be trained with the same questionnaires that are used for formal serious game validation. Suppose the results obtained in the traditional formal-validation questionnaires show a significant improvement on players. In that case, the serious game is formally validated – and models can be built immediately, without the need for further experiments. Instead of only proving the game's efficacy in the chosen educational scenario, we have also built a set of prediction models to be used for students' assessment, and identified a subset of user interaction data that is to be collected from the game in order to make assessment predictions.

The number of users to include in this validation phase is not clear, but considering the reported number of users in other data-based research on serious games (Alonso-Fernández, Calvo-Morata, et al., 2019), we recommend including at least 100 users. The information gathered during the validation phase can also be used to improve the game, if the data shows behaviors that do not align with the game design or learning design. The game can be adapted based on players' characteristics, for example, by creating player profiles based on their game behaviors, and then specifically adapting the game to each profile. Collected data can also be used to provide more targeted or personalized feedback to help players progress in the game. The resulting game, updated with features and personalization based on the feedback from the previous round, would then be subjected to another validation phase, leading either to further iterations or to a fully validated game.

3.4. Assessing players in large-scale game deployment

Once the serious game has been formally validated, the deployment phase can start, with the game applied in classrooms and other real-world educational settings. To be able to gather information from users' experience and to assess them based on their interactions, this application should include the collection of data from relevant interactions. The deployment process for large-scale scenarios reads as follows:

1. Students access the SG and play the game from beginning to end. We have used anonymous identifiers that allow only teachers to de-anonymize student data to ensure that privacy requirements are met while still linking questionnaire responses to each student's game-interaction data.
2. A tracking component integrated in the game sends the relevant traces generated from player interactions to the analytics tool while students are playing. The user interaction traces should follow a well-defined format (for instance, the xAPI-SG Profile), as required by the analytics tool that will receive it.
3. The analytics tool takes interaction data as input, uses it to fill the pre-defined GLA variables, and uses them as input to the previously created prediction models, to derive prediction outputs for the students' assessment.
4. Once students have finished their gameplay, teachers will receive the predicted score based on each students' in-game interactions, possibly together with another analytics information. They can then use this information, together with any other evaluation of their own, to obtain the final students' assessment.

Note that the prediction models provide the assessment output for students once they have finished playing the game, and therefore once all the input data required by the models is available.

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

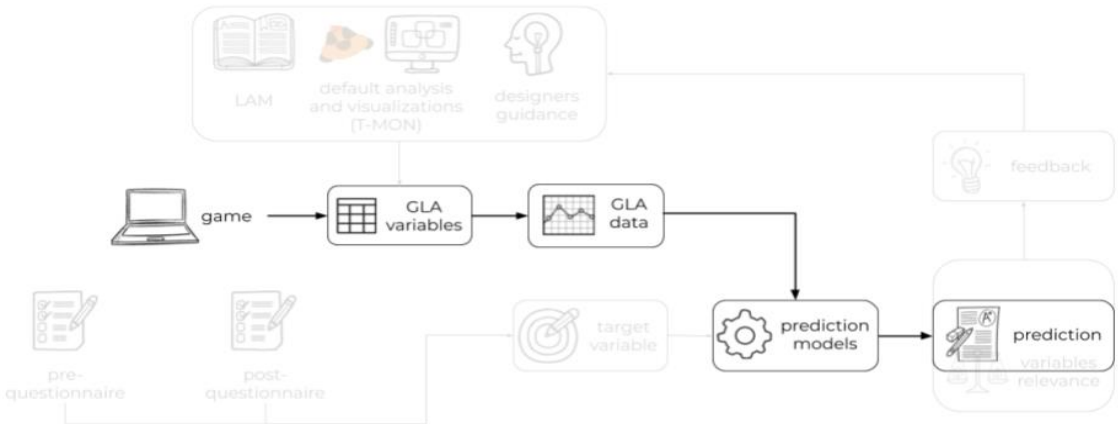

Fig. 2. Full process for evidence-based assessment of serious game players (tasks during game deployment phase): the GLA data derived from game interaction traces is input to the prediction models to obtain a measure to assess players. External questionnaires are no longer required.

