

## Discovery of Research Gaps in Game Learning Analytics Applications on Learning Processes

Gabriel Silva<sup>1,4</sup>, Rodrigo L. Rodrigues<sup>1,2</sup>, Amadeu Sá C. Filho<sup>3</sup>, Américo Amorim<sup>4</sup>

<sup>1</sup>Programa de Pós-Graduação em Ensino de Ciências (PPGEC/UFRPE)

<sup>2</sup>Departamento de Educação, Universidade Federal Rural de Pernambuco (UFRPE)

<sup>3</sup>Departamento de Medicina Clínica, Centro de Ciências Médicas (UFPE)

<sup>4</sup>Escribo Inovação para o Aprendizado

{gabccandidods, rodrigomuribec, amadeu.campos}@gmail.com,  
americo@escribo.com

**Abstract.** *The benefits of game applications in the field of education have been increasingly evident due to the use of Game Learning Analytics, which is the collection, analysis and extraction of information from data obtained through Serious Games. In this article, we aim to carry out a systematic mapping in order to identify research gaps in studies that use Game Learning Analytics to analyze an educational process, analyzing the observed educational phenomenon, the algorithms or techniques used and the characteristics and number of participants. We understand that even though the Game Learning Analytics area is still very new, it has enormous potential in helping what we value most, teaching.*

### 1. Introduction

Video games are already one of the biggest industries in the entertainment industry, and all the power they have in engaging and immersing players draws a lot of attention, generating diverse ideas and research that seek to apply their use in several areas, including education.

From the junction between games and education, the term Serious Games emerges, used to define games that have a purpose beyond entertainment, which includes educational games, training simulation games, advergames (games created to promote a product, a service or a company), among others [Michael & Chen, 2006].

In parallel to these events in the games industry, we have the use of Data Science spreading through the most diverse areas and with multiple objectives. The application of data science techniques blend perfectly with interactive digital environments, and in education we already have the fields of Learning Analytics (LA) and Educational Data Mining (EDM) that seek to understand students and their environment to improve the process of learning through data analysis [Long & Siemens, 2011; Romero & Ventura, 2010].

In order to better understand how players behave, find errors and improve gameplay, the game industry applies analyzes using data science techniques in its research and development, thus creating the field of Game Analytics [Seif El-Nasr, Drachen , & Canossa, 2013].

When creating a Serious Game, and using the educational objectives of the Learning Analytics area combined with the Game Analytics tools and techniques, a new study area appears, called Game Learning Analytics, or just GLA. This study area is defined as the collection, analysis and extraction of information from data obtained through Serious Games [Alonso-Fernández et al., 2019].

From the different elements that make up the use of GLA, other forms are opened for different types of analysis, which can range from the evaluation of the educational impact of Serious Games to their design and validation. In addition, GLA has great potential to compare and contrast the educational results obtained by different groups, which can be formed from different factors, such as socioeconomic status, cultural differences or game profiles (which includes habits and behavior of the players) , among others [Freire, M. et al., 2016].

Therefore, this research aims to conduct a systematic mapping, in search of identifying research gaps, on the applications of Game Learning Analytics that sought to analyze an educational process.

The rest of the research is structured as follows: Section 2 discusses the concept of Game Learning Analytics and its importance in the student's learning process; Section 3 describes the methodology used for this systematic mapping; Section 4 presents the results and discussions on the subject; and, in section 5, the conclusions of this mapping are made.

## **2. The importance of Game Learning Analytics**

The games have always had a tendency to be used to improve the mind and mediate the process of learning or training, such as military strategy chess. But it was in the 1980s, with the advent of authoring systems (eg, Flash) that educators started creating games to teach [Loh, C.S. et al, 2015].

There are several attempts to define the term Serious Games, but in summary, Serious Games are “digital games created without entertainment as the main objective, but with the intention of serious use in training, education, or health care” [Loh, CS et al, 2015]. From this definition, the following question arises: "How can we guarantee the effectiveness of these games in a study?".

