This document is a draft of the article accepted for publication in the Special Issue "Analytics for Game-Based Learning" of the Journal of Learning Analytics to be published in November 2022.

Game Learning Analytics: Blending visual and data mining techniques to improve serious games and to better understand players learning

Cristina Alonso-Fernandez[1], Antonio Calvo-Morata[2], Manuel Freire[2], Ivan Martinez-Ortiz[2], Baltasar Fernandez-Manjon[2]

Abstract
Game Learning Analytics comprise the collection, analysis, and visualization of players' interactions with serious games. The information gathered from these analytics can help both to improve serious games and to better understand players' actions and strategies, as well as to improve players' assessment. However, the application of analytics is a complex and costly process that is not yet generalized in serious games. To systematize the application of game learning analytics in serious games, the use of a standard data format to collect players' interactions is essential: the standardization allows to simplify and systematize every step (from the definition of the interaction data collected, to the creation of result visualizations) developing tools and processes compatible with multiple games. In this paper, we explore a combination of (1) an exploratory visualization tool that analyzes players' interactions in the game and provides an overview of their actions, and (2) an assessment approach, based on the collection of interaction data for players' assessment, to better understand players' strategies in serious games and improve games during their lifecycle. Both approaches are based on a standard data format, the xAPI-SG Profile that simplifies the definition and collection of interaction data and guides the creation of visualizations, as well as the information to be used for players' assessment. This combination is tested through a case example with a serious game to teach first aid maneuvers to high school players. The visualization tool for players' interactions unveils players' behaviors in the game establishing possible links between their actions and their learning outcomes that are later confirmed with the assessment process results. With the present work, we describe some of the different opportunities offered by analytics in game-based learning, the relevance of systematizing the process by using standards and game-independent analyses and visualizations, and the different techniques (visualizations, data mining models) that can be applied to yield meaningful information to better understand learners' actions and results in serious games.

Notes for Practice (research paper)

- Game Learning Analytics collect, analyze, and visualize interactions in serious games to provide information about players' actions for different stakeholders.
- The Experience API for Serious Games Profile is a standard data format to collect players' interactions with serious games. Using the standard simplifies definition and collection of interaction data, guides creation of visual feedback and helps integration with other systems.
- The combined approach starts with a visual tool that allows to explore the interaction data collected. The visually hinted hypotheses are then accepted or rejected with the results of an evidence-based assessment process.
- Relevance of players' actions for learning is obtained analyzing prediction results of white-box models, whose results accuracy are sufficient in the context of learning in serious games.

Keywords
Serious games, game learning analytics, game-based learning, stealth assessment, visualization.

Submitted: 10/11/12 — **Accepted:** 10/11/12 — **Published:** 10/11/12

Corresponding author [1] Email: cristina.alonsof@uam.es Address: School of Engineering, Autonomous University of Madrid, Madrid, Spain.
[2] Email: {toni, manuel.freire, imartinez, balta}@ucm.es Address: Computer Science faculty of Complutense University, Madrid, Spain.

1. Introduction

Serious Games (Dörner et al., 2016) are videogames that address issues beyond entertainment: their main purpose may be to raise awareness about social issues or change players' behaviors, although the most common purpose is that of games for learning (educational games).

Game Learning Analytics (Freire et al., 2016) comprise the collection, analysis and visualization of players' interactions with serious games (Figure 1). Visual analytics and dashboards are commonly used to provide different stakeholders with information about players' actions in the games. The information gathered from these analytics can help both to improve serious games and also to better understand players' actions and strategies, as well as to improve players' assessment (Shute, Rahimi, et al., 2021).

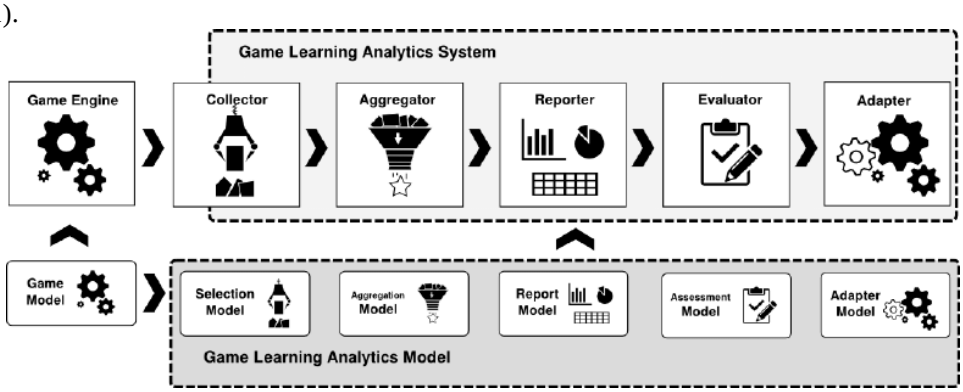

Figure 1. Game Learning Analytics system.

However, the application of learning analytics is a complex and costly process that is not yet generalized in serious games. The inclusion of analytics is usually created *ad-hoc* for each scenario/game, following no standard or systematized processes. In some cases such analytics are not included for educational purposes; that is the case of telemetry, part of the Game Analytics applied in games in general (not only serious games), that aims to help with the design and development process or to increase revenue of the games (Seif El-Nasr et al., 2013).

In this paper, we present a case study to exemplify a combined approach to systematize the proccess and to better understand players' learning using: (1) a visual exploratory tool of the interaction data; and (2) an evidence-based assessment process. The presented case study is a simple narrative game about first aid techniques that allows to easily follow all steps of the analysis process.

The rest of this paper is structured as follows: Section 2 reviews some literature on game learning analytics, visualization analytics, and assessment in serious games; Section 3 presents the methodology including the serious game used in the case study, the research questions and the combined approach, using a standard data collection format xAPI-SG, and the exploratory visualization tool T-MON together with an evidence-based assessment approach; Section 4 describes the results of the combined approach in the case study; Section 5 discusses the results, and their implications; and finally, Section 5 summarizes the main conclusions of our work, and points out its limitations and some future lines of work.

2. Literature review

Numerous studies have stated the positive effect of educational games on learning outcomes in various fields (Yu et al., 2021). Particularly, role-playing games tend to be used in research and education, to immerse students in each narrative. The scenarios present in these narrative games allow multiple chances for interactions that can then be collected as game metrics, such as: interactions with game elements, paths followed, game times, and scores (Tlili et al., 2021). Different data analytics have been applied to the collected data with three main goals: obtaining information about players' learning process; determining differences between individual players; and assessing learning (Tlili & Chang, 2019). Data visualization, data mining, and sequential analysis are the analytics techniques most applied in educational games (Tlili & Chang, 2019).

To help the different stakeholders (mainly educators and students) better understand the collected learner data, Learning Analytics dashboards are commonly used (Khosravi et al., 2021). As a communication tool, dashboards are used to project learning progress and outcomes, identify learning difficulties, and promote engagement and collaboration (Liu et al., 2021). For serious games, visualization of collected gameplay traces is also crucial to understand player strategies (Nguyen et al., 2015) and behaviors (Eagle et al., 2013); and to identify game areas that are confusing and may need refinement (Andersen et al., 2010). The information depicted in such visualizations can also help developers and researchers to develop new hypotheses as they obtain a better understanding of the collected data (Eagle et al., 2013), while real-time visualizations can help teachers to better track their students' learning progress (Minović & Milovanović, 2013).

Assessment is another common field of application of analytics in serious games, for instance, in disciplines such as stealth assessment (Shute, 2011). Using evidence trace files (Hao & Mislevy, 2018) and other techniques to capture learners' actions in the game, those actions can be analyzed to identify meaningful play patterns in order to provide evidences for player assessment (Andrews-Todd et al., 2021).

To systematize the application of game learning analytics in serious games, standardization simplifies the definition of the interaction data collected, the creation of results visualizations, and the development of tools and processes compatible with different games. The Experience API (xAPI) is a standard to collect data from learning environments (ADL, 2012), that is soon to have a new version released as an IEEE open official standard (ADL, 2021). In the standard, each learning activity is captured as a JSON-based trace (statement) with fields containing the information about the learning activity (verb), the target of the activity (object) and who performed it (actor). Statements may include additional fields such as a timestamp indicating when the action occurred, and the context of the action. For particular domains, xAPI allows to create application-specific profiles. The Experience API Profile for Serious Games (xAPI-SG) is a standard data format to collect players' interactions within serious games (Serrano-Laguna et al., 2017), and due to its simplicity and completeness to capture all relevant actions in our games, it is the standard used in our approach.

