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

Cristina Alonso-Fernández*, Antonio Calvo-Morata, Manuel Freire, Iván Martínez-Ortiz, Baltasar Fernández-Manjón

Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid, Spain

ARTICLE INFO

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.

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

---

* Corresponding author.
  E-mail address: calonsofernandez@ucm.es (C. Alonso-Fernández).

gathered from serious games to guide data-based decision-making (Loh, Sheng, & Ifenthaler, 2015a, 2015b). 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, Alonso-Fernández, Pérez-Colado, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019; Freire et al., 2016a, 2016b; 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.

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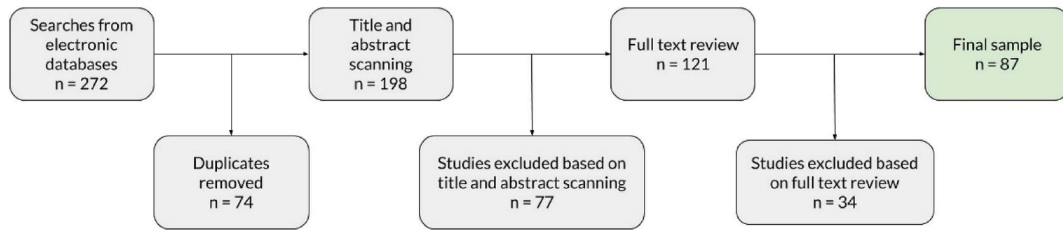
Fig. 1. Studies selection process for the systematic literature review.

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.

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

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.

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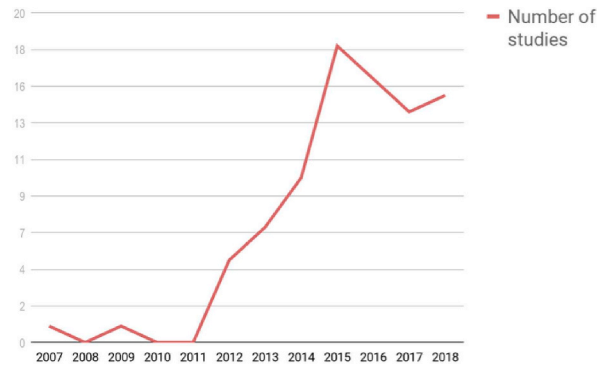
After considering all qualifying studies, we mapped their main data science application purpose into one of 5 categories: learning assessment, study of in-game behaviors, game design or evaluation, student profiling, and interventions. Some studies focus on more than one of the previous purposes. Since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

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.

Table 1

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

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

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

Table 2

Categories of purposes of data analysis application for the selected studies, with definition and two example studies.

| Purpose category	Definition of purpose	Example studies
Assessment	Assess learning, predict performance	(Ke & Shute, 2015; R. S.; Baker, Clarke-Midura, & Ocumpaugh, 2016)
In-game behaviors	Study in-game players behaviors (e.g. persistence, engagement)	(Dicerbo, 2013; Kang, Liu, & Qu, 2017)
Game design	Validate game design	(Cano, Fernández-Manjón, & García-Tejedor, 2018; Tlili, Essalmi, Jemni, & Kinshuk, 2016)
Student profiling	Stablish categories of players profiles, differentiate players characteristics	(Denden, Tlili, Essalmi, & Jemni, 2018; Loh & Sheng, 2014)
Interventions	Study effect of in-game interventions (e.g. feedback messages, notification of performance)	(DeFalco et al., 2018; McCarthy, Johnson, Likens, Martin, & McNamara, 2017)
Framework proposals	Propose a framework for specific contexts	(Halverson & Owen, 2014; Nguyen, Gardner, & Sheridan, 2018)
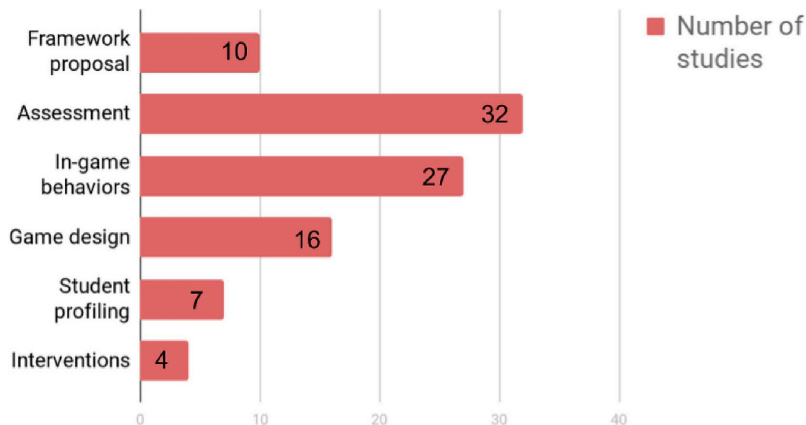
Fig. 3. Main purpose of data analysis in the selected studies. Some studies focus on several purposes.