The assessment obtained with this process is therefore automatic and non-intrusive, and simplified from both ends: teacher preparation and execution times typically required for post-game assessment are reduced, and students will simply play a game without the added time, disruption, and pressure of completing the questionnaires. Game-based assessment can also provide institutions and managers means of evaluating the efficacy of games for education and simplify the assessment process of their students. The previously used pre-post questionnaires are no longer required during the large-scale deployment, which simplifies the application of SGs in real-world larger settings. This allows students to play the game for longer periods, and/or teachers to include additional activities related with the gameplay (e.g. discussion, post-game questions), instead of the traditional student assessment. Fig. 2 depicts the deployment phase of a serious game using our process, once questionnaires are no longer required.

4. Case studies

We summarize two case-studies that we have carried out to test the previous process with two different serious games. For each case-study, we briefly describe the game goal and main mechanics, the interaction data captured (corresponding to the step of the process described in section 3.1), the analysis of those data traces to derive GLA variables (corresponding to section 3.2), and the prediction models created to assess players automatically based on their game interactions (section 3.3) in the consequent deployment phase. Additionally, the case studies have different target variables to test the full process when predicting after-game performance (case study 1) or increase in the measured characteristic (case study 2).

4.1. First Aid Game, a serious game to teach first aid techniques

The *First Aid Game* is a serious game developed to teach first aid maneuvers to teenagers (Marchiori et al., 2012). The game presents three different levels, each one depicting a different medical emergency. In each level, players can choose among several courses of action (presented as textual or visual options to select

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

from) to assist the in-game character during the emergency. Multiple interactions with the in-game character are available, as are several in-game tools, such as a defibrillator or a smartphone with which (simulated) emergency services can be contacted. After each level is completed, a score for the level is provided as user feedback, based on the errors made and their relevance. Levels can be replayed to improve the score and, consequently, knowledge of its contents.

We collected all relevant interactions in the game using the verbs and activity-types of the xAPI-SG standard: interactions with in-game elements (character and items) were traced with *interacted* traces with the trace object corresponding to the specific element; selections in multiple-choice situations and questions were collected as *selected* traces with the corresponding object and the result indicating whether the response was correct or not; *initialized* and *completed* traces were used to track start and end times of the game and each level, and scores in game levels were tracked in the result extension of the *completed* traces of the corresponding level.

The xAPI-SG traces were then analyzed to derive GLA variables. Some variables were directly obtained following the xAPI-SG based analysis given in Table 1: game completion, the number of interactions with specific in-game elements, and whether specific multiple-choice questions or situations were failed or not. We defined additional variables based on the game designers' decision, as stated in the LAM, that levels could be replayed: first and maximum scores achieved in each game level, and the number of times that each level was repeated.

Next, we used the GLA variables to predict knowledge of first aid techniques (the ones covered in the game) after playing (Alonso-Fernández, Martínez-Ortiz, Caballero, Freire, & Fernández-Manjón, 2020), both as a binary pass/fail category, and as the exact knowledge score. The *First Aid Game* had already been validated in a previous experiment, so we knew the game was effective. We tested different prediction models taking as input the GLA data. Highly accurate results were obtained predicting after-game performance as given in the post-questionnaire. The prediction model that provided the best results was a logistic regression (achieving 90% precision, 98% recall, and 10% misclassification rate), which also allowed the results to be interpreted, and provided a measure of the predictive relevance of each variable. The two most relevant variables turned out to be the scores in the first level played, and the total number of interactions with the in-game character. Based on the LAM, the relevance of these two variables seems to be due to the specific strategies followed by players when playing (trial-and-error, exploratory, …), and their overall engagement during the game – while final playthrough scores were surprisingly not that useful when predicting actual learning.