To further democratize the Game Learning Analytics process, the data collection format and the means to communicate the formatted data (e.g. through an API) need to be considered so that they are not entirely dependent on a specific infrastructure (Pérez-Colado et al., 2021). Following a standard data format, default game-independent visualizations and dashboards can be provided for different games. For instance, by using commonly available fields such as scores or gameplay times, which can be represented with the fields available in standard formats like xAPI. Such analyses and visualizations could be created by default as they do not require any specific information about particular games. Game-dependent analyses and visualizations are still possible, and can provide richer feedback regarding player actions, but require a much higher investment to create and maintain those analyses.

3. Methods

3.1. Game sessions

The First Aid Game (Marchiori et al., 2012) is a game-like simulation that aims to teach step-by-step procedures in three emergency situations (unconsciousness, choking, and chest pain) depicted as game levels (Figure 2, left). In each scenario, the game allows you to choose among several options, presented as visual or text, stating whether the option is correct or not (Figure 2, right). During the game levels, short videos show the correct way to execute procedures and provide explanations to better learn the techniques. Such videos are also available at the end of each level as a summary in case players want to watch them again. A score is given at the end of each level based on the errors made and their relevance. Players can replay levels to improve their scores. Possible interactions include choices in multiple-choice situations and questions, as well as with the different game items: the victim suffering from one of the conditions, a phone to call the emergency services, and a defibrillator sometimes available (randomly) to help the character. Questions in the game include, for instance: "Where should you place your head?" (to check if the victim is breathing, with options given visually in the game) or "What is the number to call emergency services?" (with possible answers given in text fields). The game was validated in 2012, and the use case presented in this section corresponds to a new experiment.

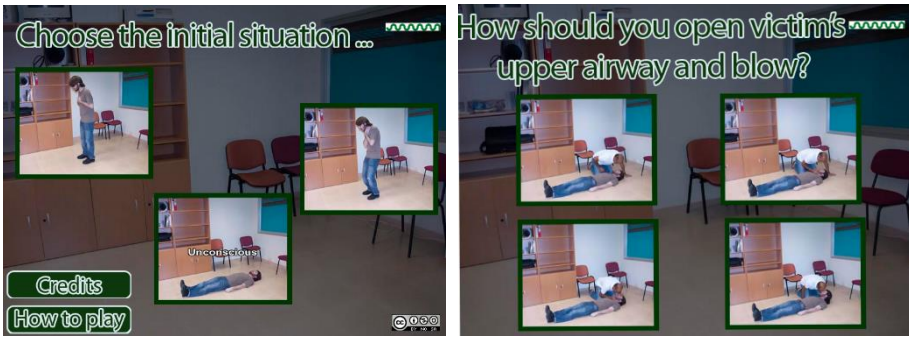

Figure 2. First Aid Game captures including the three game levels / situations (left) and visual multiple-choice questions (right).

The data used for the case study corresponds to a set of N=112 players, ages 12-16. Sessions were carried out in January-February 2017, in a school in Madrid, Spain, and lasted 50 minutes. In that time, players completed the pre-post questionnaires about the first-aid content of the game, and played the game repeating game levels as many times as wanted during the available time. All data collection complied with the ethics (Carter & Egliston, 2021) and regulations (e.g. GDPR (European

Commission, 2016)) including pseudo-anonymization on source, by using randomly assigned tokens for players to access the game. No personal information was gathered from players, and the collected data cannot be traced back to specific students.

3.2. Research questions

In the case study, we aim to apply our proposed combined approach to the First Aid Game, exploring particularly the following player behaviors:

- **Game and level duration**. We want to study players' sessions to ensure that all gameplays fit within the duration of an average lesson. Additionally, we want to study how long each player stays in each of the three game levels.
- **Questions difficulty**. We are interested in the questions that were most difficult for players and wish to compare them to those that were easier to answer.
- **Items interactions**. We want to study the interactions made by players with the different game items.
- **Video behavior**. We want to explore players' behavior with the in-game videos: whether they watch them or not, and any differences in behavior depending on the type of videos shown.

Based on such behaviors, we propose the following research questions:

RQ1. How long did players take to complete the game and each level?
RQ2. Which are the easiest questions? Which are the most difficult ones?
RQ3. What is the most interacted-with game element?
RQ4. What are players' behaviors related to videos?

3.3. Combined approach

The combined approach proposed aims to better understand players' strategies in serious games and improve the full lifecycle of games. The approach is based on the standard data collection format xAPI-SG, which simplifies the definition and collection of interaction data and guides the creation of visualizations, as well as the identification of the possible GLA variables to be used for players' assessment. After collecting the data, the approach has two main steps: an initial visual analysis of the interaction data carried out with an exploratory visualization tool, called T-MON, and a more in-depth analysis of the GLA variables and their prediction relevance with a data-mining evidence-based assessment approach. The process of the combined approach is described in more detail in the following sections, describing the: (1) data collection process of interactions in the game using the xAPI-SG standard format (section 3.2.1); the use of the visualization tool T-MON to explore the collected xAPI-SG traces to create hypotheses about players' results, actions and their effect on players' learning, as well as GLA variables for later analysis (section 3.2.2); and the evidence-based assessment approach, using the GLA for prediction models whose results provide the evidences about the effect of players' actions towards learning, to accept or reject the proposed hypotheses (section 3.2.3).

3.3.1. Data collection in xAPI-SG

The xAPI-SG Profile is the data standard used to collect players' interactions. The Profile defines a set of verbs, activity types and extensions that comprise most common interactions within serious games. This way, such interactions can be tracked and described following the format and vocabulary available in the xAPI-SG standard profile. xAPI-SG statements contain three main fields: the *actor* who performed the action, *verb* capturing the action itself and the *object* that receives the action. Traces typically contain a *timestamp* to capture when the action occurred. Additional fields can be included as extensions of the traces. If needed, it is also possible to extend the default available traces including ad-hoc fields for better describing other specific user interactions in serious games. An example xAPI-SG statement can be seen in Figure 3 stating that "John Doe (*actor*) has completed (*verb*) a programming course (*object*)".

The xAPI-SG Profile defines the following key concepts to related verbs with their corresponding types:

- **Completables:** each *level* or the full *serious game* can be *initialized*, *progressed* and *completed*. Possible extensions may capture the exact *progress* in *progressed* traces, or the final *score* in *completed* traces.
- **Alternatives:** *questions* or *menus* can have specific options *selected* or *unlocked*. The exact choice taken may additionally be included as extension in the traces and, for *questions*, whether the response is correct or not (in a *success* field).
- **Accessibles:** game *areas* or *videos* can be *accessed* and *skipped*.
- **Game objects:** *enemies* and *non-playable characters* can be *interacted* with, while *items* can also be *used*.

```
{
  "actor": {
    "name": "John Doe",
    "mbox": "mailto:johndoe@example.com"
  },
  "verb": {
    "id": "http://adlnet.gov/expapi/verbs/completed",
    "display": { "en-US": "completed" }
  },
  "object": {
    "id": "http://example.com/activities/programming-course",
    "definition": {
      "name": { "en-US": "Programing course" }
    }
  }
}
```


Figure 3. Example xAPI-SG statement.

3.3.2. Traces visualization with T-MON

T-MON (xAPI-SG Traces Monitor) is an exploratory visualization tool that analyzes players' interactions in the game and provides a default set of visualizations with an overview of their actions. The tool takes as input data a JSON file with the raw xAPI-SG traces collected from players' interactions with the game. T-MON then analyzes such xAPI-SG statements creating a set of game state metrics for each player that gets updated with the following statements of the player. That information per player encapsulates all relevant actions, taking the necessary information from the fields of the xAPI-SG traces, and includes: if the whole game (*serious-game*) has been started; if it has been completed; all interactions with items/NPCs/enemies/etc.; the progress in the whole game per time (given in an extension field *progress*, up to a maximum of 100); for each completable, including the whole game, the last score (given in an extension field *score*), the last progress achieved, the times of start and ending; all responses in alternatives, and whether they are correct or not (given in an extension field *success*); all accessibles accessed; all videos seen and skipped; and all selections in menus.