Table 3
Data science techniques used in the studies, and number of papers that use each technique.

Data science technique	Number of papers using the technique
Supervised models	31
Linear/logistic regression	18
Regression/decision trees	7
Bayesian networks	6
Neural networks	4
Naïve Bayes	3
Bayesian knowledge tracing	3
Support vector machines	2
Unsupervised models	35
Correlation	17
Clustering	16
Factor analysis	2
Visualization	36
Performance metrics	15
Gameplay pathways	7
Use of in-game tools	5
Learning curves	4
Heatmaps of interactions	2

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[1], 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.

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).

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%)

---

[1] Plot of speed or accuracy over game levels (Martin, Koedinger, Mitrovic, & Mathan, 2005).
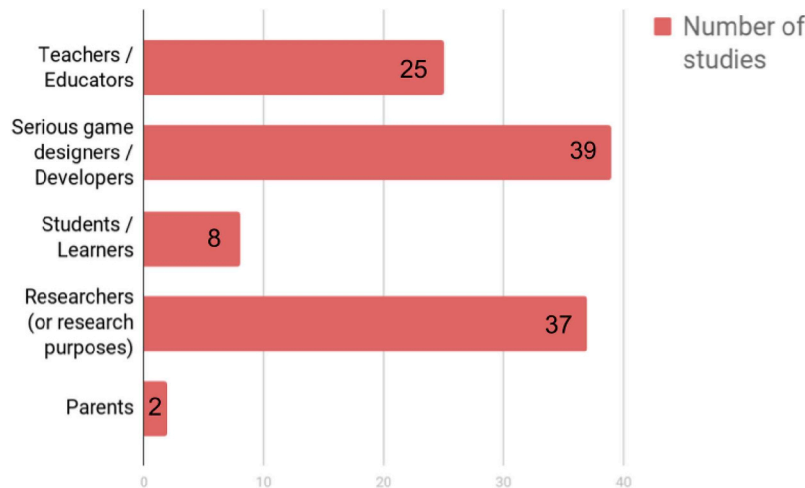
Fig. 4. Main stakeholder recipient of the analysis results in selected studies.

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 games 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/collected, 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, & Steiner, 2016), and for all stakeholders (Alonso-Fernández et al., 2019, Alonso-Fernández, Pérez-Colado, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019). The analysis of interaction data can provide a means to measure the proven positive impact of games (Kosmas, Ioannou, & Retalis, 2018; Mavridis, Katmada, & Tsiatsos, 2017). However, most data is still captured after the game (Smith, Blackmore, & Nesbitt, 2015). Authors also point out the need for specific game learning analytics (Freire et al., 2016a, 2016b), or so-called serious games analytics (Loh et al., 2015a, 2015b), 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 et al., 2016), as the achievement system built into games may not be the most informative indicator of learning (Heeter, Lee, Medler, & Magerko, 2013). Some papers point out that their best predictors for measuring learning are based on the analysis of the player's exploration strategies (Horn et al., 2016; Kang et al., 2017; Käser, Hallinen, & Schwartz, 2017; Owen, Anton, & Baker, 2016; Smith, Hickmott, Southgate, Bille, & Stephens, 2016), or on the characteristics of player failures (Halverson & Owen, 2014). One study explored the relation between learning and students' facial behavior (Z. Xu & 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, & 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, Xie, Rohrer, & Prince, 2014), while team cohesion and psychological safety may be good performance indicators in multiplayer serious games (Mayer, van Dierendonck, van Ruijven, & Wenzler, 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, & 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, Habgood, Ainsworth, & Corbett, 2007).

