Challenges in educational game data modeling from the perspective of computing students: an empirical study

Fabrizio Honda[1,2], Fernanda Pires[1], Marcela Pessoa[1], Elaine H. T. Oliveira[2]

[1]Higher School of Technology – Amazonas State University (EST-UEA)
ThinkTEd Lab - Research, Development and Innovation in emerging technologies

[2]Postgraduate Program in Computer Science (PPGI)
Institute of Computing – Federal University of Amazonas (IComp-UFAM)

{fabrizio.honda,elaine}@icomp.ufam.edu.br,{fpires,msppessoa}@uea.edu.br

Abstract. *The use of Game Learning Analytics (GLA) is fundamental in assessing educational serious games, providing evidence of the player's learning path through collecting and analyzing interaction data records. The GLBoard model facilitates the implementation of these techniques through a generic and flexible data template for any educational game. However, learning designers consider a previous step, data modeling, complex. In this regard, this work presents an empirical study to investigate the emerging difficulties of this process. Results of the interaction of computer science students indicate that the main challenges are process abstraction, game appropriation, time to carry out the process, and the need for prior knowledge of computer science disciplines - programming, database, and systems modeling.*

1. Introduction

Approaches that use data records are fundamental in educational games, providing evidence to analyze player behavior and their learning path, such as Game Learning Analytics (GLA) [Freire et al. 2016]. The objective of GLA or Learning Analytics (LA) applied to Serious Games is the collection, analysis and visualization of player interaction data, an area that has been gaining more and more recognition and bringing contributions to several stakeholders: (i) student – to monitor their progress; (ii) developer – helping to identify game design inconsistencies to improve the game; and (iii) teacher – to identify learning gaps in a class, monitor student progress, etc. [Melo et al. 2020, Banihashem et al. 2024].

However, GLA presents the following challenges: (i) the complexity of implementation by developers and (ii) the lack of standardization, resulting in specific and non-replicable capture strategies [Saveski et al. 2016, Alonso-Fernandez et al. 2017]. Furthermore, in Learning Analytics, data capture is typically applied to virtual learning environments [Almeida and Brennand 2016], MOOCS [İnan and Ebner 2020] or online judges [Oliveira et al. 2022] and to exercises with exact answers, in the form of objective assessments, but in GLA the focus is on games, which are complex learning environments [Plass et al. 2015]. Therefore, these structures can have different shapes and styles, the learning elements can be classified in various ways, and the player's path can influence their results. To minimize these obstacles, models such as xAPI-SG and GLBoard were designed to implement GLA techniques for collecting and analyzing data from educational games.

Despite the contributions of GLA techniques, using xAPI-SG, GLBoard, or another model, a process before implementing these techniques is considered highly complex by learning designers and programmers: data modeling – or data selection [Hauge et al. 2014]. This process refers to the definition of the data that will be collected from players and the justification for choosing this data, whose objective is to assist in identifying evidence of learning, whether or not related to other variables. The collection of time records can, for example, indicate how much time players require to make a decision and whether or not this time is related to the best choice in the game scenario, whether it is selection - related to the embedded curriculum material - or strategy - about game design. Furthermore, this definition must occur at the beginning of the development of the artifact project [Hauge et al. 2014, Alonso-Fernández et al. 2021, Kitto et al. 2020], as they support the implementation of data collection for a specific template, such as GLBoard. This directly impacts the game's development because, if it is already partially implemented or finalized, incorporating the capture structure becomes challenging due to factors such as game abstraction, complex mechanics, adaptation of the structure, and code navigation, even for developers with expertise in educational games and GLA [Macena et al. 2024].

