Applications of data science to game learning analytics data: 
a systematic literature review

Cristina Alonso-Fernándeza,*, Antonio Calvo-Morataa, Manuel Freirea,
Iván Martínez-Ortiza, Baltasar Fernández-Manjóna

*aDepartment of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid, Spain

Abstract

Data science techniques, nowadays widespread across all fields, can also be applied to the wealth of information derived from student interactions with serious games. Use of data science techniques can greatly improve the evaluation of games, and allow both teachers and institutions to make evidence-based decisions. This can increase both teacher and institutional confidence regarding the use of serious games in formal education, greatly raising their attractiveness. This paper presents a systematic literature review on how authors have applied data science techniques on game analytics data and learning analytics data from serious games to determine: (1) the purposes for which data science has been applied to game learning analytics data, (2) which algorithms or analysis techniques are commonly used, (3) which stakeholders have been chosen to benefit from this information and (4) which results and conclusions have been drawn from these applications. Based on the categories established after the mapping and the findings of the review, we discuss the limitations of the studies analyzed and propose recommendations for future research in this field.

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

1. Introduction

Use of data science (e.g. artificial intelligence) techniques has spread over many fields, with a wide range of purposes. The huge, and constantly growing, amounts of data being captured allow complex techniques to provide insights which are potentially deeper than those that can be found by applying only traditional, and often simpler, methods.

The application of data science techniques perfectly fits interactive environments, where multiple data can be generated. One of these environments that allows for multiple interactions is games. In particular, the use of games with purposes beyond entertainment (e.g. learning, raising awareness or changing attitudes and behaviors), that is, so called serious games (SGs), has also increased in the last years. These types of games are especially popular in domains such as medicine or the military, and have proven their effectiveness for children
and adolescents, as the familiarity of these users with gaming environments and the characteristics of games (interactivity, motivation, engagement) facilitate their interactions with serious games.

The collection and analysis of data has reached a great number of fields: in education, the fields of educational data mining (EDM) and learning analytics (LA), sometimes used interchangeably, are widely spread (Long & Siemens, 2011; Romero & Ventura, 2010). Their aim is to understand learners and their environments and improve the learning process through analysis of the data collected from students’ interactions with the learning environment. As with any other highly interactive system, a lot of data can also be gathered from serious games to guide data-based decision-making (Loh, Sheng, & Ifenthaler, 2015). Building up from the fields of educational data mining and learning analytics, which focus in education in general, game learning analytics (GLA) is defined as the collection, analysis and extraction of information from data collected from serious games (Alonso-Fernández et al., 2019; Freire et al., 2016; Owen & Baker, 2019).

The aim of the current paper is to conduct a systematic literature review on the applications of data science techniques to analyze game analytics data and/or learning analytics data from serious games. The rest of the paper is structured as follows: Section 2 provides a summary on related work; Section 3 describes the methodology used for the systematic literature review; Section 4 presents the results obtained; finally, Section 5 discusses the results, and presents the limitations and conclusions of the review.

2. Related work

The fields of serious games, learning analytics and data science have attracted considerable interest and attention in the last decade. While many works have been published related to these topics, we have not found any existing systematic literature reviews that combine the three topics together. The present work seeks to bridge this gap. In this section, we briefly present related literature reviews that involve at least one of the fields of serious games, learning analytics and data science; we also describe several that combine two of these topics.

We have found several literature reviews that examine serious games, each focusing on different aspects. That of (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012) focuses on the potential positive impact of gaming with respect to learning, skill enhancement and engagements, finding that the most frequently occurring outcomes and impacts were knowledge acquisition/content understanding, and affective and motivational outcomes. The review by (Petri & Gresse von Wangenheim, 2017) focused on the evaluation of serious games, finding that there are few approaches to systematically evaluate educational games. Special issues have also been published for different related fields like game visual analytics (Wallner, Canossa, & El-Nasr, 2018) or learning assessment (Berta & Moreno-Ger, 2018).

The possible applications of data science for games have also been studied. For instance, the work of (Yannakakis & Togelius, 2018) presents the major application areas of artificial intelligence methods within games: game-playing, content generation and player modeling. Regarding the possible applications of learning analytics on serious games, the literature review presented in the work of (Liu, Kang, Liu, Zou, & Hodson, 2017) focuses on uses of LA for assessment, but differs from our work in that we also focus on the specific data science techniques used and consider a broader set of purposes – not only assessment. Their results showed that SGs had a positive impact on learning and highlighted the importance of game design.