4.2. Conectado, a serious game to raise bullying and cyberbullying awareness

Conectado is a serious game created to raise awareness about bullying and cyberbullying (Calvo-Morata et al., 2020). The game presents a narrative story where players take on the role of a student during the first week at a new high school. Players then experience a bullying and cyberbullying situation in first person during five in-game days. Players can interact with different in-game characters (classmates, teacher, and parents) and in-game objects, including several electronical devices (mobile phone, computer) on which cyberbullying takes place. Some player's choices, related to talking to the parents and/or teachers about the situation, determine which of the three endings is reached.

Again, we captured relevant interactions following the xAPI-SG format, whose verbs and activity-types sufficed to represent them: *interacted* traces were used to collect all relevant interactions with the game

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

characters and items; *selected* traces tracked choices made in decisions and conversations with other characters; *accessed* traces were used to track scene changes; and *initialized* and *completed* traces informed about the starts and ends of the different game days. The ending reached was encoded in the result extension of the *completed* trace for the full game.

From these xAPI-SG traces, we again derived a set of GLA variables, following the xAPI-SG based analysis of Table 1: count of interactions with each character and item, times each of the game areas was accessed, total time spent in each part of the game, and specific decisions in the game, such as conversational choices and those affecting the ending. We also used an additional variable based on the game LAM: the specific ending reached.

The final set of GLA variables was then used as input for models to predict an increase in bullying and cyberbullying awareness due to playing *Conectado* (Alonso-Fernández, Calvo-Morata, Freire, Martínez-Ortiz, & Fernández-Manjón, 2020), given by the difference in pre-post questionnaires scores. The prediction model that obtained the best results was a Bayesian regression model (achieving 0,54 mean absolute error with 0,053 standard deviation, both normalized to scale [0-10]), which also provided insights into the influence of specific variables on the predictions. In this case, the most significant variables were those representing the specific in-game ending reached, as well as two variables that reported on the time spent in two in-game days. Spending time in one of these days was correlated with greater awareness, while spending additional time in the other exhibited a negative correlation. Again, we examined the design of the game to understand these results. A better ending reached predicted higher awareness increase, indicating that those players who examined a better behavior in the game also learned more. The positive-correlation day was partially spent dealing with an in-game episode of identity theft. At the same time, the negative-correlation one included an interaction where players could attempt to defend themselves against cyberbullying in a simulated in-game social network. However, since reactions to any player comments were scripted, spending time in the in-game social network acted as an indicator of players becoming distracted away from the main plot. These results showed the importance of aligning the game content to the target group age and interests, something that is especially relevant in a game about (misuse of) technology targeted at technology-savvy teenagers.

5. Discussion

The proposed evidence-based process for assessing players using serious games is based on collecting in-game interaction data, as well as evaluation questionnaire results during the game validation to create the prediction models. We have provided a set of guidelines for the critical steps of choosing interaction data to collect and of deriving useful GLA variables from that data. This process is significantly simplified by using a standard format to represent the data (e.g. the xAPI-SG Profile), as a good baseline set of variables with game learning analytics information can be easily derived from interaction traces using this standard (see Table 1). We have additionally provided a set of analyses and visualizations to be applied to any given set of xAPI-SG traces for further insight using our trace monitor, *T-Mon*, which is freely and openly available at our GitHub repository. *T-Mon* also allows exploratory analysis using well-known data-science tools, allowing GLA variable candidates to be identified and evaluated. Additional game-specific variables may be more informative but require in-depth knowledge of the game, or even better, access to the game's LAM.

During the game validation process, both pre-post questionnaires and interacted data are collected, so the prediction models can be created and validated. Note that, although in our approach we have focused on the commonly used pre-post questionnaires, other instruments could also serve as long as they are validated. The GLA information obtained could also serve to model players, improve the game's evaluation process, and

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

adapt it to users' characteristics; even while students are still playing, the partial information collected could be used to adapt the game to players' progress and needs. With the information extracted from GLA data, a deeper insight can be gained about the game design (Loh, Sheng, & Ifenthaler, 2015). The analysis of such data could also support the design of more effective learning environments (Liu, Li, Pan, & Pan, 2019). Game validations could also be improved as, for instance, the interaction data could discover players not taking the questionnaires seriously and, this way, such questionnaires could and should be discarded from the validation process (Rowley, 2014).