To make it possible to collect evidence of learning in serious educational games, data modeling plays a fundamental role in the implementation of GLA techniques, yet in the educational game design or learning design phase. However, despite being mentioned in GLA works – which reinforce that it is a necessary step and exemplify the definition of some data [Perez-Colado et al. 2018, Alonso-Fernández et al. 2019a], it is not described in detail, that is, it does not contain the step-by-step process for modeling data. Furthermore, the stakeholders who implement these techniques are educators and/or professionals (game designers, developers, etc.). However, most educational games are developed in academic contexts, such as in Course Conclusion Papers, Master's Dissertations and Doctoral Theses [Cordeiro and Duarte 2020, Nascimento and Leite Bruno 2024, Tarouco et al. 2005, Portella et al. 2017]. The creators of these objects are usually learning/instructional designers who act as developers and students. Thus, no support or materials in the literature include these profiles in implementing GLA. Therefore, this work seeks to conduct an empirical study to answer the following research question: "What are the difficulties of computer science students in data modeling for educational games?" The work is organized as follows: Section 2 presents the theoretical foundations and related works; Section 3 includes the methodology adopted in the study; Section 4 presents the results and discussions, and Section 5 highlights the final considerations.

2. Foundations and related work

GLA combines techniques from Game Analytics (GA) – which aims to improve and adjust game design to increase post-sale revenues – and Learning Analytics (LA) – which focuses on measuring, collecting, analyzing, and reporting data from learning environments, aiming to extract information from students and analyze learning progress [Loh et al. 2015, Alonso-Fernández et al. 2019b, Siemens and Long 2011, Sclater 2017]. This strategy involves (i) storing player interactions (logs) in a non-intrusive way so as not to interrupt the gameplay rhythm, (ii) visualizing the collected data through dashboards, and (iii) analyzing interactions, which can provide valuable insights to stakeholders. Furthermore, the intersection of these interactions with heuristic evaluations and data science/machine learning techniques can provide more robust inputs on the players' learning process [Alonso-Fernández et al. 2021, Silva et al. 2021, Alonso-Fernández et al. 2022].

One of the existing models for implementing GLA techniques is the xAPI-SG Profile, which is based on xAPI – a specification that tracks data (statements) in JSON format, representing learning activities through three main fields: actor (who acted), verb (the action), and object (target of the action). On the other hand, xAPI-SG is a specific profile for serious games to collect data in a standardized way, commonly including timestamps and may include extensions, such as ad-hoc fields that reflect specific player interactions. Therefore, this model defines verbs and activities involving common interactions in serious games, such as completables, alternatives, and general variables. An example of this template is: a player (actor) progressed (verb) in question 1 (object), with a performance of 0.35 (outcome, extended field) in a specific time [Alonso-Fernandez et al. 2017, Alonso-Fernández et al. 2019b, Alonso-Fernández et al. 2021].

However, xAPI-SG have some limitation which include: (i) a high learning curve, requiring a thorough study of the xAPI structure and vocabulary for data capture; (ii) flexibility: despite the Serious Games profile focusing on interactions for serious games, xAPI was initially designed for generic educational environments; (iii) the need to configure an LRS (Learning Record Store) for data storage, which requires technical knowledge of back-end and integration with game engines; and (iv) the absence of a built-in tool for analysis and visualization of the collected data (such as a Dashboard, for example) [Serrano-Laguna et al. 2017, Alonso-Fernández et al. 2022].

The obstacles of GLA and the limitations of xAPI-SG – until then, the most common model for implementing these techniques – motivated the creation of GLBoard: a model for collecting and analyzing data from educational games, which provides a generic standard structure (template) in JSON format for collecting data [Silva et al. 2022]. The model is composed of four main modules: (i) API – which manages requests and communicates with other modules; (ii) Database – stores game and player interaction data; (iii) Dashboard – responsible for displaying graphs from the collected data; and (iv) Unity Package – a package in the Unity game engine where it is possible to access GLBoard. In this way, the developer can incorporate GLBoard into their game and fill in the data template in JSON format, which is collected asynchronously. This template contains two types of variables: *(i)* generic – common to any game (player data, number of levels, performance, login and logout times, etc.), where the developer only needs to assign values; and *(ii)* flexible – such as $path\_player$, which allows storing sets of data, giving the developer freedom to define and assign values according to the data they want to capture.