Although there are some similarities between the works described above and the systematic literature review presented in this paper, our work is different in that it focuses on serious games and the application of data science algorithms to game analytics data and/or learning analytics data coming from these types of games.
3. Method

3.1. Research questions

The main goal of this systematic literature review is to explore the applications of data science to game analytics data and/or learning analytics data from serious games. For this purpose, we have stated the following main research questions:

**RQ1.** What are the purposes for which data science has been applied to game analytics data and/or learning analytics data from serious games?

**RQ2.** What data science algorithms or techniques have been applied to game analytics data and/or learning analytics data from serious games?

**RQ3.** What stakeholders are the intended recipients of the analysis results?

**RQ4.** What results and conclusions have been drawn from these applications?

Additionally, we intend to extract some further information from the studies, to complement the results:
- The main purpose of the games (e.g. teaching, change behavior) and their domain (e.g. biology, math)
- The sample size of the studies, and the educational level of their participants
- The general characteristics of the in-game interaction data collected, and the data format used

3.2. Data collection

We follow a standard systematic literature review methodology, using a fixed set of queries on a pre-identified list of bibliographical databases, and clear inclusion/exclusion criteria.

3.2.1. Databases searched

We have queried 9 databases, including some of the main databases for education, computer science, and general science research. Specifically, we have searched: Association for Computing Machinery (ACM), Cambridge Journals Online, Education Resources Information Center (ERIC), IEEE Computer Society Digital Library (CDSL), IngentaConnect, Oxford University Press (journals), Science Direct, Scopus and Springer.

3.2.2. Search terms

To perform the searches on the databases, we focus on our three main terms of interest: data science, game analytics and learning analytics, and games. As seen in the Introduction, the terms “learning analytics” and “educational data mining” are sometimes used interchangeably, so we conducted two parallel searches, one for each of these terms. All searches are restricted to title, abstract and keywords.

- **Search query for game analytics**
  We included the term “game analytics” and several alternative terms for “data science” and specific analysis techniques. Final search query:
  (“game analytics”) AND (“artificial intelligence” OR “data mining” OR “machine learning” OR “data analysis” OR “deep learning”)

- **Search query with learning analytics**
  We included the terms “learning analytics”, “games”, and alternative terms for analysis techniques. Final search query:
  (“learning analytics”) AND (“games”) AND (“artificial intelligence” OR “data mining” OR “machine learning” OR “data analysis” OR “deep learning”)
• **Search query with educational data mining**
  As the term “educational data mining” includes the analysis of the data, we do not include additional terms of data analytics, and therefore used:
  (“educational data mining”) AND (“games”)

• **Additional search query on Journal of Artificial Intelligence in Education**
  Finally, from our previous research, we encountered a specific journal on artificial intelligence in education, the International Journal of Artificial Intelligence in Education (Springer). We performed an additional search for papers of this journal which included the term “games”.

3.2.3. **Study selection**

After removing duplicates, we scanned the title and abstract of all papers, comparing them against the inclusion and exclusion criteria below. After this first scanning, studies were classified as either possible or excluded. Clearly irrelevant publications were excluded, while those classified as possible were read (conclusions or even full text) to ensure relevancy. We examined possibly-relevant papers with our research questions in mind, to ensure that they provided enough information about purposes, techniques and stakeholders regarding the application of data science techniques to GLA data from serious games. Additionally, we looked for studies which included information on the games from which data was collected, sample sizes, and details on the interaction data collected from games; although this information was not mandatory for papers to be included in the review. No time restrictions were set.

**Inclusion criteria**

- Journals, conference papers or book chapters
- Include empirical evidence relating the outcomes of applying data science techniques to game analytics data and/or learning analytics data from serious games

**Exclusion criteria**

- Publications whose full text is not available
- Publications not written in English

3.3. **Data analysis**

For each of the studies selected for the literature review, we collected data on each of our research questions and conducted a mapping study to categorize the results of each research question. When available, we also collected additional information that could provide a more in-depth review of the applications of data science techniques to GLA data from SGs. We classified the data of the selected studies according to the following criteria:

- The main purposes of the analysis of the data collected from SGs (addressing RQ1)
- The algorithms or techniques used to analyze the data collected from SGs (addressing RQ2)
- The stakeholder that is the main beneficiary of the extracted information (addressing RQ3)
- The results and conclusions of the analysis of the data collected from SGs (addressing RQ4)
- The purpose of the serious game
- The domain of the serious game
- The sample size of the study
- The educational level or specific characteristics of the participants
- The in-game interaction data captured
- The format of the in-game interaction data captured
4. Results

4.1. Studies identified by search terms

Studies were retrieved in December 2018 using the search terms. In this first search, 272 studies were found. After analyzing the results, we excluded 74 duplicate publications, yielding 198 unique studies.