- Recommendations to improve predictions: feature engineering improves performance of models with simple raw data (Owen & 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, & McNamara, 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, & Albert, 2015). Complex assessment conditions can be created by combining some of the basic sets of traces (Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 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, Elouaai, Elaachak, Yedri, & Bouhorma, 2018; Freitas & Gibson, 2014; Forsyth et al., 2012; Lazo, Anareta, Duremdes, & Red, 2018; Polyak, von Davier, & Peterschmidt, 2017; Martinez-Garza & Clark, 2017; Chung, 2015) and in-game choices (Cutumisu, Blair, Chin, & Schwartz, 2017), which are also related to prior knowledge (Martinez-Garza & Clark, 2017). Some studies explore methods to differentiate experts from novice users (Loh and Sheng 2014, 2015a, 2015b). Once students are classified in a performance group, scores can be inferred when time or action sequences are added to the analysis (Gibson & Clarke-Midura, 2015). Tactics that lead to success can also be discovered with cluster analysis (Sharples & Domingue, 2016).

- Importance of understanding learners' characteristics: students with different learning characteristics may exhibit different learning behaviors (Liu, Lee, Kang, & Liu, 2016), for instance, different exploration strategies (Martin et al., 2013) or, in some cases, their age and gender (Wallner & Kriglstein, 2015). It is key to model students for effective adaptive instruction (Koedinger, McLaughlin, & 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, Shores, Mott, & Lester, 2013) and high-performance students tend to use tools more appropriately (Liu, Kang, Lee, Winzeler, & Liu, 2015). Behaviors also depend on student background (Jaccard, Hulaas, & 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, & 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 (Ketamo, 2015; Roberts, Chung, & Parks, 2016).

- 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 & 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 et al., 2018; Harpstead, MacLellan, Aleven, & Myers, 2015; Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 2012; Tlili et al., 2016), specific design choices (Cheng, Rosenheck, Lin, & Klopfer, 2017) and even to create new game mechanics (Ninaus, Kiili, Siegler, & Moeller, 2017) or to automatically discover speech act categories for dialogue-based educational games (Rus, Moldovan, Niraula, & Graesser, 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 & Shute, 2015). One study proposes a design approach to integrate data-driven assessment in game design (Ke, Shute, Clark, & Erlebacher, 2019). Debriefing via visualizations can improve understanding of outcomes (De Troyer, Helalouch, & 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, Silvervarg, Gulz, & Haake, 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 & 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 et al., 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 & 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, & 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, & Fernández-Manjón, 2016; Nguyen et al., 2018), a game-based assessment model (Halverson & Owen, 2014), a framework to integrate design of event-stream features for analysis (Owen & 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, & Carron, 2016), an interoperable adaptivity framework (Streicher & 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, Johnson, Lee, Moore, & Brewer, 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, Manero, Freire, & Fernández-Manjón, 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

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

ADL. (2012). *Experience API.* Retrieved March 20, 2016, from https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/.

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019a). 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.

Berta, R., & Moreno-Ger, P. (2018). Introduction: Intelligent learning assessment in serious games. *International Journal of Serious Games, 5*(1)https://doi.org/10.17083/ijsg.v5i1.237.

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.

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016a). Game learning analytics: Learning analytics for serious games. *Learning, Design, and Technology,* 1–29. https://doi.org/10.1007/978-3-319-17727-4_21-1.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education, 102*, 202–223. https://doi.org/10.1016/j.compedu.2016.09.001.

Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning analytics as an assessment tool in serious games: A review of literature. *Serious games and edutainment applications* (pp. 537–563). . https://doi.org/10.1007/978-3-319-51645-5_24.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015a). In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.). *Serious games analytics*https://doi.org/10.1007/978-3-319-05834-4.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review,* 31–40.

Martin, B., Koedinger, K., Mitrovic, A., & Mathan, S. (2005). On using learning curves to evaluate ITS. *International Conference on Artificial Intelligence in Education,* 419–426.

Owen, E., & Baker, R. (2019). *Learning analytics for serious games.* (February). Retrieved from http://www.galanoe.eu/index.php/home/365-learning-analytics-for-serious-games.

Petri, G., & Gresse von Wangenheim, C. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & Education, 107*, 68–90. https://doi.org/10.1016/j.compedu.2017.01.004.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532.

Seif El-Nasr, M., Drachen, A., & Canossa, A. (2013). In A. Drachen, & A. Canossa (Eds.). *Game analytics (M. Seif el-Nasr*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.

Wallner, G., Canossa, A., & El-Nasr, M. S. (2018). Introduction to the special issue on visual game analytics. *Information Visualization, 17*(3), 181–182. https://doi.org/10.1177/1473871617722040.