We consider that the integration of automatic non-intrusive player assessment in serious games can greatly simplify and increase their application in education. Game-based assessment provides multiple benefits but still presents several limitations and drawbacks, including the use of immersion-breaking questionnaires to verify whether students are learning or not. Since the step of formally validating a game is essential before deployment to prove that it works as intended, our process is based on re-using data from the validation step to create valid and accurate prediction models for assessment. Once built, those models can be used during the deployment phase for game-based assessment, providing an automatic means of predicting students' knowledge using data science techniques. This way, the actual large-scale deployment of serious games in real settings can be greatly simplified, as game-based assessments can be obtained without the need to conduct pre-post questionnaires. The predicted knowledge can then be used directly for assessment, or as an additional data-point for the teachers or trainers.

5.1. Limitations

Our process has some limitations. A first limitation is that GLA variables created solely from traces represented using the xAPI-SG Profile (see Table 1) may ignore important game-specific details that could lead to more accurate predictions. However, we consider that these variables provide a good baseline on the information that can be extracted from any SG. Additionally, if accuracy of predictions is not deemed to be high enough, re-analysis of the traces to build better GLA variables is certainly possible – and we have developed *T-Mon* to make such re-analysis much more accessible. Note that it is possible to reuse the non-profile specific aspects of *T-Mon* as a starting point to process other statements that are compliant with the xAPI standard.

A second limitation is that prediction models must be created ad-hoc for each serious game based on their formal validation process; and are only valid for players that are sufficiently similar to those it was validated with. Although this means that our process is not generalizable to all kinds of serious games, we consider that it is generalizable to games that share the same genre or mechanics, as similar games can be expected to report similar interaction data, amenable to similar analysis and likely to yield similar results. In a related context, some authors have pointed out that the choice of variables has a greater impact than that of prediction models (Gardner & Brooks, 2018); therefore, the baseline set of variables proposed following the xAPI-SG standard, possibly with additional game-dependent variables, is expected to yield accurate-enough results. Quality of results will be tightly linked to the quality of the games and the selection of variables, which is in turn driven by their Learning Analytics Models, which define how the game and learning designs map to collected interaction data. To ensure the validity of the whole process, it is essential that all games undergo the validation step where prediction models are created, and their validity is tested.

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

5.2. Conclusions

Serious games have proven to be a useful tool for learning. Availability of a systematic and generalized processes to assess their players would greatly increase serious games' applicability and adoption in real settings, such as schools. It is also time for teachers, educators, and institutions to start trusting these educational tools for assessment purposes. Embedding automatic non-intrusive assessment directly into the game experience provides several advantages, such as reducing both costs and students' stress towards paper-based formal evaluations, while also reducing costs in terms of both time and effort for teachers. Research on automatic assessment using machine learning techniques is been conducted in similar fields, for instance, in language learning through gamified applications (Settles & Laflair, 2020).

Our evidence-based process to assess players using serious games based on tracked in-game interaction data can provide a general standards-based process for other assessment. First, a standard format such as the xAPI-SG Profile can represent the most common interactions present in serious games. Then, their analysis to yield a default set of GLA variables as described on this paper, including their refinement in *T-Mon* starting from the default analysis and visualizations, provides a baseline of the information that can be extracted from any serious game. With the resulting GLA variables, interpretable prediction models (Carvalho, Pereira, & Cardoso, 2019) can be tested during the game validation phase and, in the deployment phase, used to assess players. If possible, we recommend building a Learning Analytics Model early on during the game design, thus ensuring that representative data will be captured, and will therefore be available for later analysis, ready to yield information with educational value (Ke & Shute, 2015). Availability of a LAM can then be used to inform better choices for GLA variables and build better player models or to adapt and personalize games based on the characteristics of their players.