4.2. Studies selected using inclusion criteria

The selection process began with 198 studies. After scanning their titles and abstract, 77 studies were excluded for not meeting our inclusion criteria. An additional full text review was performed to ensure the suitability of the papers. On this final review, 34 studies were excluded for not meeting one or more of the inclusion criteria, such as not mentioning serious games or not collecting any data from the games. The final sample consists of 87 studies. Fig. 1. summarizes the full selection process.

Fig. 1. Studies selection process for the systematic literature review.

Table 1 shows the total number of studies identified in the search process and meeting inclusion criteria from each database considered.

Table 1. Number of studies identified in search and meeting inclusion criteria from each database.

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of studies identified in search</th>
<th>Number of studies meeting inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Cambridge</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>ERIC</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>IEEE CDSL</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>IngentaConnect</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Oxford</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Science Direct</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Scopus</td>
<td>96</td>
<td>28</td>
</tr>
<tr>
<td>Springer</td>
<td>112</td>
<td>33</td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>87</td>
</tr>
</tbody>
</table>

Regarding the year of publication of the selected studies, Fig. 2. shows the number of papers selected for the literature review for each year of publication. Note that two papers included in the review, to be published in
2019, were available online when the search was conducted; in the figure they are therefore considered as published in 2018. The figure shows an increased interest in the topics of the review from 2011 onwards.

![Figure 2](image.png)

**Fig. 2.** Number of selected studies found per year of publication.

4.3. **Main purpose of studies**

This subsection responds to RQ1 based on the studies that met all inclusion criteria:

**RQ1.** What are the purposes for which data science has been applied to game analytics data and/or learning analytics data from serious games?

After considering all qualifying studies, we mapped their main data science application purpose into one of 5 categories: learning assessment, study of in-game behaviors, game design or evaluation, student profiling, and interventions. Some studies focus on more than one of the previous purposes. Since several studies proposed

<table>
<thead>
<tr>
<th>Purpose category</th>
<th>Definition of purpose</th>
<th>Example studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>Assess learning, predict performance</td>
<td>(Ke and Shute 2015; R. S. Baker, Clarke-Midura, and Ocumpaugh 2016)</td>
</tr>
<tr>
<td>In-game behaviors</td>
<td>Study in-game players behaviors (e.g. persistence, engagement)</td>
<td>(Dicerbo 2013; Kang, Liu, and Qu 2017)</td>
</tr>
<tr>
<td>Game design</td>
<td>Validate game design</td>
<td>(Tlili et al. 2016; Cano, Fernández-Manjón, and García-Tejedor 2018)</td>
</tr>
<tr>
<td>Student profiling</td>
<td>Establish categories of players profiles, differentiate players characteristics</td>
<td>(Loh and Sheng 2014; Denden et al. 2018)</td>
</tr>
<tr>
<td>Interventions</td>
<td>Study effect of in-game interventions (e.g. feedback messages, notification of performance)</td>
<td>(McCarthy et al. 2017; DeFalco et al. 2018)</td>
</tr>
<tr>
<td>Framework proposals</td>
<td>Propose a framework for specific contexts</td>
<td>(Halverson and Owen 2014; Nguyen, Gardner, and Sheridan 2018)</td>
</tr>
</tbody>
</table>
frameworks to simplify the application of GLA data for SGs in specific contexts, but their primary purpose was unrelated to data science, we framed those studies under an additional, 6th category named framework proposals. Table 2 details the purposes of each data science related category, and lists two example studies for each.

![Bar chart showing purposes of data science related categories](chart.png)

**Fig. 3.** Main purpose of data analysis in the selected studies. Some studies focus on several purposes.

Fig. 3 shows the main purpose of the selected studies. Assessment (32 studies, 36.8%) followed by in-game behaviors (27 studies, 31%), are the two main purposes.

4.4. *Data science algorithms or techniques*

This subsection provides answers to RQ2:

**RQ2.** What data science algorithms or techniques have been applied to game analytics data and/or learning analytics data from serious games?

The data science algorithms and techniques used in the reviewed studies can be grouped into three main categories:

- Supervised algorithms: linear and logistic regression, regression and decision trees, support vector machines, Bayesian networks, neural networks, naive Bayes, and Bayesian knowledge tracing.
- Unsupervised algorithms: correlation, clustering, factor analysis.
- Visualization techniques: display of gameplay pathways, performance metrics, learning curves*, heatmaps of interactions, use of in-game tools (frequency or duration).