Yannakakis, G. N., & Togelius, J. (2018). *Artificial intelligence and games.* https://doi.org/10.1007/978-3-319-63519-4.

Coded papers

Alonso-Fernández, C., Pérez-Colado, I., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019b). Improving serious games analyzing learning analytics data: Lessons learned. *Games and learning alliance: 7th international conference, GALA 2018, Palermo, Italy, December 5–7, 2018, proceedings: Vol. 10653*, (pp. 287–296). . https://doi.org/10.1007/978-3-030-11548-7_27.

Baker, R. S., Clarke-Midura, J., & Ocumpaugh, J. (2016). Towards general models of effective science inquiry in virtual performance assessments. *Journal of Computer Assisted Learning, 32*(3), 267–280. https://doi.org/10.1111/jcal.12128.

Baker, R. S. J. D., Habgood, M. P. J., Ainsworth, S. E., & Corbett, A. T. (2007). Modeling the acquisition of fluent skill in educational action games. *User modeling 2007* (pp. 17–26). . https://doi.org/10.1007/978-3-540-73078-1_5.

Cano, A. R., Fernández-Manjón, B., & García-Tejedor, Á. J. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology, 49*(4), 659–672. https://doi.org/10.1111/bjet.12632.

Cheng, M.-T., Rosenheck, L., Lin, C.-Y., & Klopfer, E. (2017). Analyzing gameplay data to inform feedback loops in the Radix Endeavor. *Computers & Education, 111*, 60–73. https://doi.org/10.1016/j.compedu.2017.03.015.

Chung, G. K. W. K. (2015). Guidelines for the design and implementation of game telemetry for serious games analytics. *Serious games analytics* (pp. 59–79). . https://doi.org/10.1007/978-3-319-05834-4_3.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2017). Assessing whether students seek constructive criticism: The design of an automated feedback system for a graphic design task. *International Journal of Artificial Intelligence in Education, 27*(3), 419–447. https://doi.org/10.1007/s40593-016-0137-5.

De Troyer, O., Helalouch, A., & Debruyne, C. (2016). Towards computer-supported self-debriefing of a serious game against cyber bullying. In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.). *Games and learning alliance: 6th international conference, GALA 2017, Lisbon, Portugal, December 5–7, 2017, proceedings* (pp. 374–384). . https://doi.org/10.1007/978-3-319-50182-6_34.

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., et al. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education, 28*(2), 152–193. https://doi.org/10.1007/s40593-017-0152-1.

Denden, M., Tlili, A., Essalmi, F., & Jemni, M. (2018). Implicit modeling of learners' personalities in a game-based learning environment using their gaming behaviors. *Smart Learning Environments, 5*(1), 1–19. https://doi.org/10.1186/s40561-018-0078-6.

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

DiCerbo, K. E., Bertling, M., Stephenson, S., Jia, Y., Mislevy, R. J., Bauer, M., et al. (2015). An application of exploratory data analysis in the development of game-based assessments. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.). *Serious games analytics* (pp. 319–342). . https://doi.org/10.1007/978-3-319-05834-4_14.

Eagle, M. (2009). Level Up : A frame work for the design and evaluation of educational games. *Artificial Intelligence, 434*(1), 339–341. https://doi.org/10.5507/bp.2011.016.

Elaachak, L., Belahbibe, A., & Bouhorma, M. (2015). Towards a system of guidance, assistance and learning analytics based on multi agent system Applied on serious games. *International Journal of Electrical and Computer Engineering (IJECE) Journal, 5*(2), 2088–8708. Retrieved from http://iaesjournal.com/online/index.php/IJECE.

Forsyth, C., Pavlik, P., Graesser, A., Cai, Z., Germany, M.-L., Millis, K., et al. (2012). Learning gains for core concepts in a serious game on scientific reasoning. *Proceedings of the 5th international conference on educational data mining, 1–4*. Retrieved from http://w.optimallearning.org/people/Articles/edm2012_short_2.pdf.

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016b). Game learning analytics: Learning analytics for serious games. *Learning, Design, and Technology, 1–29*. https://doi.org/10.1007/978-3-319-17727-4_21-1.

Freitas, S. de, & Gibson, D. (2014). Exploratory learning analytics methods from three case studies. Rhetoric and Reality: Critical Perspectives on Educational Technology. *Proceedings of Ascilite Dunedin, 2014, 383–388*.