If we want the educational authorities accept and legitimize the serious games as an important learning tool, it will be necessary that the process of collecting and analyzing information to be taken seriously [Crookall, D., 2010]. As Smith, S.P. et al, (2015) tells in his work, good data analysis depends on the combination of the strengths of using Big Data and data mining, but information does not necessarily mean evidence.

The accuracy of the information will depend on the validity and quality of the data and the filtering, abstraction and simplification processes. If the data collected is of

poor quality, all subsequent processes will be compromised and the attempt to validate any identified pattern will be weakened [Smith, S.P. et al, 2015].

The Foundation of American Scientists (2006), understands that a game-based learning process, better known by the term “Game-Based Learning”, has the following characteristics: clear objectives, repeatable tasks, monitoring student progress, encouraging increase task time (through motivation), and adjusting the difficulty to suit the student's learning progress.

Understanding the various characteristics necessary for Game-Based Learning to occur, it is essential to search for concrete evidence of their effectiveness, since studies based on data obtained only through questionnaires, participants' statements or pen and paper tests are not well convincing [Loh, CS et al, 2015].

To demonstrate the effectiveness of this process, the areas of Learning Analytics and Game Analytics come together, forming the area of Game Learning Analytics. The methods in a Game Learning Analytics process are used for different purposes, but the main ones are: 1) Evaluate students based on data obtained from Serious Games; 2) Predict student learning based on their interaction with the game; 3) Validate the effectiveness of the game in student learning and; 4) identify possible improvements in the game's design [Alonso-Fernández, C. et al, 2019].

### **3. Method**

#### **3.1. Research Questions**

The main objective of this systematic mapping was to explore the applications of Game Learning Analytics in the student's learning process. For this purpose, the following central research questions were defined:

- **Q1:** What educational phenomena does the research intend to analyze?
- **Q2:** What techniques / algorithms are used to analyze the data collected?
- **Q3:** What is the educational level of the participants and the sample size?

Additionally, to complement the results, information was also collected on the educational areas of the games, year of publication and conclusions presented in the research.

#### **3.2. Data Collection**

The Google Scholar research platform was used to obtain the studies. Choose this justified by the fact that this repository already aggregates the works carried out on other platforms, such as IEEE Digital Library, ScienceDirect, SpringerLink, among others.

To conduct the survey, we used a string, which indicates the keywords and search specifications to be held. The words “game” and “analytics” were defined as mandatory, and in addition to them, a word more related to education must be found. Below is the search string used, note that the Google Scholar searcher doesn't need the AND instruction.

allintitle: (“game” AND “analytics”) AND (“school” OR “education” OR “learning” OR “teaching” OR “students”)

The search results for these keywords were restricted to the title, for searches published between 2010 to 2020 and the options "include patents" and "include citations" were unchecked, thus returning 64 results.

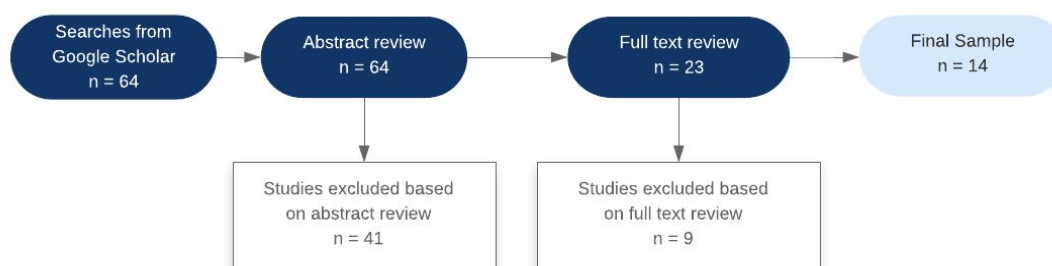
### 3.3. Data analysis

After obtaining the initial set of works, the exclusion criteria were defined and applied to remove studies that do not fall within the scope defined for the mapping.