Note that some studies present results of the application of various techniques and algorithms. Table 3 summarizes the techniques used and the number of studies that use each technique. The three main categories are used in a similar number of studies. For each of the three categories, we have specified the methods that are used in more than one study. We can see that linear models are the most used methods for supervised models (in 18 studies), while correlation and cluster analysis are the most widely used unsupervised methods in the studies (in 17 and 16 studies, respectively). Among the visualizations presented in the studies, a majority (15 studies) focus on displaying performance information.

*Plot of speed or accuracy over game levels (Martin, Koedinger, Mitrovic, & Mathan, 2005).*
Table 3. Data science techniques used in the studies, and number of papers that use each technique.

<table>
<thead>
<tr>
<th>Data science technique</th>
<th>Number of papers using the technique</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supervised models</strong></td>
<td></td>
</tr>
<tr>
<td>Linear/logistic regression</td>
<td>18</td>
</tr>
<tr>
<td>Regression/decision trees</td>
<td>7</td>
</tr>
<tr>
<td>Bayesian networks</td>
<td>6</td>
</tr>
<tr>
<td>Neural networks</td>
<td>4</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>3</td>
</tr>
<tr>
<td>Bayesian knowledge tracing</td>
<td>3</td>
</tr>
<tr>
<td>Support vector machines</td>
<td>2</td>
</tr>
<tr>
<td><strong>Unsupervised models</strong></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>17</td>
</tr>
<tr>
<td>Clustering</td>
<td>16</td>
</tr>
<tr>
<td>Factor analysis</td>
<td>2</td>
</tr>
<tr>
<td><strong>Visualization</strong></td>
<td></td>
</tr>
<tr>
<td>Performance metrics</td>
<td>15</td>
</tr>
<tr>
<td>Gameplay pathways</td>
<td>7</td>
</tr>
<tr>
<td>Use of in-game tools</td>
<td>5</td>
</tr>
<tr>
<td>Learning curves</td>
<td>4</td>
</tr>
<tr>
<td>Heatmaps of interactions</td>
<td>2</td>
</tr>
</tbody>
</table>

4.5. Stakeholders

This subsection addresses RQ3:

**RQ3.** What stakeholders are the intended recipients of the analysis results?

The five stakeholders considered in the studies are: teachers/educators, serious game designers/developers, students/learners, researchers (or studies with research purposes) and parents. Fig. 4. shows the number of studies that focus on each of the stakeholders. Game designers or game developers are the main target of most studies (39 studies), closely followed by researchers, or studies with research purposes (37 studies); and teachers or educators (25 studies). We notice that when researchers are the main target of studies, they usually have additional roles (for instance, they also act as game designers or developers).

![Fig. 4. Main stakeholder recipient of the analysis results in selected studies.](image-url)
4.6. General information of the studies

Before moving to RQ4, this subsection looks at additional information of the studies regarding the serious games used, the participants and sample sizes, and the nature and contents of the captured interaction data.

4.6.1. Serious games used

Regarding the purpose of the serious games used, 55 studies (63.2%) use games that aim to teach, 8 (9.2%) to train and 6 (6.9%) to assess. 2 studies use games to raise awareness and 1 to change behaviors. The remaining 15 studies (17.2%) do not clearly state the purpose of the game used.

The domain that most serious games in the selected studies focus on is mathematics (20 studies). 10 studies use science games and 4 studies focus on problem solving. Biology and physics are the domain of games in 3 studies each, while computer architecture, military, memory, language/reading and ability to design scientific investigations are the focus of games in 2 studies each. Other domains mentioned only in one study each include: business, programming, project management, ecology, research methodology, algorithmic and critical thinking, strategy planning or team work.

4.6.2. Participants and sample size

Regarding samples (N) used on the studies: 28 studies (32%) used fewer than 100 participants, while 27 studies (31%) use between 100 and 1000 participants. 7 studies (8%) used more than 1000 participants and only one study reported more than 10000 participants. The remaining 24 studies (27%) did not report data about participants. The samples of the 63 studies who reported participants are skewed by the highest values (mean = 1643 participants with SD = 10200; however, median = 116). Excluding outlier values from the calculations, that is, studies with over 1000 participants, the mean sample size for the remaining 55 studies drops to 161 participants (SD = 191). The high value of the standard deviation can be explained by the large number of papers with less than 100 participants (28 of these remaining 55 studies).

Analyzing the educational level of participants, 18 studies (20.7%) focus on primary school (up to 12 years old), 22 (25.3%) on secondary school (from 12 to 18 years old), 11 (12.6%) on undergraduates, 2 (2.3%) on graduates, and 8 (9.1%) on adults. 26 studies (30%) did not state this information.