The full lifecycle of serious games is considered in our proposal. Games undergo a first validation phase to prove their efficacy as learning tools, while game learning analytics data is also collected. At the end of this phase, prediction models are created and validated based on in-game interactions and external pre-post questionnaires. When games move to the deployment phase in actual educational settings, the prediction models created can be used to provide a (predicted) score as assessment for each student/player. This way, the cost and time required to deploy the games is greatly reduced, as questionnaires are no longer required to assess students. Use of our process also greatly simplifies large-scale deployment of serious games in real settings and provides a deep insight into the learning experiences of its players.

Acknowledgements

This work has been partially funded by the Regional Government of Madrid (eMadrid P2018/TCS4307, co-funded by the European Structural Funds FSE and FEDER), by the Ministry of Education (TIN2017-89238-R), by the European Commission (Erasmus+ IMPRESS 2017-1-NL01-KA203-035259), by MIT-La Caixa (MISTI program, LCF/PR/MIT19/5184001) and by the Telefónica-Complutense Chair on Digital Education and Serious Games. Also thanks to Julio Santilario Berthilier and Ana Rus Cano for their contribution to Conectado and T-Mon.

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, 103612. https://doi.org/10.1016/j.compedu.2019.103612

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2020). Evidence-based evaluation of a serious game to increase bullying awareness. *Interactive Learning Environments*, 1–11. https://doi.org/10.1080/10494820.2020.1799031

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior*, *99*, 301–309. https://doi.org/10.1016/j.chb.2019.05.036

Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., & Fernández-Manjón, B. (2020). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning*, *36*(3), 350–358. https://doi.org/10.1111/jcal.12405

Boada, I., Rodriguez-Benitez, A., Garcia-Gonzalez, J. M., Thió-Henestrosa, S., & Sbert, M. (2016). 30 : 2: A Game Designed to Promote the Cardiopulmonary Resuscitation Protocol. *International Journal of Computer Games Technology*, *2016*, 1–14. https://doi.org/10.1155/2016/8251461

Brinton, C. G., Buccapatnam, S., Chiang, M., & Poor, H. V. (2016). Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance. *IEEE Transactions on Signal Processing*. https://doi.org/10.1109/TSP.2016.2546228

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education*, *87*, 396–422. https://doi.org/10.1016/j.compedu.2015.07.011

Calvo-Morata, A., Rotaru, D. C., Alonso-Fernandez, C., Freire-Moran, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2020). Validation of a Cyberbullying Serious Game Using Game Analytics. *IEEE Transactions on Learning Technologies*, *13*(1), 186–197. https://doi.org/10.1109/TLT.2018.2879354

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, *8*(8), 832. https://doi.org/10.3390/electronics8080832

Charleer, S., Vande Moere, A., Klerkx, J., Verbert, K., & De Laet, T. (2017). Learning Analytics Dashboards to Support Adviser-Student Dialogue. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2017.2720670

Devin Soni. (2007). Supervised vs. Unsupervised Learning •. Retrieved from https://www.kdnuggets.com/2018/04/supervised-vs-unsupervised-learning.html

Dicerbo, K. E. (2013). Game-based assessment of persistence. *Educational Technology and Society*, *17*(1), 17–28.