#### Exclusion Criteria

- The article does not collect students' behavioral characteristics, nor does it aim to analyze this aspect;
- Game Learning Analytics is used for purposes other than identifying characteristics of the learning processes;
- The article is not part of the “education” domain;
- Is it a systematic review, book, or is it written in a language other than English

The study selection process for this systematic mapping can be seen in Figure 1.



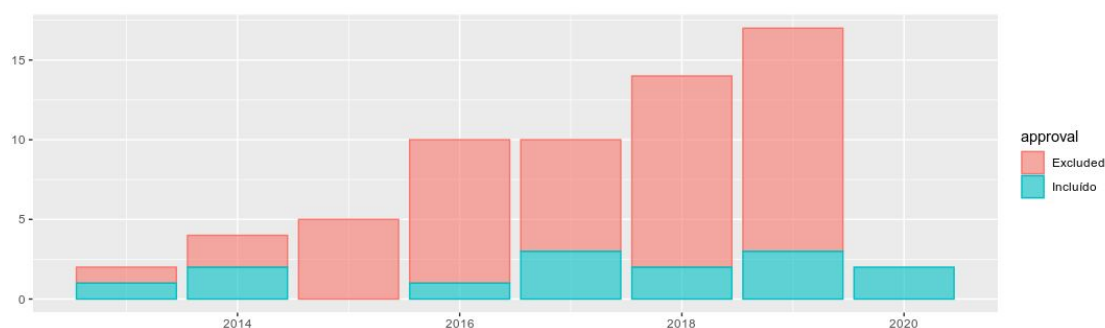
**Figure 1. Selection process for studies of systematic mapping**

We analyzed all 64 abstracts, comparing them to the exclusion criteria described above, and after that, a full-text analysis was performed with the research questions in mind, to ensure that the final sample would meet defined specifications. One of the limitations of this research, with regard to the data collection and analysis process, was the absence of the peer verification step.

## 4. Results and discussion

The distribution of the works returned by the research is shown in Figure 2. It is possible to notice a clear increase in the number of publications over the years, thus showing an increasing interest in the area of Game Learning Analytics since the year 2013.

Regarding the researches that specifically meet the criteria of this systematic mapping, understood by the blue color in the graph, they have shown themselves stable over the years, even so with some increase.



**Figure 2. Distribution, by year, of the works returned by the search criteria.**

At the end of the abstract and full text analysis process, a final sample of 14 selected works was reached.

authors	type	year
Callaghan, MJ., McShane, N. & Gómez Eguíluz, A.	Conference	2014
Peddycord-Liu, Z., Cody, C., Karamarkovich, S., Barnes, T., Lynch, C. & Rutherford, T.	Conference	2017
Martin, T., Aghababayan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., Philips, R. & Smith, C.	Conference	2013
Vahdat, M., Carvalho, M., Funk, M., Rauterberg, M., Hu, J. & Anguita, D.	Conference	2016
Akram, B., Min, W., Wiebe, E., Mott, B., Boyer, K. & Lester, J.	Conference	2018
Sanchez, E. & Mandran, N.	Conference	2017
Westera, W., Nadolski, R. & Hummel, H.	Conference	2014
Alonso-Fernandez, C., Martinez-Ortiz, I., Caballero, R., Freire, M. & Fernández-Manjón, B.	Journal	2019
Wen, C., Chang, C., Chang, M., Chiang, S., Liu, C., Hwang, F. & Tsai, C.	Journal	2018
Vahldick, A., Mendes, A. & Marcelino, M.	Journal	2017
Feng, X. & Yamada, M.	Conference	2019
Niemelä, M., Kärkkäinen, T., Äyrämö, S., Ronimus, M., Richardson, U. & Lyytinen, Heikki.	Journal	2020
Flores, R., Silverio, R., Feria, R. & Cariaga, A.	Book Chapter	2019
Priyaadharshini, M., Natha N., Dakshina, R., Sandhya, S. & Bettina, R.	Journal	2020

**Table 1. Works by author, selected for this systematic mapping.**

From these works, its objectives, theoretical foundation, methodologies and results were analyzed in order to obtain the answers to the research questions described in section 3.1.