4.6.3. Interaction data captured

An additional goal was to characterize data collected from the serious games. The game analytics data and/or learning analytics data captured in the selected studies include: completion times (in 30 studies); actions/interactions in general (28 studies); scores (14); correct/incorrect answers (11); clicks (9); attempts/tries (8); choices, errors/mistakes, answers (7); start/end, variables, completion, duration (4); events, performance, action sequences, contents accessed (3); phase/level changes, posts, location, context, success, items used/collection, progress, number of players (2). Other studies mention preferences, health, or use of in-game hints.

The format of the collected data is not specified in most studies: only 1 study reported the use of CSV as a format, 3 studies reported use of Experience API (xAPI) (ADL, 2012), 3 used XML, 3 used ad-hoc strings, and 6 studies used tables. The remaining 71 out of the 87 selected studies (81.6%) did not report any specific format of the collected data.

4.7. Results and conclusions of data analysis

This subsection addresses RQ4:
RQ4. What results and conclusions have been drawn from these applications?

The results and conclusions of the studies have been grouped based on the topics they are related to.

4.7.1. Results on assessment and student profiling

Several studies focus on assessment and learning predictions, also relating these with learners’ characteristics and in-game behaviors. These results correspond to studies with purposes tagged in RQ1 as assessment, in-game behaviors and student profiling. A summary of the results of these studies is that:

**GLA data can accurately predict games’ impact:**

- The application of GLA data can be useful both at real-time (online) and after the intervention is completed (offline) (Wiemeyer, Kickmeier-Rust, and Steiner 2016), and for all stakeholders (Alonso-Fernández et al. 2019). The analysis of interaction data can provide a means to measure the proven positive impact of games (Kosmas, Ioannou, and Retalis 2018; Mavridis, Katmada, and Tsiatsos 2017). However, most data is still captured after the game (Smith, Blackmore, and Nesbitt 2015). Authors also point out the need for specific game learning analytics (Freire et al. 2016), or so-called serious games analytics (Loh, Sheng, and Ifenthaler 2015), that differ from games analytics in general.

- Learning predictors: predictions of player success can often be accurately created based on log data (R. S. Baker, Clarke-Midura, and Ocumphaugh 2016), as the achievement system built into games may not be the most informative indicator of learning (Heeter et al. 2013). Some papers point out that their best predictors for measuring learning are based on the analysis of the player’s exploration strategies (Käser, Hallinen, & Schwartz, 2017; Owen, Anton, and Baker 2016; Horn et al. 2016; Smith, Hickmott, Southgate, Bille, & Stephens, 2016; Kang, Liu, and Qu 2017), or on the characteristics of player failures (Halverson and Owen 2014). One study explored the relation between learning and students’ facial behavior (Z. Xu and Woodruff 2017). One study found, analyzing interactions in an online discussion forum, that the content that best explained and predicted learning was related to uncertainty, decision-making, time, collaboration and communication (Hernández-Lara, Perera-Lluna, and Serradell-López 2019). In crowd-sourced serious games, three game-play metrics (active users, session counts and session time) were found to be good indicators of productivity by (Tellioglu et al. 2014), while team cohesion and psychological safety may be good performance indicators in multiplayer serious games (Mayer et al. 2014). Also, behaviors such as avoiding a concept indicated poor performance (Ketamo 2013). Implicit learning can also be adequately measured through behaviors (Rowe et al. 2017) and game log data (Rowe, Asbell-clarke, and Baker 2015). Learning curves provide insights into learning (Peddycord-Liu et al., 2018) and can be studied for speed and accuracy (Eagle 2009; R. S. J. D. Baker et al. 2007).

- Recommendations to improve predictions: feature engineering improves performance of models with simple raw data (Owen and Baker 2018). Additional information, such as the domain structure and the weights of competencies, improves accuracy of prediction models (Kickmeier-Rust 2018). Exploratory data analysis (DiCerbo et al. 2015) and dynamical analysis (Snow, Allen, and McNamaara 2015) can uncover unexpected patterns and provide richer information about students’ interactions.

- Creating assessment conditions: additional information may be extracted by letting teachers define assessment rules based on and combining generic game trace variables to obtain new information (Steiner, Kickmeier-Rus, and Albert 2015). Complex assessment conditions can be created by combining some of the basic sets of traces (Serrano-Laguna et al. 2014).
Performance is related to players’ characteristics:

- Clusters of players in performance groups: players can be clustered into performance groups based on in-game actions (Martin et al. 2015; Slimani et al. 2018; Freitas and Gibson 2014; Forsyth et al. 2012; Lazo, Anareta, Duremdes, & Red, 2018; Polyak, von Davier, and Peterschmidt 2017; Martinez-Garza and Clark 2017; Chung 2015) and in-game choices (Cutumisu et al. 2017), which are also related to prior knowledge (Martinez-Garza and Clark 2017). Some studies explore methods to differentiate experts from novice users (Loh and Sheng 2015b; Loh and Sheng 2014; Loh and Sheng 2015a). Once students are classified in a performance group, scores can be inferred when time or action sequences are added to the analysis (Gibson and Clarke-Midura 2015). Tactics that lead to success can also be discovered with cluster analysis (Sharples and Domingue 2016).