Therefore, GLBoard becomes replicable for any game and can store the players' gameplay trajectory, the analysis of which can provide a detailed assessment of how learning is occurring [Silva et al. 2021, Wallner and Kriglstein 2013]. Among the games that use GLBoard to collect data, the following stand out: "MyName" [Nascimento et al. 2023], "Cadê minha Pizza?" [Honda et al. 2023] and "Hello Food" [Macena et al. 2024] . However, whether to assign values in the GLBoard data template, xAPI-SG, or another GLA tool, it is essential to perform data modeling (or selection), which consists of defining which data will be captured and the justification of why they are relevant to help identify evidence of learning – according to the educational objectives of the game. However, this complex process requires a thorough analysis of the educational game and is not covered in detail in GLA studies in the literature. The works described below mention this process, listing collection variables for some educational games.

In Perez-Colado et al. [2018], an extension of a GLA system is described, whose objective is to manage multilevel analyses through analysis models and technical infrastructure. LAMs (Learning Analytics Models) are proposed to help structure data collection, analysis, and visualization based on well-defined learning objectives. The authors also describe meta-LAMs, which focus on integrating and analyzing data from multiple games or mini-games, providing an overview of progression and learning. The First Aid Game is used as a case study, in which the authors reinforce that the data (traces) must be selected sufficiently to reflect the student's learning progress – alluding to the data modeling (or selection) process. In the game, the selected data were level progress (0 to 1), answers to questions, points, and events ("initialized", "progressed", "completed" and "selected"). Furthermore, the system includes data analysis and visualization for teachers and developers, allowing them to track the students' performance. It also consists of an API that allows the games to access the analysis in real-time.

The work of Alonso-Fernández et al. [2019] dealt with the application of GLA to three serious games, aiming to (i) validate/implement a game about cyberbullying awareness; (ii) validate the design of a game focused on improving the independent life of users with intellectual disabilities and; (iii) improve the evaluation of a game about first aid techniques. The Learning Analytics Model (LAM) was used to connect the games' objectives according to the data to be analyzed, whose model helps to identify which data should be collected (referring to the data modeling/selection processes). The xAPI-SG data template was used, whose data selected for collection were: (i) Connected – decision-making in dialogues, changes in friendship patterns, interactions with objects, etc.; (ii) DownTown – attempts per mini-game, clicks on the interface, timestamps, etc.; and (iii) First Aid Game – completion status, total and level score, correct/incorrect answers, etc. The results were positive, allowing the achievement of the defined objectives and bringing lessons learned, such as standardization of data collected according to xAPI-SG, the effectiveness of GLA to achieve different goals, and the benefit of stakeholders with data from GLA.

The focus of the work by Emerson et al. [2020] was to investigate how the use of multimodal data (data records, facial expressions, and eye tracking) can be used to predict student performance and interest in the game Crystal Island. An experiment was conducted with 61 undergraduate students, whose sessions had up to 3 hours of gameplay, involving pre- and post-tests, questionnaires and motivation, video sensors, and gaze tracking. Data modeling is not comprehensively covered, but the authors indicate that the data selection included (i) gameplay traces (movement, responses, time, etc.); (ii) gaze (eye fixation on eight categories of in-game objects, such as NPCs); and (iii) facial expressions, encoded in 20 action units (AUs). The results showed higher accuracy than unimodal models, helping to predict student performance and interest in the game, whose authors point out a potential to guide the development of real-time adaptive strategies.

The works cited above contribute to the GLA area and mention the data modeling process, highlighting it as necessary for data collection. However, despite indicating the data selected for collection and informing that the decisions of this process are based on the learning objectives, data modeling is not described in detail. In addition, the stakeholders responsible for this process are game designers (including developers) and educators. Thus, the contributions of this work and differences with other related works are: (i) an analysis focused on data selection/modeling, not currently present in the literature; (ii) a survey of difficulties of this process and; (iii) the perspective of students (specifically of computing) as stakeholders of the data modeling process.