#### 4.1. Q1: What educational phenomena does the research intend to analyze?

Considering the classified studies, the educational phenomena that the research sought to analyze were mapped and classified in 5 categories: Collaboration, Student Performance, Motivational Factors, Learning Processes and Problem Solving.

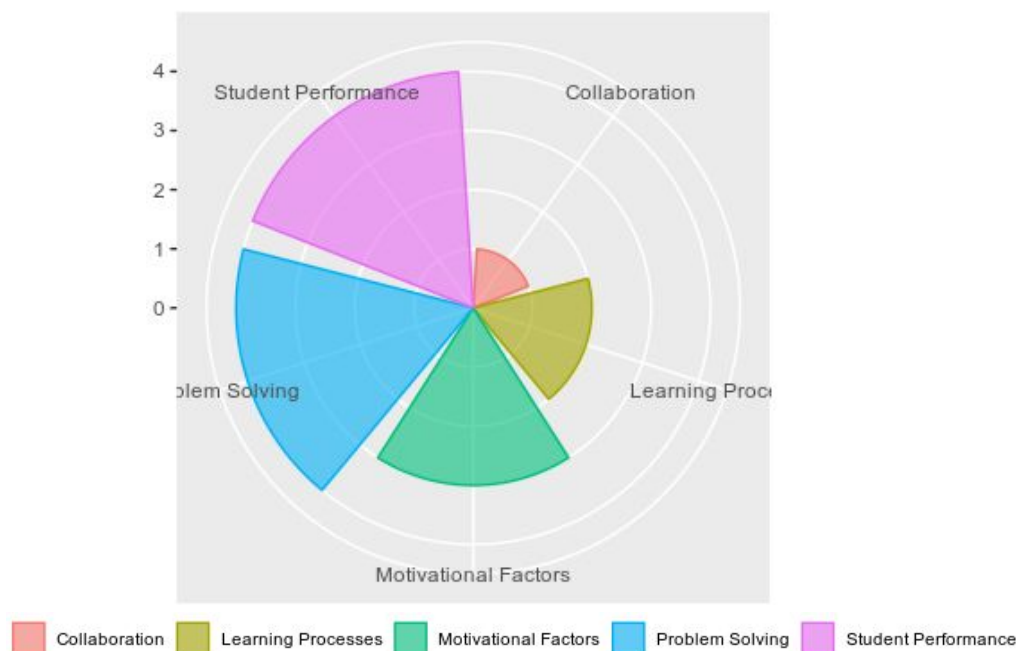


Figure 3. Educational phenomena of the selected studies.

According to graph 3, we can observe that the distribution of the quantitative of the educational phenomena found in the selected studies is quite diverse, with Collaboration being the least present phenomenon among these studies, being present in only one of them, in the research by Sanchez, E. & Mandran, N. (2015) who sought to understand how the interaction between competition and collaboration influences learning.

As for the other observed phenomena, the analysis and application of Learning Processes was present in the research by Martin, T. et al. (2013) and Peddycord-Liu, Z. et al. (2017), while the research by Callaghan, MJ. et al. (2014), Vahldick, A. et al. (2017) and Flores R.L. et al. (2019) sought to understand the Motivational Factors involved in the use of games for learning.

Among the most observed phenomena, we have the Student Performance analysis, which was analyzed in 4 of the selected surveys, those of Alonso-Fernández, C. et al. (2019), Feng, X. & Yamada, M. (2019), Niemelä, M. et al (2020) and Westera W. et al. (2014). Finally, we have the research by Vahdat, M. et al. (2016), Akram, B. et al. (2018), Priyaadharshini, M. et al. (2020) and Wen, C. et al. (2018), who analyzed how players use Problem Solving in the difficulties encountered during the games.