- Importance of understanding learners’ characteristics: students with different learning characteristics may exhibit different learning behaviors (Liu et al. 2016), for instance, different exploration strategies (Martin et al. 2013) or, in some cases, their age and gender (Wallner and Kriglstein 2015). It is key to model students for effective adaptive instruction (Koedinger, McLaughlin, and Stamper 2012), for instance, self-regulated learners tend to make better use of in-game curricular resources and may be more deliberate in their actions (Sabourin et al. 2013) and high-performance students tend to use tools more appropriately (Liu et al. 2015). Behaviors also depend on student background (Jaccard, Hulaas, and Dumont 2017). Personalities can be identified based on actions and in-game choices (Denden et al. 2018).

Further information can be extracted from GLA data:

- Real-time information to stakeholders: in (Elaachak, Belahbibe, and Bouhorma 2015) a system is presented that combines information for teachers, displaying a pie chart of students’ performance, and for students, with assistance messages displayed on the screen according to their performance and progress. Some studies also provide systems that allow parents to receive real-time information about their children’s learning (Roberts, Chung, and Parks 2016; Ketamo 2015).

- Measure other students’ characteristics: some studies include additional applications of analysis of GLA data to track students’ progress (Gweon et al., 2015), assess persistence (Dicerbo 2013) or detect engagement (Ghergulescu and Muntean 2016).

4.7.2. Results on serious game design

Several studies focus on applications to obtain further insight and improve serious game design and implementation. These results correspond to studies with purposes tagged in RQ1 as game design, interventions, and framework proposal. Studies have drawn several conclusions and pointed out recommendations for serious game design based on findings of the analyzed game learning analytics data, including:

GLA data can validate serious game design:

- Several studies use game learning analytics data to validate serious game design (Cano, Fernández-Manjón, and García-Tejedor 2018; Serrano-Laguna et al. 2012; Tlili et al. 2016; Harpstead et al. 2015), specific design choices (Cheng et al. 2017) and even to create new game mechanics (Ninaus et al. 2017) or to automatically discover speech act categories for dialogue-based educational games (Rus et al. 2012).

Assessment can and should be integrated in serious game design:

- Recommendations for creating assessment in SGs: assessment design and learning context/task design should be considered in the early phase of game development (Ke and Shute 2015). One study proposes
a design approach to integrate data-driven assessment in game design (Ke et al. 2019). Debriefing via visualizations can improve understanding of outcomes (De Troyer, Helalouch, and Debruyne 2016).

- Data to be collected: before applying GLA, and as part of the game design, it is highly recommended to specify and determine the game traces that will be collected (Tlili et al. 2016; Serrano-Laguna et al. 2018).
- Teachable agents: in educational games, teachable agents can help achieve deeper levels of learning that transfer outside the game (Pareto 2014) and have a significant impact on in-game performance, preferably when designed to have low self-efficacy (Tärning et al. 2018).

Importance of serious games characteristics:
- Difficulty: for games without adaptive difficulty, it is especially important to present a smooth difficulty curve. For (Hicks et al. 2016), a high difficulty level increased dropout. It is also important to classify students and modulate difficulty (Martinez-Garza and Clark 2017). Allowing players to return to game areas with lower difficulty significantly reduced error rates and increased learning rate according to (Käser et al. 2013).
- Engagement and motivation: it is important to design for engagement by matching challenges with incentives and motivating activities (Pareto 2014). Games should include motivational elements (Tlili et al. 2016). Engagement seems to decrease in internet experiments (Stamper et al. 2012).
- Feedback and interventions during play: too many metacognitive prompts during play may be detrimental (McCarthy et al. 2017). Mixed results obtained for self-efficacy enhancing interventions based on interaction-based affect detectors to enhance outcomes (DeFalco et al. 2018).

Identified challenges when designing serious games:
- Designing games for assessment: assessment routines are usually black boxes that teachers cannot inspect. Studies have identified a need for transparent and reliable assessment in educational games, based on assessment models that are ideally valid, easy to use, and provide meaningful educational information, while giving game industry evidence on game quality (Steiner, Kickmeier-Rus, and Albert 2015). Different design decisions may be considered to explore how they affect learning outcomes (Plass et al. 2013). Adaptivity is also desired, although there is still a lack of real applications due to its high costs (Streicher and Smeddinck 2016).
- Designing games with tracking features: game manufacturers are resistant to include data recording of learning evidence in their games, as they think it will increase costs and hamper the entertainment that encourages consumers to buy their games (Pereira, De Souza, and De Menezes 2016).