García-Tejedor, Á. J., Cano, A. R., & Fernández-Manjón, B. (2016). GLAID: Designing a game learning analytics model to analyze the learning process in users with intellectual disabilities. *6th EAI international conference on serious games, interaction and simulation*. Retrieved from http://sgamesconf.org/2016/show/technical-session.

Ghergulescu, I., & Muntean, C. H. (2016). ToTCompute: A novel EEG-based TimeOnTask threshold computation mechanism for engagement modelling and monitoring. *International Journal of Artificial Intelligence in Education, 26*(3), 821–854. https://doi.org/10.1007/s40593-016-0111-2.

Gibson, D., & Clarke-Midura, J. (2015). Some psychometric and design implications of game-based learning analytics. E-Learning systems, environments and approaches. CELDA247–261. https://doi.org/10.1007/978-3-319-05825-2_17.

Gweon, G.-H., Lee, H.-S., Dorsey, C., Tinker, R., Finzer, W., & Damelin, D. (2015). Tracking student progress in a game-like learning environment with a Monte Carlo Bayesian knowledge tracing model. *Proceedings of the fifth international conference on learning analytics and knowledge - LAK '15* (pp. 166–170). . https://doi.org/10.1145/2723576.2723608.

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

Harpstead, E., MacLellan, C. J., Aleven, V., & Myers, B. A. (2015). Replay analysis in open-ended educational games. *Serious games analytics* (pp. 381–399). . https://doi.org/10.1007/978-3-319-05834-4_17.

Hauge, J. B., Berta, R., Fiucci, G., Manjon, B. F., Padron-Napoles, C., Westra, W., et al. (2014). Implications of learning analytics for serious game design. *2014 IEEE 14th international conference on advanced learning technologies* (pp. 230–232). . https://doi.org/10.1109/ICALT.2014.73.

Heeter, C., Lee, Y.-H., Medler, B., & Magerko, B. (2013). Conceptually meaningful metrics: Inferring optimal challenge and mindset from gameplay. *Game analytics* (pp. 731–762). . https://doi.org/10.1007/978-1-4471-4769-5_32.

Hernández-Lara, A. B., Perera-Lluna, A., & Serradell-López, E. (2019). Applying learning analytics to students' interaction in business simulation games. The usefulness of learning analytics to know what students really learn. *Computers in Human Behavior, 92*, 600–612. https://doi.org/10.1016/j.chb.2018.03.001.

Hicks, D., Eagle, M., Rowe, E., Asbell-Clarke, J., Edwards, T., & Barnes, T. (2016). Using game analytics to evaluate puzzle design and level progression in a serious game. *Proceedings of the sixth international conference on learning analytics & knowledge - LAK '16* (pp. 440–448). . https://doi.org/10.1145/2883851.2883953.

Horn, B., Hoover, A. K., Barnes, J., Folajimi, Y., Smith, G., & Harteveld, C. (2016). Opening the black box of play. *Proceedings of the 2016 annual symposium on computer-human interaction in play - CHI PLAY '16* (pp. 142–153). . https://doi.org/10.1145/2967934.2968109.

Jaccard, D., Hulaas, J., & Dumont, A. (2017). In J. Dias, P. A. Santos, & R. C. Veltkamp (Eds.). *Using comparative behavior analysis to improve the impact of serious games on students' learning experience*https://doi.org/10.1007/978-3-319-71940-5.

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior, 72*, 757–770. https://doi.org/10.1016/j.chb.2016.09.062.

Käser, T., Busetto, A. G., Solenthaler, B., Baschera, G. M., Kohn, J., Kucian, K., et al. (2013). Modelling and optimizing mathematics learning in children. *International Journal of Artificial Intelligence in Education, 23*(1–4), 115–135. https://doi.org/10.1007/s40593-013-0003-7.

Käser, T., Hallinen, N. R., & Schwartz, D. L. (2017). Modeling exploration strategies to predict student performance within a learning environment and beyond. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17* (pp. 31–40). . https://doi.org/10.1145/3027385.3027422.

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

Ke, F., Shute, V., Clark, K. M., & Erlebacher, G. (2019). *Interdisciplinary design of game-based learning platforms*. https://doi.org/10.1007/978-3-030-04339-1.