Once all data has been analyzed, T-MON displays a default set of visualizations summarizing the received data. The output visualizations included in T-MON can be created just by processing the xAPI-SG statements generated by the game, with no further game-dependent information (e.g. game structure, game elements, etc.). The visualizations displayed include information about game and levels progress and completion, including scores and duration, accessibles accessed and skipped, selections in alternatives (including success), and interactions with game items. For instance, Figure 4 depicts some of the default visualizations available in T-MON: the left-most line chart presents the progress of players over time in the game, calculated by taking the extension field *progress* and the timestamp of *progressed* traces; the middle figure presents maximum and minimum completion times per *completable*, calculates by taking the difference between timestamps in *initialized* and *completed* traces for each *completable*; the final bar chart present correct and incorrect responses per player, calculated by taking the selected traces in questions, and the extension field *success* to determine whether they were correct. Some of these visualizations relate to the ones provided in LRSs (Learning Record Stores), for instance, those which show player activity over time (similar to progress in the case of games) or completed objects (completables in the case of the xAPI-SG Profile).

Currently, T-MON is implemented as a set of Python Jupyter Notebooks, therefore providing an entry-point for data scientist to work with game analytics without requiring them to have an extensive background on the game development or the data collection standard. T-MON is available freely as open code on GitHub (e-UCM, 2020).

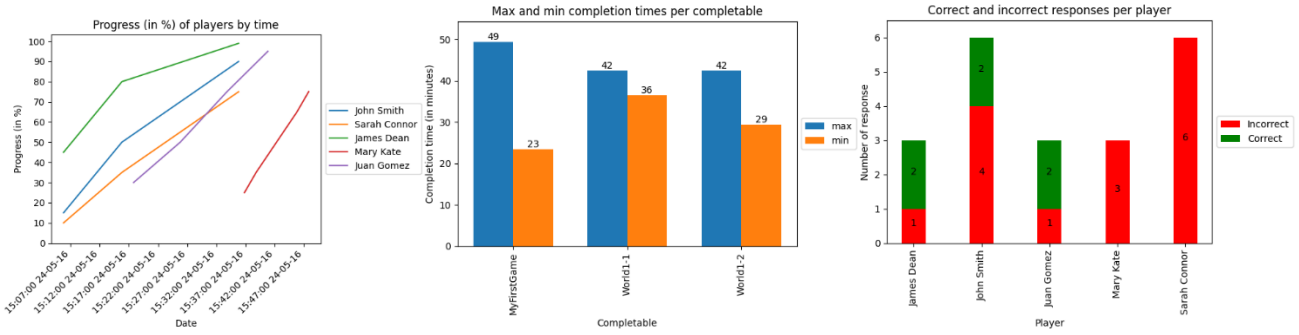

Figure 4. Some of the default visualizations included in T-MON (from left to right): progress of players over time, maximum and minimum completion times in completables, and correct and incorrect answers per player.

For the combined approach, T-MON provides an initial overview of the data, and hints about some possible players' behaviors, establishing links between their actions and learning outcomes studying issues like questions with higher rate of failure, items with higher number of interactions, levels scores, progress and completion, etc. Those outcomes can be useful

for some game stakeholders, for instance, for game developers to improve the quality of the final game. But they can also be used to inform more advanced analyses. With this visual information, different hypotheses can be suggested about players' results and the relevance of players' behaviors in the game towards their learning. Once those hypotheses are stated, we can move to the following step to carry out the assessment approach.

3.3.3. Evidence-based assessment approach

The evidence-based assessment approach builds from the collected interaction data from gameplays to create relevant variables for players' assessment. The collected xAPI-SG statements are analyzed to create GLA variables encapsulating the information collected from players' interactions. The derived GLA variables are then used as input to different prediction models. Results of predictions are analyzed to extract the most relevant variables, and their relevance for predictions.

The prediction models require some additional data as the target variable to compare predictions with. Traditionally, this is accomplished with the pre-post questionnaires used in formal serious games validation, which measure the actual players' state (i.e. knowledge, awareness) before and after the game (Alonso-Fernández et al., 2021). As such, if we want to predict players' learning outcome, the score obtained in the post-questionnaire will be our target variable in the prediction models.

For the combined approach proposed, the results of the assessment can be analyzed to accept or reject the previously stated hypotheses. This can be achieved based on the obtained relevance of the GLA variables towards the assessment predictions. That is, the prediction results will indicate the player behaviors in game (encapsulated in those created GLA variables) that have a greater impact on learning, therefore, providing the required evidence to test the hypotheses.

The combined approach described could be iterated, creating new hypotheses, and further accepting or rejecting them. In the following section, we test the approach through a case study with a simple serious game for which we explore four different players' behaviors with T-MON to create the corresponding hypotheses, and then conducting a single iteration of the assessment approach to test their validity.

For the case study, we selected an already validated game that is simple enough so that all steps in the combined approach can be easily followed, but that it also contains enough features so that interesting results may be obtained. We have already started to apply similar approaches to other serious games in different domains (Alonso-Fernández et al., 2019).

4. Case study results

This section describes the results of the case study to exemplify the combined approach with the First Aid Game: the data collected from the game (4.1), and the steps of the combined approach with T-MON (4.2), and the evidence-based assessment process to obtain the results (4.3).

4.1. Data collection

The interaction data from the N=112 players playing the First Aid Game was captured using the xAPI-SG Profile. All interactions with the game were captured:

- For each of the three game levels and for full game, the start and ending were captured using *initialized* and *completed* traces. The *score* in each level was captured as an extension of the *completed* traces for each level.
- Progress in each game level and the full game was captured with *progressed* traces, with the specific *progress* as an extension field.
- All interactions with game *items* (phone, defibrillator) and *non-playable-characters* (the in-game victim) were captured with *interacted* traces and the specific item or character.
- Answers in in-game *questions* (whether those were presented in text, or as visual options) were captured with *selected* traces, with two additional extension fields: *response* to store the specific option selected, and *success* to store whether that option was the correct one.
- Access to *menus* and *videos* were stored with *accessed* traces. For the case of *videos*, it was also collected whether players *skipped* them.

A total of 24354 xAPI-SG traces were collected, which translates into ~215 data traces per player.

4.2. T-MON

As explained in the combined approach, the visualization tool T-MON analyzes players' interactions (given in the xAPI-SG statements), unveiling players' behaviors in the game and helping the researchers to establish possible links between the actions of players and their learning outcomes. In each of the following subsections, we focus on one of the previously stated questions, and explore the default visualizations available in T-MON related to that question, trying to obtain some visual information to answer them. Note that through T-MON it is possible to analyze the collected data globally (considering all players) or to drill down to a specific player or set of players.

4.2.1. Game and level duration: RQ1. How long did players take to complete the game and each level?

The progress of players is shown in T-MON (Figure 5). T-MON allows for view in relative time, where, for each player, 0

represents the instant when they started to play; therefore, it is possible to compare all session lengths. Notice that this visualization is game-independent, created by simply taking the timestamp of all traces related to progress in the complete serious game. As seen in Figure 5 (left), it seems that faster players finished in around 5-6 minutes, while slower players took up to 18 minutes to complete the game. We additionally notice some players who may have entered and exited the game without progressing (bottom line of Figure 5). This graph shows all their replays in the same progress line, so each actual replay may be even shorter than 5-6 minutes. Most players finished the game in 8-12 minutes (it was expected that players would replay the game in the class session to improve their final scores and to learn to save the patient efficiently).

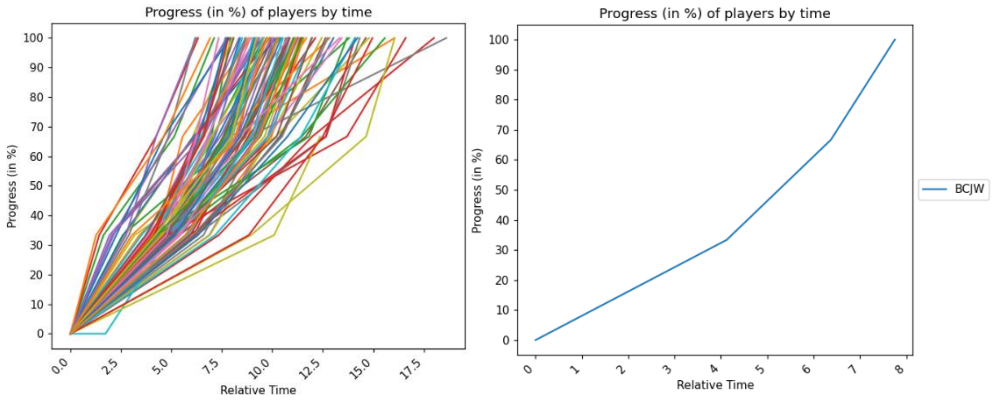

Figure 5. T-MON cropped screen shot visualizations: line chart of progress over time for all players (left) and filtered for a single player (right).