Proposed frameworks:
- Some frameworks have been proposed to simplify tasks in serious game design, including: two game analytics frameworks for people with intellectual disabilities (García-Tejedor, Cano, and Fernández-Manjón 2016; Nguyen, Gardner, and Sheridan 2018), a game-based assessment model (Halverson and Owen 2014), a framework to integrate design of event-stream features for analysis (Owen and Baker 2018), a framework to support tracking and analysis of learners in-game activities (Hauge et al. 2014), a framework to help designers model experts’ solving process almost automatically (Muratet, Yessad, and Carron 2016), an interoperable adaptivity framework (Streicher and Roller 2017), a framework for internet-scale experiments to inform and be informed by classroom and lab experiments (Stamper et al. 2012), an open-source SGs framework for sustainability (Y. Xu et al. 2014) and a framework for a mobile game application for adults with cystic fibrosis (Vagg et al. 2018).
5. Discussion

We have found that most studies focus on assessment and learners’ behaviors (RQ1). This suggests that, having established that games are indeed a useful tool for purposes beyond entertainment, there is an interest in analyzing interaction data to measure how much impact serious games have on players (mainly focusing on learning), and how that impact relates to players’ in-game behaviors.

From our analysis of the methods used (RQ2), we found that visualization, supervised and unsupervised techniques are present in a similar number of papers. Among the data science techniques, the most widely used are linear models, correlations, and cluster techniques. All these methods are classical techniques, which may be surprising as newer, more complex and powerful techniques, in particular neural networks, are experiencing an important surge in popularity. A reason that may explain this result is the need of further evidence on how to widely and reliably apply these new complex techniques and the common difficulty to explain the results obtained, which has opened the debate about explainable AI (XAI) (Adadi & Berrada, 2018).

The main stakeholders considered, in a similar number of studies, are game designers/developers, and researchers, followed by teachers/educators (RQ3). This suggests that the analysis of data from games is used for several purposes including research, improving or validating game design, and providing information when applying games in educational scenarios. Although students only appear in 8 of the papers as the main direct stakeholders to benefit from the results of data analysis, they are always indirect recipients of the results, as the research, improvement and adaptivity of games and assessment techniques will make the use of games more effective and efficient for the ones who actually play the games, that is, the students/learners.

From the general information of the studies, we have found that most of the games used in the studies teach science-related topics, in particular mathematics. This result shows the intention to benefit from games’ advantages to improve learning in a subject typically considered difficult for children and adolescents, and aligns with previous research which found that mathematics and science were the main areas for games that target primary education (Hainey, Connolly, Boyle, Wilson, & Razak, 2016). This may also be related with the fact that these domains have a clearly defined underlying model that simplifies assessment.

Sample sizes used in the studies are, in general, quite low (32% of the studies used less than 100 participants). This can potentially restrict the significance and generalization of their results, as well as the application of more complex algorithms such as deep neural networks, which require large amounts of data points to be adequately applied. The low sample size used in experiments is an important issue already pointed out by authors (Petri & Gresse von Wangenheim, 2017). Most participants were from primary and secondary schools, which aligns with the fact that the most commonly used games aim to teach mathematics.

Data collected from students’ interactions included mainly completion times, actions or interactions in general, and scores. All these are common information that can be collected from any game but are, however, basic data that do not take full advantage of the rich interactions produced in games, as described in works on game analytics in entertainment games (Seif El-Nasr, Drachen, & Canossa, 2013). As some studies pointed out, the data to be collected is best identified at early stages of the game development, to ensure that it provides information with educational value. Most papers did not report the format in which they collected the data, so that we cannot know if they were using a standard or relying on their own data-formats. The latest scenario is less desirable, as it restricts the open sharing of the data for other purposes and requires an extra effort to replicate results with other techniques (Serrano-Laguna et al., 2017). We have not found reports of any open data set of game analytics data or learning analytics data from serious games; this hinders research in this area, as testing out new data science techniques requires not only choosing the techniques themselves, but also developing a serious game and performing the experiments to collect its interaction data.

The analysis of GLA data from serious games has yielded, as expected, wide and varied results (RQ4). We can, however, extract some general findings from the conclusions and discussions of the studies analyzed:
• **Predicting games impact with GLA data:** raw data can be used to accurately predict impact (e.g. learning), including simple values from interactions (e.g. completion times, scores) but also more complex information such as kind of failures or exploration strategies. Adding information of the context is also recommended, as it can improve the models’ accuracy. Also, the choice of data to analyze should ideally be taken during game design, to ensure that as much educationally-relevant data as possible is actually captured.