#### 4.2. Q2: What techniques / algorithms are used to analyze the data collected?

The algorithms and techniques used in the selected studies were grouped into 5 categories: Cluster Analysis, Classification, Text Mining, Regression Models and Data Visualization. Some studies apply their own techniques and algorithms, or do not present their defined analysis method, for these we use the term Undefined.



**Figure 4. Algorithms and techniques used in the selected studies.**

Based on the survey of the algorithms and techniques identified in the selected studies, a preference for the use of Cluster Analysis becomes evident, a technique that classifies participants into different groups, based on their behavior and preferences in the game.

The Cluster Analysis technique was present in 5 of the 14 studies, which is equivalent to about 36% of them, being used mainly to analyze the Learning Processes [Martin, T. et al., 2013; Peddycord-Liu, Z. et al., 2017] and Student Performance [Feng, X. & Yamada, M., 2019; Niemelä, M. et al., 2020], also being present in a study that sought to analyze the Problem Solving [Vahdat, M. et al., 2016].

The Classification, Regression Models and Data Visualization techniques were found, each, in 2 of the studies. Vahldick, A. et al. (2017) used the Rating to focus on the student's performance, seeking an adjustment of the game's difficulty based on the ratings generated by this technique, while Akram, B. et al. (2018) used it to assess the predictive power of this technique in students' post-test performance.

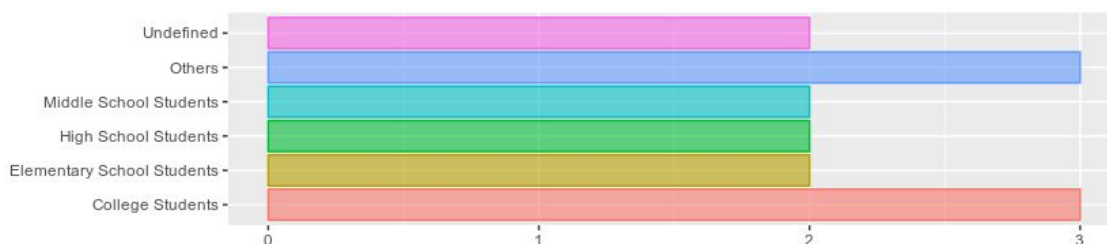
In research by Westera, W. et al. (2014) and Alonso-Fernández, C. et al. (2019), the Regression Models techniques were used to verify Student Performance, seeking to understand their behavior and predict the acquired learning and using Data Visualization techniques, Sanchez, E. & Mandran, N. (2015) were able to visualize the collaborative behavior of the participants, while in the research by Callaghan, MJ. et al. (2014), this same technique was used to measure student engagement.

Finally, the least used technique among the selected studies was Text Mining, being present only in the research by Wen, C. et al. (2018) who sought to understand student progress in problem solving.

#### **4.3. Q3: What is the educational level of the participants and the sample size?**

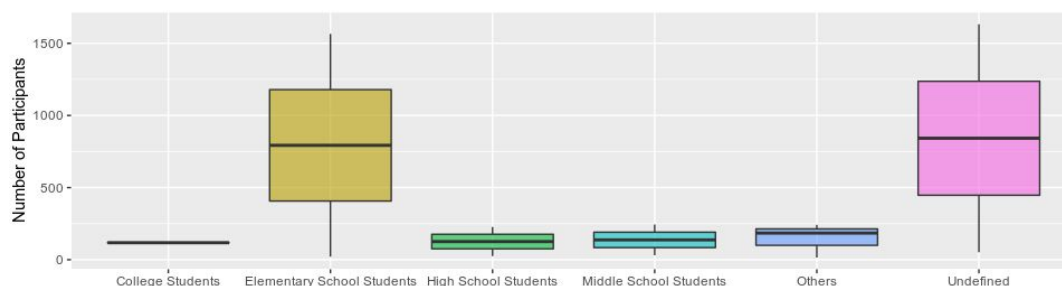
We considered four educational levels for this classification, as follows: Elementary School, Middle School, High School and College. Surveys in which participants are not

classified in one of these categories were classified as others, and surveys that do not make clear their educational level were classified as Undefined.