Ketamo, H. (2013). Agents and analytics - a framework for educational data mining with games based learning. *Proceedings of the 5th international conference on agents and artificial intelligence* (pp. 377–382). . https://doi.org/10.5220/0004331403770382.

Ketamo, H. (2015). User-generated character behaviors in educational games. *Healthcare Informatics Research, 21*, 57–68. https://doi.org/10.1007/978-981-287-408-5_5.

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, (Eds.). *Serious games* (pp. 133–144). . https://doi.org/10.1007/978-3-030-02762-9_14.

Koedinger, K., McLaughlin, E., & Stamper, J. (2012). Automated student model improvement. *Proceedings of the 5th international conference on educational data mining*

(pp. 17–24). . ISBN: 978-1-74210-276-4. Available at: http://educationaldatamining.org/EDM2012/uploads/procs/EDM_2012_proceedings.pdf.

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends, 62*(6), 594–601. https://doi.org/10.1007/s11528-018-0294-5.

Lazo, P. P. L., Anareta, C. L. Q., Duremdes, J. B. T., & Red, E. R. (2018). Classification of public elementary students' game play patterns in a digital game-based learning system with pedagogical agent. *Proceedings of the 6th international conference on information and education technology - ICIET '18* (pp. 75–80). . https://doi.org/10.1145/3178158.3178160.

Liu, M., Kang, J., Lee, J., Winzeler, E., & Liu, S. (2015). Examining through visualization what tools learners access as they play a serious game for middle school science. *Serious games analytics* (pp. 181–208). . https://doi.org/10.1007/978-3-319-05834-4_8.

Liu, M., Lee, J., Kang, J., & Liu, S. (2016). What we can learn from the data: A multiple-case study examining behavior patterns by students with different characteristics in using a serious game. *Technology, Knowledge and Learning, 21*(1), 33–57. https://doi.org/10.1007/s10758-015-9263-7.

Loh, C. S., & Sheng, Y. (2014). Maximum similarity index (MSI): A metric to differentiate the performance of novices vs. multiple-experts in serious games. *Computers in Human Behavior, 39*, 322–330. https://doi.org/10.1016/j.chb.2014.07.022.

Loh, C. S., & Sheng, Y. (2015a). Measuring expert performance for serious games analytics: From data to insights. *Serious games analytics* (pp. 101–134). . https://doi.org/10.1007/978-3-319-05834-4_5.

Loh, C. S., & Sheng, Y. (2015b). Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies, 20*(1), 5–19. https://doi.org/10.1007/s10639-013-9263-y.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015b). Serious games analytics: Theoretical framework. *Serious games analytics* (pp. 3–29). . https://doi.org/10.1007/978-3-319-05834-4_1.

Martin, T., Aghababyan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., et al. (2013). Nanogenetic learning analytics. *Proceedings of the third international conference on learning analytics and knowledge - LAK '13* (pp. 165). . https://doi.org/10.1145/2460296.2460328.

Martinez-Garza, M. M., & Clark, D. B. (2017). Investigating epistemic stances in game play with data mining. *International Journal of Gaming and Computer-Mediated Simulations, 9*(3), 1–40. https://doi.org/10.4018/ijgcms.2017070101.

Martin, T., Petrick Smith, C., Forsgren, N., Aghababyan, A., Janisiewicz, P., & Baker, S. (2015). Learning fractions by splitting: Using learning analytics to illuminate the development of mathematical understanding. *The Journal of the Learning Sciences, 24*(4), 593–637. https://doi.org/10.1080/10508406.2015.1078244.

Mavridis, A., Katmada, A., & Tsiatsos, T. (2017). Impact of online flexible games on students' attitude towards mathematics. *Educational Technology Research & Development, 65*(6), 1451–1470. https://doi.org/10.1007/s11423-017-9522-5.

Mayer, I., van Dierendonck, D., van Ruijven, T., & Wenzler, I. (2014). Stealth assessment of teams in a digital game environment. *Lecture Notes in Computer Science, 8605*, 224–235. https://doi.org/10.1007/978-3-319-12157-4_18.

McCarthy, K. S., Johnson, A. M., Likens, A. D., Martin, Z., & McNamara, D. S. (2017). *Metacognitive prompt overdose: Positive and negative effects of prompts in iSTART. Grantee submission.* 404–405 Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED577125&site=ehost-live.