Donker, T., Van Esveld, S., Fischer, N., & Van Straten, A. (2018). 0Phobia - towards a virtual cure for acrophobia: Study protocol for a randomized controlled trial. *Trials*. https://doi.org/10.1186/s13063-018-2704-6

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (Eds.). (2016). Serious Games. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-40612-1

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

European Commission. (2018). 2018 reform of EU data protection rules. Retrieved from https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

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

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*. https://doi.org/10.1007/s11257-018-9203-z

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Halverson, R., & Owen, V. E. (2014). Game-based assessment: an integrated model for capturing evidence of learning in play. *International Journal of Learning Technology*, *9*(2), 111. https://doi.org/10.1504/ijlt.2014.064489

Homer, B. D., Ober, T. M., & Plass, J. L. (2018). Digital Games as Tools for Embedded Assessment. In *The Cambridge Handbook of Instructional Feedback* (pp. 357–375). Cambridge University Press. https://doi.org/10.1017/9781316832134.018

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for Game-Based Learning. In *Assessment in Game-Based Learning* (pp. 1–8). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3546-4_1

Jupyter Team. (2020). Jupyter Projects. Retrieved November 1, 2020, from https://jupyter.readthedocs.io/en/latest/projects/content-projects.html

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

This original article has been published in Telematics and Informatics with DOI
https://doi.org/10.1016/j.tele.2021.101583

Ke, F., & Shute, V. J. (2015). *Serious Games Analytics*. (C. S. Loh, Y. Sheng, & D. Ifenthaler, Eds.), *Serious Games Analytics*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4

Kim, Y. J., Ruipérez-Valiente, J. A., Tan, P., Rosenheck, L., & Klopfer, E. (2019). Towards a Process to Integrate Learning Analytics and Evidence-Centered Design for Game-based Assessment. *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, (March), 8–10.

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

Liu, M., Li, C., Pan, Z., & Pan, X. (2019). Mining big data to help make informed decisions for designing effective digital educational games. Interactive Learning Environments. https://doi.org/10.1080/10494820.2019.1639061

Loh, C. S., & Sheng, Y. (2015). Measuring Expert Performance for Serious Games Analytics: From Data to Insights. In Serious Games Analytics (pp. 101–134). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4_5

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.

Michael, D. R., & Chen, S. L. (2005). Serious Games: Games That Educate, Train, and Inform. *Education*, *October 31*, 1–95. https://doi.org/10.1145/2465085.2465091

Perez-Colado, I., Alonso-Fernandez, C., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2018). Game learning analytics is not informagic! In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1729–1737). IEEE. https://doi.org/10.1109/EDUCON.2018.8363443

Project Jupyter. (2020). Jupyter. Retrieved November 1, 2020, from https://jupyter.org/

Rowley, J. (2014). Designing and using research questionnaires. *Management Research Review*, *37*(3), 308–330. https://doi.org/10.1108/MRR-02-2013-0027

Ruipérez-Valiente, J. A., Cobos, R., Muñoz-Merino, P. J., Andujar, Á., & Delgado Kloos, C. (2017). Early Prediction and Variable Importance of Certificate Accomplishment in a MOOC. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 263–272). https://doi.org/10.1007/978-3-319-59044-8_31

Serrano-Laguna, Á., Manero, B., Freire, M., & Fernández-Manjón, B. (2017). A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. *Multimedia Tools and Applications*, *77*(2), 2849–2871. https://doi.org/10.1007/s11042-017-4467-6

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

Settles, B., & Laflair, G. T. (2020). Machine Learning Driven Language Assessment, *8*, 247–263.

Shoukry, L., Göbel, S., & Steinmetz, R. (2014). Learning Analytics and Serious Games: Trends and Considerations. In *Proceedings of the 2014 ACM International Workshop on Serious Games*. https://doi.org/10.1145/2656719.2656729

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1–19. https://doi.org/10.1111/jcal.12172

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton's Playground. *The Journal of Educational Research*, *106*(6), 423–430. https://doi.org/10.1080/00220671.2013.832970

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

Sliney, A., & Murphy, D. (2011). Using Serious Games for Assessment. In *Serious Games and Edutainment Applications* (pp. 225–243). London: Springer London. https://doi.org/10.1007/978-1-4471-2161-9_12

Wallner, G., & Kriglstein, S. (2015). Comparative Visualization of Player Behavior for Serious Game Analytics. In *Serious Games Analytics* (pp. 159–179). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-05834-4_7