In another visualization (Figure 6), we can further explore the maximum and minimum completion times for the complete game (right-most bars) and the three game scenarios (unconsciousness, choking, and chest pain) for all players. This visualization is also game-independent, created by taking the specific completables' names used in each game: here, the 3 game scenarios, and the complete game. We verify now that shortest gameplays were less than 4 and a half minutes. Although total gameplays varied between 4 and 18 minutes, individual levels were often under 1 minute, due to much faster replays, which may indicate that players had already learned the procedures.

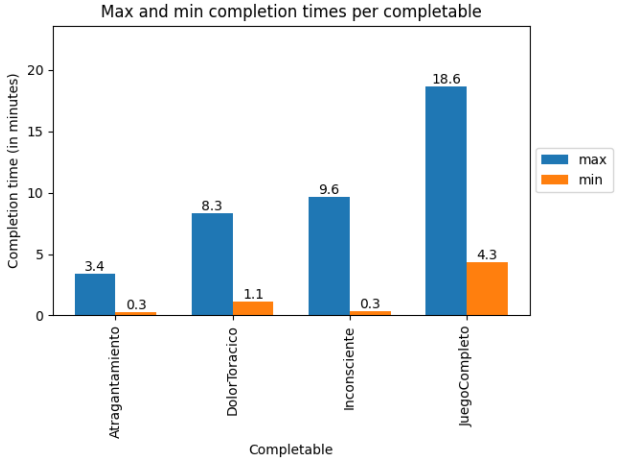

Figure 6. T-MON cropped screen shot visualization: bar chart with maximum (blue-left) and minimum (orange-right) completion times per completable.

To verify if players indeed carried out faster replays, we checked that in the collected interaction traces: for level "choking", out of the 44 players who repeated the level at least once (i.e. played it at least twice), 40 players constantly decreased their level completion times in all subsequent attempts, 3 players lowered their completion times in some replay but also increased their completion times in other replay(s), and only 1 player took longer to complete the level in replays than in the first attempt; for level "chest pain", out of the 70 players who replayed the level, 59 lowered their completion times in all attempts, while 8 had both shorter and longer completion times in replays, and only 3 had strictly longer times in replays; finally, for level "unconsciousness", out of the 93 players who replayed the level, 56 players decreased their completion time in all attempts, while 34 players completed attempts with longer and shorter completion times than previous attempts, and only 3 players took more time in replays. Therefore, we can conclude that, as suspected, players tended to carry out faster replays. Also, it should

be noted that a longer time in the first attempt may also be influenced by learning to use the videogame and to learn its mechanics.

From this question and visualizations, we see a variety of players' durations in the different levels and the full game. Considering the almost-linear structure of the game, in all 3 game levels, shorter gameplays (under 5 minutes) possibly point out to replays, when players are no longer taking the needed time to read in detail all questions/options in the game. A longer time in the game may lead to players better learning the content (even from repeating after making some errors and taking time to reflect before choosing an option), therefore, we propose the following hypothesis:

- H1: Players who took longer to complete game levels / the game learned more.

4.2.2. Questions difficulty: RQ2. Which are the easiest questions? Which are the most difficult ones?

One of the default visualizations in T-MON shows the correct (in green) and incorrect (in red) number of responses per alternative (i.e., question) for all players. This game-independent visualization (Figure 7) is created by considering all alternative names used in the game.

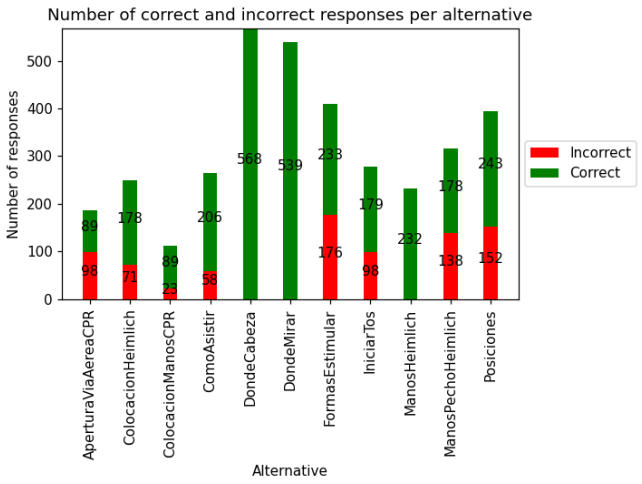

Figure 7. T-MON cropped screen shot visualization: bar chart with correct (green-top) and incorrect (red-bottom) number of responses per alternative.

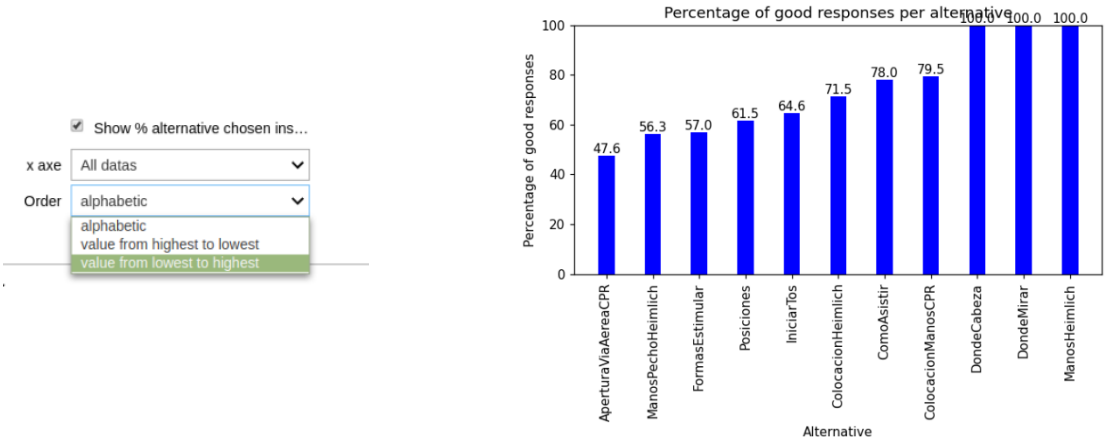

Figure 8. T-MON cropped screen shot visualization: configuring visualization in Figure 7 by using the available options (left), bar chart with alternatives ordered from lowest to highest correct response rate (right).

For better understanding of the visualization, we can make two changes in the available configuration options in T-MON (Figure 8, left): (1) use percentages (by activating a checkbox in T-MON UIs), and then sort from lowest to highest correct response rate. The result after making these configuration changes is shown in the right part of Figure 8.

With these changes, we can see that the three easiest questions (100% success rate) correspond to the positions of the hands in Heimlich maneuver, and where to place the head and where to look to verify if the victim is breathing. On the other hand, the question with lowest success rate (47.6%) corresponds to how to open the airway when performing CPR.

For these questions and visualizations, we see generally high success rates in questions, but some of them had an important percentage of failure (30-40%), and even some of them had a success rate below 50%. If players select the correct answer in

the game (although this could be from previous knowledge), we speculate that they are learning more with the game, therefore, we propose the following hypothesis:

- H2: Players who answered more questions correctly / answered less questions incorrectly learned more.

4.2.3. Items interactions: RQ3. What is the most interacted-with game element?

T-MON displays in another visualization the number of interactions of each player with each in-game element, providing a heatmap of such interactions (Figure 9). This visualization is game-independent, created by considering all game object names interacted with in the game, and all actors' names. (For greater clarity, Figure 9 shows a subset of 20 players not to expand too much horizontally for the same items). "Victima" (the in-game victim character, shown in the last row of the heatmap) was the element with the highest number of interactions for almost all players.

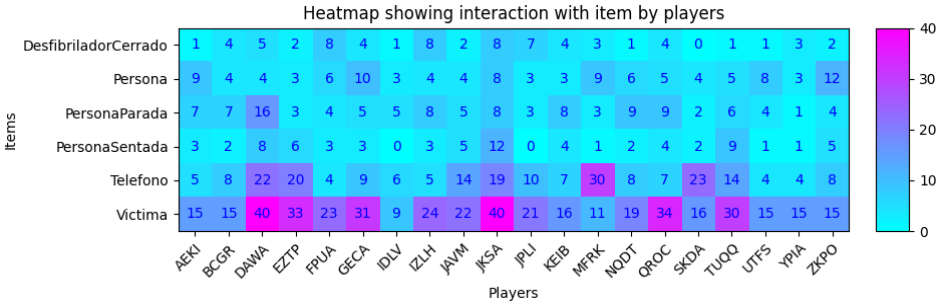

Figure 9. T-MON cropped screen shot visualization: heatmap with number of interactions of players (x-axis) with the different game items (y-axis), darker colors represent a higher number of interactions.