Muratet, M., Yessad, A., & Carron, T. (2016). Understanding learners' behaviors in serious games. In F. W. B. Li, R. Klamma, M. Laanpere, J. Zhang, B. F. Manjón, & R. W. H. Lau (Eds.). *Advances in web-based learning - ICWL 2015* (pp. 195–205). . https://doi.org/10.1007/978-3-319-47440-3_22.

Nguyen, A., Gardner, L. A., & Sheridan, D. (2018). A framework for applying learning analytics in serious games for people with intellectual disabilities. *British Journal of Educational Technology, 49*(4), 673–689. https://doi.org/10.1111/bjet.12625.

Ninaus, M., Kiili, K., Siegler, R. S., & Moeller, K. (2017). Data-driven design decisions to improve game-based learning of fractions. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics): Vol. 10653LNCS.* (pp. 3–13) https://doi.org/10.1007/978-3-319-71940-5_1.

Owen, V. E., Anton, G., & Baker, R. (2016). Modeling user exploration and boundary testing in digital learning games. *Proceedings of the 2016 conference on user modeling adaptation and personalization - UMAP '16* (pp. 301–302). . https://doi.org/10.1145/2930238.2930271.

Owen, V. E., & Baker, R. S. (2018). Fueling prediction of player decisions: Foundations of feature engineering for Optimized behavior modeling in serious games. *Technology, Knowledge and Learning*0123456789. https://doi.org/10.1007/s10758-018-9393-9.

Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education, 24*(3), 251–283. https://doi.org/10.1007/s40593-014-0018-8.

Peddycord-Liu, Z., Harred, R., Karamarkovich, S., Barnes, T., Lynch, C., & Rutherford, T. (2018). In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.). *Learning curve analysis in a large-scale, drill-and-practice serious math game: Where is learning support needed?* (pp. 436–449). . https://doi.org/10.1007/978-3-319-93843-1_32.

Pereira, H. A., De Souza, A. F., & De Menezes, C. S. (2016). A computational architecture for learning analytics in game-based learning. *Proceedings - IEEE 16th international conference on advanced learning technologies, ICALT 2016* (pp. 191–193). . https://doi.org/10.1109/ICALT.2016.3.

Plass, J. L., Homer, B. D., Kinzer, C. K., Chang, Y. K., Frye, J., Kaczetow, W., et al. (2013). Metrics in simulations and games for learning. *Game analytics* (pp. 697–729). . https://doi.org/10.1007/978-1-4471-4769-5_31.

Polyak, S. T., von Davier, A. A., & Peterschmidt, K. (2017). Computational psychometrics for the measurement of collaborative problem solving skills. *Frontiers in Psychology, 8*(NOV), 1–16. https://doi.org/10.3389/fpsyg.2017.02029.

Roberts, J. D., Chung, G. K. W. K., & Parks, C. B. (2016). Supporting children's progress through the PBS KIDS learning analytics platform. *Journal of Children and Media, 10*(2), 257–266. https://doi.org/10.1080/17482798.2016.1140489.

Rowe, E., Asbell-clarke, J., & Baker, R. S. (2015). Serious games analytics to measure implicit science learning. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.). *Serious games analytics* (pp. 343–360). . https://doi.org/10.1007/978-3-319-05834-4.

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., et al. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior, 76*, 617–630. https://doi.org/10.1016/j.chb.2017.03.043.

Rus, V., Moldovan, C., Niraula, N., & Graesser, A. (2012). Automated discovery of speech act categories in educational games. *Proceedings of International Conference on Educational Data Mining,* 25–32.

Sabourin, J. L., Shores, L. R., Mott, B. W., & Lester, J. C. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education, 23*(1–4), 94–114. https://doi.org/10.1007/s40593-013-0004-6.

Serrano-Laguna, Á., Manero, B., Freire, M., & Fernández-Manjón, B. (2018). A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. *Multimedia Tools and Applications, 77*(2), 2849–2871. https://doi.org/10.1007/s11042-017-4467-6.

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2012). Tracing a little for big improvements: Application of learning analytics and videogames for student assessment. *Procedia Computer Science, 15,* 203–209 Elsevier.

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of learning analytics in educational videogames. *Entertainment Computing, 5*(4), 313–322. https://doi.org/10.1016/j.entcom.2014.02.003.

Sharples, M., & Domingue, J. (2016). *Adaptive and adaptable learning.* Switzerland: Lecture Notes in Computer Science13–16. 9891 https://doi.org/10.1007/978-3-319-45153-4.