**Figure 5. Educational level of participants in selected studies**

The diversity of different educational levels among the participants becomes very evident with this graph, the research in the area of Game Learning Analytics analyzed aimed to not only analyze educational phenomena with students, but also with teachers [Sanchez, E. & Mandran, N., 2015], and adults [Vahdat, M. et al., 2016], while in the research by Feng, X. & Yamada, M. (2019) the performance of participants from different educational backgrounds was analyzed.



**Figure 6. Number of participants by educational level**

In addition to the diversity of education levels between surveys, it is also notable huge discrepancies in the number of participants who were subjected to these analyzes, the average being equal to 363 participants.

Research by Martin, T. et al. (2013) and Peddycord-Liu, Z. et al. (2017) aimed to observe the same educational phenomenon, use similar techniques, both games are in the area of mathematics and the participants also have the same education, but the number of students analyzed in the first is 21, while in the second survey it is 1565. On the other hand, two of the three surveys whose participants are from higher education "College Students" do not even make clear in their texts the amount of the sample analyzed [Callaghan, MJ. et al., 2014; Westera, W. et al., 2014].

## 5. Conclusion

This study presented a systematic mapping whose objective was to identify research gaps on Game Learning Analytics applications that sought to analyze an educational process.

A total of 14 studies were selected for the construction of this mapping, which were analyzed according to the research questions developed and we reached the following conclusions:



Regarding the first question, used to understand that educational phenomena are being analyzed in the research, we realized that Collaboration is the least researched phenomenon, being present in only one of the works. We also note the absence of learning theories that guide research, which can weaken the educational arguments of these studies.

As for the techniques used, the Cluster Analysis has a notable prominence, being present in 5 articles, while Text Mining was present in only 1. In addition, in 2 of the analyzed studies it does not use methods of the area in its researches, having to be classified as undefined.

The third question aimed to understand who the research participants were and in what quantity they presented themselves and we noticed that this varies greatly between studies, from students who are in the early years of teaching to master students, even teachers. We also noticed how the number of participants is very different between the surveys, ranging from 15 to 1632.

Finally, we understand that the Game Learning Analytics area is still relatively new, as it must have been perceived in Figure 1, the oldest studies presented as research return date from 2013, but this is an area that has been increasingly researched and debated, as it has enormous potential to assist in what we seek to value most, teaching.

## References

- Akram, B. et al. (2018). "Improving Stealth Assessment in Game-Based Learning with LSTM-Based Analytics". International Educational Data Mining Society, Paper presented at the International Conference on Educational Data Mining (EDM) (11th, Raleigh, NC, Jul 16-20, 2018)
- Alonso-Fernández, C. et al. (2019). "Learned applying learning analytics to assess serious games." *Computers in Human Behavior*. doi:10.1016/j.chb.2019.05.036.
- Alonso-Fernández, C. et al. (2019). "Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning*." doi:10.1111/jcal.12405
- Callaghan, MJ. et al. (2014). "Using game analytics to measure student engagement/retention for engineering education" 2014 11th International Conference on Remote Engineering and Virtual Instrumentation (REV), Porto, 2014, pp. 297-302, doi: 10.1109/REV.2014.6784174.
- Crookall, D. (2010). "Serious games, debriefing, and simulation/gaming as a discipline." *Simulation & Gaming*, 41 (6), 898–920. doi:10.1177/1046878110390784.
- Feng, X. & Yamada, M. (2019). "Effects of game-based learning on informal historical learning: A learning analytics approach". *Proceedings of International Conference on Computer in Education 2019*, pp505-514 At: Kenting, Taiwan, ROC
- Flores R.L. et al. (2019). "Motivational Factors Through Learning Analytics in Digital Game-Based Learning." In: Tlili A., Chang M. (eds) *Data Analytics Approaches in Educational Games and Gamification Systems*. Smart Computing and Intelligence. Springer, Singapore
- Foundation of American Scientists. (2006). "Summit on educational games: Harnessing the power of video games for learning." Washington, DC.
- Freire, M. et al. (2016). "Game Learning Analytics: Learning Analytics for Serious Games." *Learning, Design, and Technology*, 1–29. doi:10.1007/978-3-319-17727-4\_21-1