We can further see the number of interactions with particular items for all players in a different set of visualizations. For instance, for the most interacted-with element, the in-game character, we can see that players' number of interactions ranged from 7 to 89, while most players interacted an average of 10-20 times (Figure 10, ordered from highest to lowest values). While for the telephone available in the game, interactions ranged from 3 to 34 (Figure 11).

From this question and visualizations, we see a high number of interactions with game items, particularly with the in-game victim (as expected, as there are some mandatory interactions with it to advance in the game). A higher number of interactions may also be the result of repeating the game levels, having in doing so more changes to learn. Therefore, we propose the hypothesis:

- H3: Players who interacted more with items learned more.

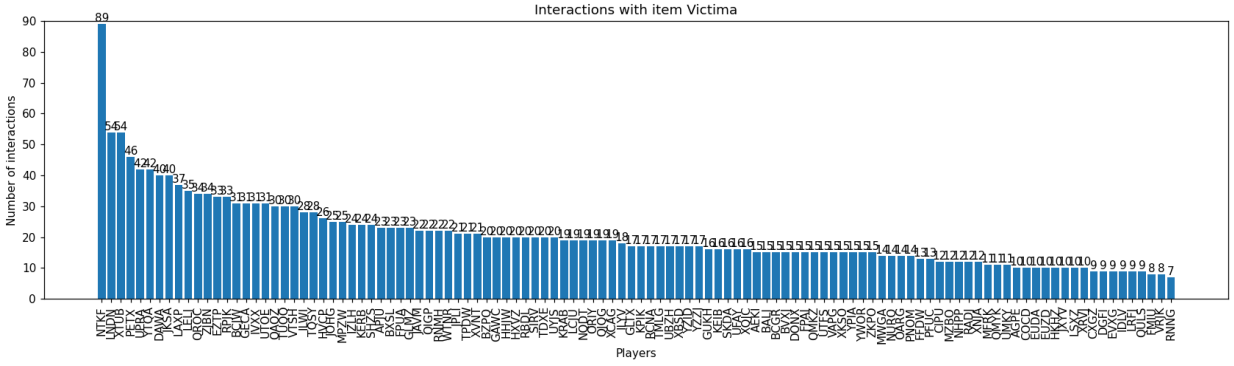

Figure 10. T-MON cropped screen shot visualization: bar chart with number of interactions of players with item "Victima" representing the in-game character.

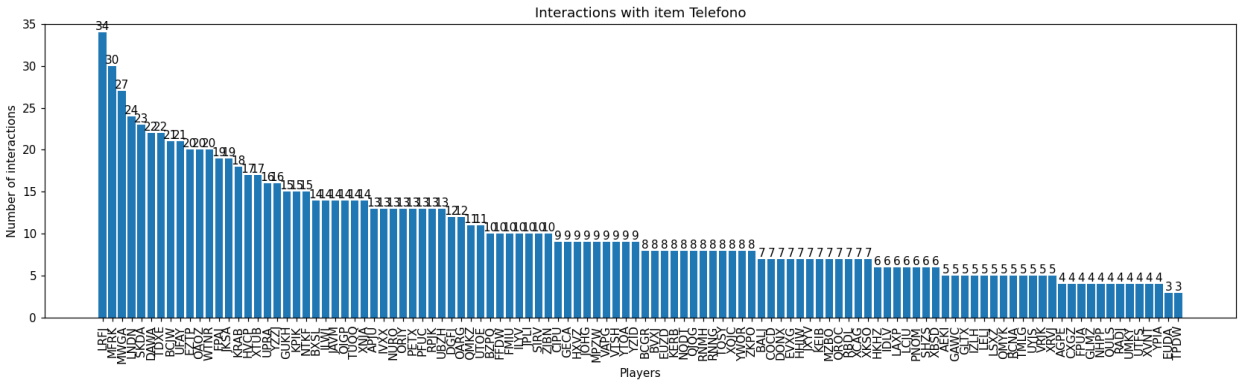

Figure 11. T-MON cropped screen shot visualization: bar chart with number of interactions of players with item "Telefono" representing the in-game phone.

4.2.4. Video behavior: RQ4. What are players' behaviors related to videos?

For this exploratory question, we analyze the visualization in T-MON showing the number of times each video has been seen and skipped (Figure 12). This game-independent visualization takes all accessibles (videos) in the game.

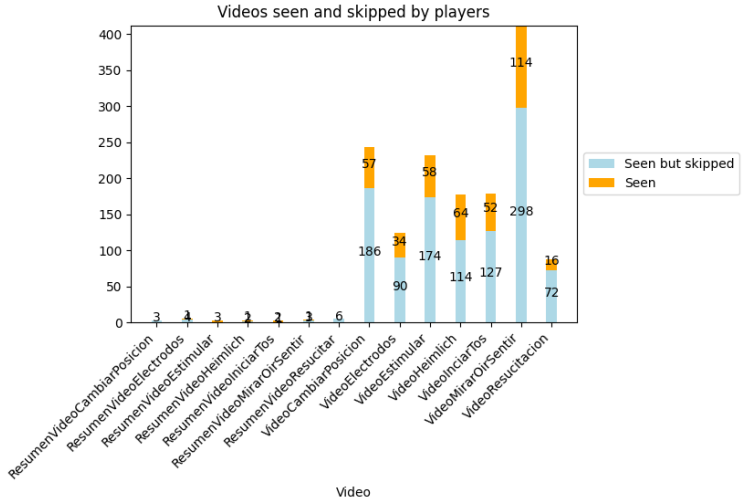

Figure 12. T-MON cropped screen shot visualization: bar chart with number of times videos (a type of accessibles) have been seen (orange-top) and skipped (blue-bottom).

From the visualization, we notice two different eye-catching behaviors.

- Lots of videos starting with "Resumen" (Spanish for "Summary", left part of the plot) are almost never accessed

Summary videos were shown at the end of attempts - when they were not that relevant any longer. Therefore, players did not access those videos.

- Why so many skipped videos (shown in blue in the plot)?

As players could play each level multiple times, improving scores as they did so – and skipped videos on successive playthroughs. They had already seen videos, so they skipped them in following attempts. The most skipped videos can be seen ordering the visualization (Figure 13).

From this visualization, we see that players watched most videos many times, but also skipped them a lot. As videos provide more information about the content to be learned from the game, we propose the hypothesis that:

- H4: Players who watched more videos / skipped less videos learned more.

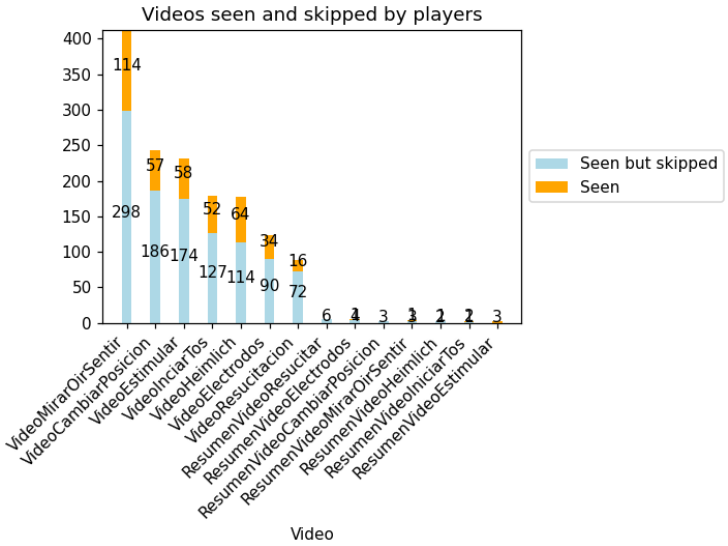

Figure 13. T-MON cropped screen shot visualization: configuring visualization in Figure 12, bar chart showing videos ordered by the most skipped ones from left to right.

4.3. Evidence-based assessment approach

After exploring the data with T-MON, we have proposed four hypotheses regarding the possible relationships between players' actions and behaviors in the game and their learning results. To accept or reject those hypotheses, we analyze the data with the evidence-based assessment approach. For that, we first need to analyze the xAPI-SG statements gathered from players' interactions. Such interactions need to be summarized into a set of GLA variables to be used as input for the prediction models.