• **Importance of student profiling:** performance appears to be highly related to students’ characteristics and behaviors, so it is recommended to create students’ profiles or clusters to improve learning, including targeted feedback and adaptive learning experiences. The need to fit users’ needs has also led authors to propose user-specific frameworks (e.g. for users with intellectual disabilities).

• **Designing serious games for assessment:** assessment needs to be formally and reliably integrated in the development phase of SGs to provide meaningful educational information. This should not damage costs or entertainment, as games need to maintain engaging and motivation features, while controlling for an adequate difficulty. GLA data can then be used to validate the game design and assessment.

5.1. **Limitations**

As in any other literature review, our work is limited by the search terms used. To try to minimize this limitation, we have included similar and commonly interchanged terms as well as an additional search on a specific journal. However, there still can be some relevant works that have not matched our search criteria.

5.2. **Conclusions**

This paper presents a systematic literature review on the applications of data science to game analytics data and/or learning analytics data collected from serious games. The goal of this literature review was to find how data science techniques have been used with interaction data from serious games, as we consider that the application of data science can increase the still-limited application of serious games in education. Games have proven to be beneficial for learning in different domains, including more authentic learning and higher student engagement, both thanks to their greater degree of interactivity and immersion. Their limited application can be partly explained by a simple cost-benefit analysis: games are certainly costlier and more complex than other contents, and their advantages (for example, in terms of evaluating learning) are hard enough to measure that stakeholders may not be convinced of their overall return on investment (ROI). We consider that the information extracted from the application of data science techniques to interaction data from serious games can both reduce costs and complexity by simplifying game design and development, and measuring games’ actual impact so the benefits of applying games are clearer for all stakeholders.

On this systematic literature review, we have identified 87 papers that reported evidence of the outcomes of the analysis of game analytics data and/or learning analytics data collected from serious games. We have classified them according to their specific purposes, methods of analysis used, stakeholders that benefit from the information and conclusions drawn from the analysis. These classifications can be used as a baseline for further research related to analysis of data from serious games.

Despite the diversity of the studies, we have been able to extract some notable common points and conclusions. The main purpose when analyzing data from serious games is assessment, most commonly with linear prediction models, simple correlation or cluster techniques, or visually displaying performance information. Learning predictions obtained are quite accurate and may be improved with some of the previous recommendations. The importance of student profiling as well as recommendations for integrating assessment into early phases of game design and development also stand out among the conclusions of the studies. Further
research in this field should also bear in mind these recommendations to effectively and accurately assess students playing serious games.

Considering the studies presented in this review, we encourage researchers to consider large-enough sample sizes to ensure significant conclusions, and to decide in advance which data is to be collected from the games. In this sense, as a baseline, typical data such as completion times, interactions, or scores can and should be included; but research can benefit from moving on to more complex data extracted from in-game interactions. Regarding algorithms, we encourage researchers to compare classical techniques with new more complex ones (e.g. neural networks), to determine which ones draw the best results in each case. Finally, authors have pointed out a clear need for specific game learning analytics (GLA), where the use of standards to collect GLA data is desirable, as it allows the creation of open data sets in standard formats, such as xAPI (Serrano-Laguna et al., 2017), for research purposes, and simplifies results reproducibility and improvement, as well as testing of new techniques and integration of analytics as a module of larger systems.

Acknowledgements

This work has been partially funded by Regional Government of Madrid (eMadrid P2018/TC54307), by the Ministry of Education (TIN2017-89238-R) and by the European Commission (RAGE H2020-ICT-2014-1-644187, BEACONING H2020-ICT-2015-687676, Erasmus+ IMPRESS 2017-1-NL01-KA203-035259).

References


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). https://doi.org/10.1007/978-3-319-51645-5_24


Coded papers


Ke, F., & Shute, V. J. (2015). Serious Games Analytics. Serious Games Analytics, (January). https://doi.org/10.1007/978-3-319-05834-4


Kickmeier-Rust, M. D. (2018). Predicting Learning Performance in Serious Games. In S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, … P. Caserman (Eds.), Serious Games (pp. 133–144). https://doi.org/10.1007/978-3-030-02762-9_14


Ninaus, M., Kilii, K., Siegler, R. S., & Moeller, K. (2017). Data-Driven Design Decisions to Improve Game-Based Learning of Fractions. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10653. LNCS (pp. 3–13). https://doi.org/10.1007/978-3-319-71940-5_1