Slimani, A., Elouaai, F., Elaachak, L., Yedri, O. B., & Bouhorma, M. (2018). Learning analytics through serious games: Data mining algorithms for performance measurement and improvement purposes. *International Journal of Emerging Technologies in Learning, 13*(1), 46–64. https://doi.org/10.3991/ijet.v13i01.7518.

Smith, S. P., Blackmore, K., & Nesbitt, K. (2015). A meta-analysis of data collection in serious games research. *Serious games analytics* (pp. 31–55). . https://doi.org/10.1007/978-3-319-05834-4_2.

Smith, S. P., Hickmott, D., Southgate, E., Bille, R., & Stephens, L. (2016). Exploring play-learners' analytics in a serious game for literacy improvement. In T. Marsh, M. Ma, M. F. Oliveira, J. Baalsrud Hauge, & S. Göbel (Eds.). *Serious games* (pp. 13–24). . https://doi.org/10.1007/978-3-319-45841-0_2.

Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). The dynamical analysis of log data within educational games. *Serious games analytics* (pp. 81–100). . https://doi.org/10.1007/978-3-319-05834-4_4.

Stamper, J. C., Lomas, D., Ching, D., Ritter, S., Koedinger, K. R., & Steinhart, J. (2012). The rise of the super experiment. *Proceedings of the 5th International Conference on Educational Data Mining,* 196–199. https://doi.org/10.1177/0003122412458508.

Steiner, C. M., Kickmeier-Rus, M. D., & Albert, D. (2015). Making sense of game-based user data: Learning analytics in applied games. *International conference E-learning* (pp. 195–198). . https://doi.org/10.1017/CBO9781107415324.004.

Streicher, A., & Roller, W. (2017). Interoperable adaptivity and learning analytics for serious games in image interpretation. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics): Vol. 10474*, (pp. 598–601). LNCS. https://doi.org/10.1007/978-3-319-66610-5_71.

Streicher, A., & Smeddinck, J. D. (2016). In R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, & K. Zweig (Eds.). *Personalized and adaptive serious games* (pp. 332–377). Springer. https://doi.org/10.1007/978-3-319-46152-6_14.

Tärning, B., Silvervarg, A., Gulz, A., & Haake, M. (2018). Instructing a teachable agent with low or high self-efficacy – does similarity attract? *International Journal of Artificial Intelligence in Education, 29*(1), 89–121. https://doi.org/10.1007/s40593-018-0167-2.

Tellioglu, U., Xie, G. G., Rohrer, J. P., & Prince, C. (2014). Whale of a crowd: Quantifying the effectiveness of crowd-sourced serious games. *2014 computer games: AI, animation, mobile, multimedia, educational and serious games (CGAMES)* (pp. 1–7). . https://doi.org/10.1109/CGames.2014.6934151.

Tlili, A., Essalmi, F., Jemni, M., & Kinshuk (2016). An educational game for teaching computer architecture: Evaluation using learning analytics. *2015 5th international conference on information and communication technology and accessibility, ICTA 2015*https://doi.org/10.1109/ICTA.2015.7426881.

Vagg, T., Tan, Y. Y., Shortt, C., Hickey, C., Plant, B. J., & Tabirca, S. (2018). MHealth and serious game analytics for cystic fibrosis adults. *2018 IEEE 31st international symposium on computer-based medical systems (CBMS), 2018-June* (pp. 100–105). . https://doi.org/10.1109/CBMS.2018.00025.

Wallner, G., & Kriglstein, S. (2015). Comparative visualization of player behavior for serious game analytics. *Serious games analytics* (pp. 159–179). . https://doi.org/10.1007/978-3-319-05834-4_7.

Wiemeyer, J., Kickmeier-Rust, M., & Steiner, C. M. (2016). Performance assessment in serious games. *Serious Games, 13*, 273–302. https://doi.org/10.1007/978-3-319-40612-1_10.

Xu, Y., Johnson, P. M., Lee, G. E., Moore, C. A., & Brewer, R. S. (2014). *Makahiki : An open source serious game framework for sustainability education and conservation, Vol. 8.* International Association for Development of the Information Society.

Xu, Z., & Woodruff, E. (2017). Person-centered approach to explore learner's emotionality in learning within a 3D narrative game. *Proceedings of the seventh international learning analytics & knowledge conference on - LAK '17* (pp. 439–443). . https://doi.org/10.1145/3027385.3027432.