To construct the GLA variables, we use some of the features available in most games (e.g., scores, timestamps to obtain durations), as well as game-dependent information (e.g., the three game levels available in the game), and the consequent combinations of both (scores per game level, duration per game level). We additionally base our creation of GLA variables on the visual information displayed in T-MON, sometimes taking simple specific xAPI fields (the field *success* in alternatives to know if the response given in a specific question is correct or not) and other times performing simple operations like aggregations (count of interactions per item, count of videos seen and skipped).

In particular, the xAPI-SG statements from the First Aid Game are analyzed and summarized into a default set of 29 GLA variables per player encapsulating the most relevant information about: game completion, scores, and level replays (11 variables as a baseline of information), maximum and minimum duration in game levels (6 variables to address H1), failure in specific questions and correct and incorrect responses (7 variables to address H2), items interactions (3 variables to address H3), and videos seen and skipped (2 variables to address H4). The full list of GLA variables created can be seen in Table 1.

As stated in section 3.3, the approach uses questionnaires that are also processed to obtain the target variable for the prediction models. Note that the questionnaires were also validated (Marchiori et al., 2012). A score is given for both questionnaires, counting the number of correct responses in each questionnaire, with a maximum possible score of 15. We have chosen the difference between post-questionnaire and pre-questionnaire scores, a continuous variable, as the target variable to predict; and have called it "learning". Table 2 provides some descriptive statistics about pre-questionnaire scores, post-questionnaire scores, and learning (difference between both scores).

The prediction models tested include some traditional simple models, that provide information about the relevance of the variables in the results (white-box models) as well as some other more complex ones to compare with (black-box models). The prediction models tested are linear regression, regression trees, Bayesian ridge regression, support vector machines for regression (SVR), k-nearest neighbors (kNN), and neural networks.

All the analysis for the evidence-based assessment approach was carried out in Python, using Pandas and Numpy for data processing and mathematical funcions, and Scikit-learn for all machine learning models.

Following the description of the combined approach, we tested the previously listed prediction models using as input data the created GLA variables (Table 1) and predicting the variable "learning" created with the questionnaire data. Data preprocessing included looking for and deleting any null values, verifying the target variable's skewness (its distribution was approximately symmetric, skewness=0.09 so no transformation was performed), shuffling data rows, and scaling input data for models that require it (kNN, support vector machines and neural networks). All models were tested with 10-fold cross validation, and with different parameters, to obtain the combination that provided best results.

Table 1. List of GLA variables derived from the First Aid Game interaction data.

Hypothesis	Variable name	Type
Basic information	Game completed	Binary (True, False)
	Score in complete game	Numerical [0-10]
	Maximum score in game level "Unconsciousness"	Numerical [0-10]
	Maximum score in game level "Choking"	Numerical [0-10]
	Maximum score in game level "Chest pain"	Numerical [0-10]
	First score in game level "Unconsciousness"	Numerical [0-10]
	First score in game level "Choking"	Numerical [0-10]
	First score in game level "Chest pain"	Numerical [0-10]
	Replays of game level "Unconsciousness"	Integer
	Replays of game level "Choking"	Integer
	Replays of game level "Chest pain"	Integer
H1	Minimum time to complete game level "Unconsciousness"	Numerical [0-50] minutes
	Minimum time to complete game level "Choking"	Numerical [0-50] minutes
	Minimum time to complete game level "Chest pain"	Numerical [0-50] minutes
	Maximum time to complete game level "Unconsciousness"	Numerical [0-50] minutes
	Maximum time to complete game level "Choking"	Numerical [0-50] minutes
	Maximum time to complete game level "Chest pain"	Numerical [0-50] minutes
H2	Failure in question "Emergency number"	Binary (True, False)
	Failure in question "Abdominal thrusts"	Binary (True, False)
	Failure in question "Heimlich maneuver name"	Binary (True, False)
	Failure in question "Heimlich maneuver position"	Binary (True, False)
	Failure in question "Heimlich maneuver hands position"	Binary (True, False)
	Total correct responses	Integer
	Total incorrect responses	Integer
H3	Interactions with in-game character	Integer
	Interactions with in-game phone	Integer
	Interactions with in-game defibrillator	Integer
H4	Total videos seen	Integer
	Total videos skipped	Integer

Table 2. Descriptive statistics about scores in pre-post questionnaires.

Variable name	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Pre score	2	6	8	7.75	9	13
Post score	3	8	10	9.643	12	14
Learning	-6	0	2	1.893	3	8

Table 3 presents the mean absolute error obtained for the most-accurate prediction model tested of each type. The prediction model that provided the best results was a neural network, obtaining the lowest mean absolute error for predictions. As neural networks are black-box models, obtaining the feature importance is, at best, not straightforward (Montavon et al., 2018). Therefore, we looked for the next best model. The prediction model with the second-best performance was a SVR, with a prediction error also low enough considering the range of predictions.

Table 3. Mean absolute error obtained for prediction models tested.

Prediction model	Mean Absolute Error (MAE)
Linear Regression	2.2358
Regression Trees	2.0196

Support Vector Machines for Regression (SVR)	0.0758
Naïve Bayes	1.9279
k-nearest neighbors (kNN)	1.4940
Neural Networks	0.0005

For that model, we could obtain the information about the input GLA variables with higher prediction relevance. Table 4 presents prediction relevance of all GLA variables. Notice that variables with highest relevance appear in both the top part of the table (variables with highest positive impact) and in the bottom part of the table (variables with highest negative impact). Variables with high impact on predictions included failure in some in-game questions. In particular, in the questions about the Heimlich maneuver and about the emergency number. In both cases, failing the question predicted lower learning.

Table 4. Prediction relevance (given as coefficient for SVR model) for GLA variables, ordered from highest to lowest values.

GLA variable	SVR coefficient
Failure in question "Abdominal thrusts"	0.2954
Replays of game level "Chest pain"	0.2796
Total videos skipped	0.2650
Maximum score in game level "Choking"	0.2100
Maximum score in game level "Chest pain"	0.1989
Replays of game level "Choking"	0.1952
Total correct responses	0.1543
Interactions with in-game defibrillator	0.1287
Failure in question "Heimlich maneuver hands position"	0.1226
Score in complete game	0.1005
Minimum time to complete game level "Choking"	0.0598
First score in game level "Unconsciousness"	0.0228
Minimum time to complete game level "Unconsciousness"	0.0086
Maximum time to complete game level "Chest pain"	0.0077
Maximum time to complete game level "Unconsciousness"	-0.0008
Minimum time to complete game level "Chest pain"	-0.0010
Maximum time to complete game level "Choking"	-0.0449
Interactions with in-game character	-0.0502
Total incorrect responses	-0.0679
Maximum score in game level "Unconsciousness"	-0.0897
Interactions with in-game phone	-0.1033
Total videos seen	-0.1792
Replays of game level "Unconsciousness"	-0.2829
Failure in question "Heimlich maneuver name"	-0.3449
First score in game level "Chest pain"	-0.3593
First score in game level "Choking"	-0.4394
Game completed	-0.5891
Failure in question "Emergency number"	-0.9652
Failure in question "Heimlich maneuver position"	-1.0602

With these results, we can accept or reject the hypotheses stated after visually analyzing the interaction data with T-MON. For each hypothesis, we analyze the prediction results obtained from the evidence-based assessment approach.

4.3.1. H1: Players who took longer to complete game levels / the game learned more.

Rejected. None of the variables containing duration (of any of the three game levels, minimum and maximum time to complete them) had a high impact on the prediction results. Therefore, we cannot conclude if game and/or level duration had an impact on learning. We hypothesized that, even if players played fast, the game has an almost-linear structure in each game level that

forces all players to go through the same scenes, therefore, all players were exposed to the same minimum content and the duration may not be so important in this case. Additionally, as mentioned during the visual exploration with T-MON, shorter durations in game levels may be explained by faster replays once players have already mastered the knowledge covered in the game level.

4.3.2. H2: Players who answered more questions correctly / answered less questions incorrectly learned more.

Accepted. The number of correct answers positively impacted (coefficient of +0.15) learning: the more correct in-game responses, the greater the difference between post-test and pre-test. Therefore, multiple-choice questions and situations seem to be a suitable learning feature to make players learn the procedures covered in the game.