- Loh, C.S. et al. (2015) "Serious Games Analytics: Theoretical Framework." In: Loh C., Sheng Y., Ifenthaler D. (eds) Serious Games Analytics. Advances in Game-Based Learning. Springer, Cham
- Long, P., & Siemens, G. (2011). "Penetrating the fog: Analytics in learning and education." *Educause Review*, 31–40.
- Martin, T. et al. (2013). "Nanogenetic learning analytics: Illuminating student learning pathways in an online fraction game". Proceedings of the Third International Conference on Learning Analytics and Knowledge" - LAK '13. doi:10.1145/2460296.2460328
- Michael, D., & Chen, S. (2006). "Serious games: Games that educate, train, and inform." Boston, MA: Thomson.
- Niemelä, M. et al. (2020). "Game learning analytics for understanding reading skills in transparent writing system". *British Journal of Educational Technology*. 10.1111/bjet.12916.
- Peddycord-Liu, Z. et al. (2017). "Using serious game analytics to inform digital curricular sequencing: What math objective should students play next?". Proceedings of the Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '17. doi:10.1145/3116595.3116620.
- Petersen, Kai et al. (2008). "Systematic Mapping Studies in Software Engineering." In: EASE. p. 68-77.
- Priyaadharshini, M. et al. (2020). "Learning Analytics: Game-based Learning for Programming Course in Higher Education". *Procedia Computer Science*. 172. 468-472. 10.1016/j.procs.2020.05.143.
- 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>
- Sanchez, E. & Mandran, N. (2017). "Exploring Competition and Collaboration Behaviors in Game-Based Learning with Playing Analytics." In: Lavoué É., Drachsler H., Verbert K., Broisin J., Pérez-Sanagustín M. (eds) *Data Driven Approaches in Digital Education*. EC-TEL 2017. Lecture Notes in Computer Science, vol 10474. Springer, Cham.
- Seif El-Nasr, M. et al. (2013). "Game analytics – maximizing the value of player data." London, UK: Springer.
- Smith S.P. et al. (2015). "A Meta-Analysis of Data Collection in Serious Games Research." In: Loh C., Sheng Y., Ifenthaler D. (eds) Serious Games Analytics. Advances in Game-Based Learning. Springer, Cham.
- Vahdat, M. et al. (2016) "Learning Analytics for a Puzzle Game to Discover the Puzzle-Solving Tactics of Players." In: Verbert K., Sharples M., Klobučar T. (eds) *Adaptive and Adaptable Learning*. EC-TEL 2016. Lecture Notes in Computer Science, vol 9891. Springer, Cham.
- Vahldick, A. et al. (2017). "Dynamic Difficulty Adjustment through a Learning Analytics Model in a Casual Serious Game for Computer Programming Learning". *EAI Endorsed Transactions on Serious Games*.
- Wen, C. et al. (2018). "The learning analytics of model-based learning facilitated by a problem-solving simulation game." *Instr Sci* 46, 847–867. <https://doi.org/10.1007/s11251-018-9461-5>.
- Westera, W. et al. (2014). "Learning Analytics in Serious Gaming: Uncovering the Hidden Treasury of Game Log Files." In: De Gloria A. (eds) *Games and Learning Alliance*. GALA 2013. Lecture Notes in Computer Science, vol 8605. Springer, Cham