Also, failing two specific questions (about the emergency number and the Heimlich maneuver) had a negative impact on learning (SVR coefficients of -0.96 and -0.10 respectively), which may point out to a higher relevance of those particular questions to overall learning. Therefore, we can conclude that correct answers in in-game questions are reflected in a higher learning (pre-post difference).

4.3.3. H3: Players who interacted more with items learned more.

Accepted. Out of the three possible items to interact with (in-game victim, phone, and defibrillator), the variable containing the number of interactions with the defibrillator (which appeared randomly in some gameplays) has a significant positive impact (coefficient of +0.12) on the prediction results. That is, a higher number of interactions with the defibrillator predicted higher learning. For the other items (in-game victim and phone), results did not show a significant relevance towards predictions, so we cannot assume how the interactions with such items affected learning if they did so. Therefore, we can conclude that interacting with some game elements leads to higher learning. The novelty of the item (defibrillator), more specific and related to the content of the game than a game character or phone, and the additional details provided to use it correctly, may be some of the reasons that made it have a higher impact on players' attention and, consequently, in their learning. And this occurs even if, as depicted in T-MON, there were fewer interactions with the defibrillator than with the other game items.

4.3.4. H4: Players who watched more videos / skipped less videos learned more.

Rejected. As opposed to what we expected, results indicate that a higher number of videos seen predicted lower learning (coefficient -0.17), while a higher number of videos skipped predicted higher learning (coefficient +0.26). A possible explanation is that players who already had the procedures clear and had acquired the knowledge, may be more prone to skip videos (as pointed out in the analysis of the default visualizations in T-MON), as well as players who were replaying the game levels only to improve their scores. In those cases, players skipped videos more times, but their learning was also higher.

5. Discussion and implications

The results of testing our hypotheses bring some new perspectives to previous literature. For H1, longer gameplay times (in multiple occasions caused by several and faster replays) did not lead to higher learning. Authors have previously discussed that elective replay after failures may increase players' learning outcomes (Syal & Nietfeld, 2020; Zhang & Rutherford, 2022), so in future work we will analyze students' replays in more detail. For H2, we found that correct answers in the game corresponded to higher learning, while failure in others to predicted lower learning. Failing questions in the game was expected to promote reflection in players, leading to a higher learning (Anderson et al., 2018), therefore, it may be that insufficient feedback is provided to students after failing certain questions; and improving those questions may promote their reflection about that content (Shute et al., 2019). For H3, interacting with game elements (particularly, with the defibrillator) increased players' learning, meshing with the generally accepted fact that interactivity is one of the key elements of educational games that promotes learning (Connolly et al., 2012; Zeng et al., 2020). For H4, the videos included in the game did not achieve their purpose of increasing learning; we think that better designed videos that more closely align to the game content, targeting specific knowledge (Shute, Smith, et al., 2021) may provide a better learning support (Yang et al., 2021).

The findings of the study can therefore provide guidelines to help educators better use a game such as the First Aid Game in their classrooms. Particularly, our results seem to point out that: (1) as maximum and minimum play time in each level does not influence learning, maybe shorter gameplays could be enough, allotting more time for an after-game discussion; (2) as in-game questions influence learning, particularly those failed, teachers may want to revise the implied concepts with students in the after-game class discussion; (3) as interactions with in-game items seem to increase learning, teachers may want to encourage students to interact more with the game, trying all options/paths/items (for instance, asking students to aim for final scores of at least an 8 out of 10 in each level); and (4) as watching the game videos did not influence learning, teachers could directly suggest that students skip those videos (and game designers should, if possible, revise their contents). The information visually displayed in T-MON could also help teachers during the use of games in classrooms: for instance, for the previous point (2), teachers may want to revise in class the questions with a lower success percentage (given by one of T-MON

visualizations). Visualizations of T-MON could further be improved by incluiding teachers in the design process to increase usability of the tool in classrooms (Kim et al., 2021).

The collected analytics can also be used to improve the serious game design. For example, by taking into account the analysis from sections 4.2.4 and 4.3.4 regarding the summary videos and their low impact in learners, we may decide to remove them from the game, to avoid any distraction of players that tend to explore all game options and, thus, require more time to complete the activity. Another insight provided by the T-MON visualizations (Figure 7 and 8) is the fact that some questions have been answered correctly by all players; for such questions, it could also be worth to reevaluate them to see if they are too easy, and could benefit from reformulating some of their answers, or if the knowledge they cover is too trivial and they may be completely removed from the game. An option to tackle this issue is to include adaptive feedback or adaptive gameplays: for instance, omitting videos in first attempts, or changing questions' difficulty based on players' results.

This study can also contribute to the body of application of educational games in the medical domain, for which some educational games already exist, particularly to train cardiopulmonary resuscitation protocols (Nery Mendes et al., 2022; Siqueira et al., 2020). Previous studies have pointed out the limited application of games especially in the medical domain since their learning outcomes have sometimes been unsatisfactory (Yu et al., 2021).

The use of standards (like xAPI-SG) can address some challenges detected in data collection of learning analytics in educational games, such as the difficulty to identify important data and the lack of reusability (Tlili et al., 2021). The use of a data collection standard simplifies the definition of the collected data and allows reusability of created systems (Pérez-Colado et al., 2021).

6. Conclusions

With the present work, we describe some of the opportunities offered by analytics in game-based learning, the relevance of systematizing the process by using standards and game-independent analyses and visualizations, and the different techniques (visualizations, data mining models) that can be applied to yield meaningful information to better understand learners' actions and results in serious games. In particular, the combination of both an exploratory and visual tool (T-MON) with data mining models has allowed us to obtain an in-depth analysis of the relationship between players' interactions in the game and their learning results.

The exploratory tool T-MON provides the first step by displaying a default set of visualizations based on the fields available in the xAPI-SG standard. All visualizations included in T-MON are game-independent, that is, no specific data from the game is required to create them. This allows to conduct such exploratory analysis instantly for any given game, as long as the collected interaction data follows the standard xAPI-SG and implies a cost reduction at the time of the first analysis.

The evidence-based assessment approach has allowed us to obtain a clearer relationship between players' actions and learning results. An important note is that we have been able to accept/reject the proposed hypotheses only because we are analyzing results of a model that provides information about variables relevance (SVR). For any other black-box model, such a discussion would not be possible or, at least, it will not be as straightforward (Dreiseitl & Ohno-Machado, 2002). It is also relevant that, even though the model with best results was a black-box model (neural network), we could analyze the results of the second-best model as the range of predictions and context is not that restricted: that is, to predict learning within the educational game, the results of the second-best model were accurate enough. In other fields, such as the medical domain, the difference between those models may become not only significative but crucial, and therefore, using only an "accurate-enough" model may not provide sufficient information on the relevance of variables.

We also need to point out some limitations and future lines of work. First, the case study simply aims to exemplify our combined approach, using a simple narrative game with enough features that allows T-MON's visualizations to be showcased and explored, and the discussion of various hypotheses. The default analyses and visualizations may omit some relevant game information, but this can be detected in the evidence-based assessment approach by adding game-specific relevant data when creating the GLA variables. As future work, we also plan to improve and extend the visual exploratory tool T-MON with more complex analyses and visualizations. In particular, we plan to address the fact that some visualizations, such as the progress of players over time, are only currently useful when displaying small class-size samples (~20-30 students). There are multiple approaches to visualize much larger datasets, such as those described by (Zhao et al., 2022). It is also possible to improve T-MON to use different default analyses according to game genre and mechanics.

Additionally, to obtain information on the relevance of variables when assessing players, we must either restrict ourselves to white-box models or provide some other means to obtain feedback on the relevance of particular variables in the results. More complex prediction models may provide more accurate results, but at the cost of harder to convey information on the relevance of specific variables. In subsequent research, we also plan to differentiate students' attempts and replays in game levels, and study learning for such situations independently, with emphasis on learning on first play, while exploring other behaviors such as persistence, struggle and other 21st century skills in the following attempts and replays. Other 21st century skills like creativity have already been assessed with learner data in educational games (Shute & Rahimi, 2021).

We believe that there is still much work to do in the field of learning analytics applied to educational and serious games. As this paper proposes, including systematic and standardized approaches from data collection (e.g., xAPI-SG Profile) to data analysis and visualization (default game-independent analyses) can simplify all steps in the process and allow all stakeholders involved (game designers and developers, educators, students) to have clearer information to improve their games and evaluate their students playing them.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

Acknowledgments

This work has been partially funded by the Regional Government of Madrid (eMadrid S2018/TCS-4307, co-fundedby the European Structural Funds FSE and FEDER), by the Ministry of Education (PID2020-119620RB-I00) and by the Telefónica-Complutense Chair on Digital Education and Serious Games.

References

ADL. (2012). *Experience API (xAPI) Standard*. https://adlnet.gov/projects/xapi/

ADL. (2021, October 20). *IEEE to Standardize xAPI v2.0 as an International Standard*. https://adlnet.gov/news/2021/10/20/IEEE-to-Standardize-xAPI-v2.0-as-an-International-Standard/

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., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2021). Improving evidence-based assessment of players using serious games. *Telematics and Informatics*, *60*, 101583. https://doi.org/10.1016/j.tele.2021.101583

Andersen, E., Liu, Y.-E., Apter, E., Boucher-Genesse, F., & Popović, Z. (2010). Gameplay analysis through state projection. *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, 1–8. https://doi.org/10.1145/1822348.1822349

Anderson, C. G., Dalsen, J., Kumar, V., Berland, M., & Steinkuehler, C. (2018). Failing up: How failure in a game environment promotes learning through discourse. *Thinking Skills and Creativity*, *30*, 135–144. https://doi.org/10.1016/j.tsc.2018.03.002

Andrews-Todd, J., Mislevy, R. J., LaMar, M., & de Klerk, S. (2021). *Virtual Performance-Based Assessments* (pp. 45–60). https://doi.org/10.1007/978-3-030-74394-9_4

Carter, M., & Egliston, B. (2021). What are the risks of Virtual Reality data? Learning Analytics, Algorithmic Bias and a Fantasy of Perfect Data. *New Media & Society*, 146144482110127. https://doi.org/10.1177/14614448211012794

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

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

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, *35*(5–6), 352–359. https://doi.org/10.1016/S1532-0464(03)00034-0

Eagle, M., Johnson, M., Barnes, T., & Boyce, A. (2013). Exploring Player Behavior with Visual Analytics. *Proceedings of the International Conference on the Foundations of Digital Games - FDG '13*.

e-UCM. (2020). *T-MON: Traces Monitor in xAPI-SG*. https://github.com/e-ucm/t-mon

European Commission. (2016). *What does the General Data Protection Regulation (GDPR) govern?* https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern_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). Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_21-1

Hao, J., & Mislevy, R. J. (2018). The Evidence Trace File: A Data Structure for Virtual Performance Assessments Informed by Data Analytics and Evidence-Centered Design. *ETS Research Report Series*, *2018*(1), 1–16. https://doi.org/10.1002/ets2.12215

Khosravi, H., Shabaninejad, S., Bakharia, A., Sadiq, S., Indulska, M., & Gašević, D. (2021). Intelligent Learning Analytics Dashboards: Automated Drill-Down Recommendations to Support Teacher Data Exploration. *Journal of Learning Analytics*, *8*(3), 133–154. https://doi.org/10.18608/jla.2021.7279

Kim, Y. J., Lin, G., & Ruipérez-Valiente, J. A. (2021). *Expanding Teacher Assessment Literacy with the Use of Data Visualizations in Game-Based Assessment* (pp. 399–419). https://doi.org/10.1007/978-3-030-81222-5_18

Liu, M., Han, S., Shao, P., Cai, Y., & Pan, Z. (2021). *The Current Landscape of Research and Practice on Visualizations and Dashboards for Learning Analytics* (pp. 23–46). https://doi.org/10.1007/978-3-030-81222-5_2

Marchiori, E., Ferrer, G., Fernández-Manjón, B., Povar Marco, J., Fermín Suberviola, J., & Giménez Valverde, A. (2012). Video-game instruction in basic life support maneuvers. *Emergencias*, *24*.

Minović, M., & Milovanović, M. (2013). Real-time learning analytics in educational games. *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality - TEEM '13*, 245–251. https://doi.org/10.1145/2536536.2536574

Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011

Nery Mendes, I., de Araújo Nogueira, M., Valente Mendes, F., Noura Teixeira, O., & Almeida dos Santos, V. (2022). The Use of Serious Games for Learning Cardiopulmonary Resuscitation Procedures: A Systematic Mapping of the Literature. In *Computer Game Development [Working Title]*. IntechOpen. https://doi.org/10.5772/intechopen.102399

Nguyen, T.-H. D., Seif El-Nasr, M., & Canossa, A. (2015). Glyph: Visualization Tool for Understanding Problem Solving Strategies in Puzzle Games. *Proceedings of the International Conference on the Foundations of Digital Games - FDG '15*. https://doi.org/https://doi.org/10.48550/arXiv.2106.13742

Pérez-Colado, V. M., Pérez-Colado, I. J., Martínez-Ortiz, I., Freire-Morán, M., & Fernández-Manjón, B. (2021). *Democratizing Game Learning Analytics for Serious Games* (pp. 164–173). https://doi.org/10.1007/978-3-030-92182-8_16

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

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

Shute, V. (2011). Stealth assessment in computer-based games to support learning. In *Computer games and instruction.* (pp. 503–524). IAP Information Age Publishing.

Shute, V., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, *116*, 106647. https://doi.org/10.1016/j.chb.2020.106647

Shute, V., Rahimi, S., & Lu, X. (2019). *Supporting Learning in Educational Games: Promises and Challenges* (pp. 59–81). https://doi.org/10.1007/978-981-13-8265-9_4

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C., Kuba, R., Liu, Z., Yang, X., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, *37*(1), 127–141. https://doi.org/10.1111/jcal.12473

Shute, V., Smith, G., Kuba, R., Dai, C.-P., Rahimi, S., Liu, Z., & Almond, R. (2021). The Design, Development, and Testing of Learning Supports for the Physics Playground Game. *International Journal of Artificial Intelligence in Education*, *31*(3), 357–379. https://doi.org/10.1007/s40593-020-00196-1

Siqueira, T. V., Nascimento, J. da S. G., Oliveira, J. L. G. de, Regino, D. da S. G., & Dalri, M. C. B. (2020). The use of serious games as an innovative educational strategy for learning cardiopulmonary resuscitation: an integrative review. In *Revista gaucha de enfermagem* (Vol. 41, p. e20190293). NLM (Medline). https://doi.org/10.1590/1983-1447.2020.20190293

Syal, S., & Nietfeld, J. L. (2020). The impact of trace data and motivational self-reports in a game-based learning environment. *Computers & Education*, *157*, 103978. https://doi.org/10.1016/j.compedu.2020.103978

Tlili, A., & Chang, M. (2019). *Data Analytics Approaches in Educational Games and Gamification Systems* (A. Tlili & M. Chang, Eds.). Springer Singapore. https://doi.org/10.1007/978-981-32-9335-9

Tlili, A., Chang, M., Moon, J., Liu, Z., Burgos, D., Chen, N. S., & Kinshuk. (2021). A systematic literature review of empirical studies on learning analytics in educational games. *International Journal of Interactive Multimedia and Artificial Intelligence*, *7*(2), 250–261. https://doi.org/10.9781/ijimai.2021.03.003

Yang, X., Rahimi, S., Shute, V., Kuba, R., Smith, G., & Alonso-Fernández, C. (2021). The relationship among prior knowledge, accessing learning supports, learning outcomes, and game performance in educational games. *Educational Technology Research and Development*, *69*(2), 1055–1075. https://doi.org/10.1007/s11423-021-09974-7

Yu, Z., Gao, M., & Wang, L. (2021). The Effect of Educational Games on Learning Outcomes, Student Motivation, Engagement and Satisfaction. *Journal of Educational Computing Research*, *59*(3), 522–546. https://doi.org/10.1177/0735633120969214

Zeng, J., Parks, S., & Shang, J. (2020). To learn scientifically, effectively, and enjoyably: A review of educational games. *Human Behavior and Emerging Technologies*, *2*(2), 186–195. https://doi.org/10.1002/hbe2.188

Zhang, Q., & Rutherford, T. (2022). Grade 5 Students' Elective Replay After Experiencing Failures in Learning Fractions in an Educational Game: When Does Replay After Failures Benefit Learning? *LAK22: 12th International Learning Analytics and Knowledge Conference*, 98–106. https://doi.org/10.1145/3506860.3506873

Zhao, Y., Wang, Y., Zhang, J., Fu, C.-W., Xu, M., & Moritz, D. (2022). KD-Box: Line-segment-based KD-tree for Interactive Exploration of Large-scale Time-Series Data. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 890–900. https://doi.org/10.1109/TVCG.2021.3114865