3. Methods

This work aims to identify the difficulties and understand the perception of computer science students (self-assessment) regarding the data modeling process of educational games. Given this, an empirical study was conducted, organized in stages that detail the experimental design, detailed and illustrated in Figure 1.

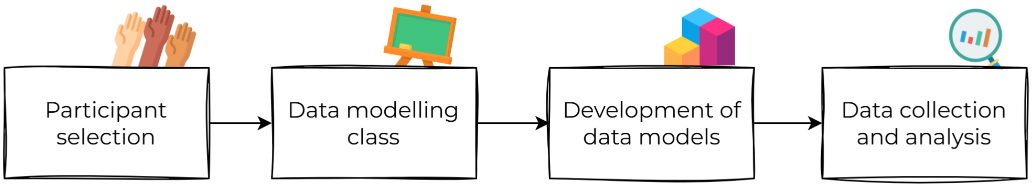

Figure 1. Steps of the study carried out on data modeling.

Participant selection: The participants of this study were selected based on two criteria: (i) undergraduate students of the Computing Education degree course at the Amazonas State University (UEA) who are taking the "Fundamentals of Educational Software" (FSE) course, carrying out an activity of building an educational game and implementing GLA techniques – and must then perform data modeling; and (ii) participants of a research and development laboratory of educational technologies, who are refactoring games that have already been implemented, currently in the data modeling stage. Therefore, 16 students participated in the study: 11 (69%) males and 5 (31%) females, aged 19 to 26 years. All participants were enrolled in computing courses: 12 (75%) in the computing education degree, 3 (19%) in Information Systems, and 1 (6%) in Engineering. 75% were in the fourth or sixth semester of the course, while the rest were in the eighth or higher. 100% had taken the three introductory programming courses, while only a few had taken Database I (62.5%), Data Structures II (56.25%), and Systems Modeling and Design (50%). Nine participants (56%) reported not having failed any of these courses, while 7 (44%) had failed at least once in one of them.

Data modeling class: To effectively start the study, a class was held within the scope of the discipline to which all participants were invited. The class took place in person at the university, lasting one hour, and its objective was to introduce the content to the participants. Thus, the following topics were addressed: (i) educational games; (ii) challenges and obstacles in the assessment of learning in games; (iii) Learning Analytics, Game Analytics, and Game Learning Analytics; (iv) GLBoard: model definition and architecture, data template, $path\_player$ variable, and capture examples; and (v) data modeling. A multimedia presentation was used with information and images about the previous topics, including dynamic examples where students could interact and practice the content.

Development of data models: Participants were assigned individually modeling data from an educational game after the presented content. The data modeling for the students in the discipline should be based on the educational games they were developing. In contrast, for the laboratory participants, the choice of educational game was free – as long as there were no duplicates. To this end, a collaborative online spreadsheet was shared with the participants, containing a table with the following columns: number, name of the variable (data), data type, example, and justification for collection. Then, the participants were asked to make a copy of the spreadsheet and begin the activity. In addition, the students were instructed to perform the modeling manually and not to use Generative Artificial Intelligence tools such as ChatGPT – as this could compromise the viability of the study. The participants then began the data modeling process, which consisted of the analysis/use of the educational games and the proposal and insertion of the capture variables in the spreadsheet, lasting approximately one hour and thirty minutes. Figure 2 illustrates some moments of the activity.

(a) Record 1. (b) Record 2. (c) Record 3.

Figure 2. Records of moments of data modeling by students.

Data collection and analysis: For data collection, participants were asked to fill out a Google Forms form after completing the activity, which contained: (i) a space to send the spreadsheets with the models; (ii) quantitative questions on a Likert-5 scale; and (iii) qualitative questions, both addressing the students' perspective on the data modeling process. The questions on the form can be found at the link[1], whose participants agreed to fill it out, with the consent that the data will be disclosed anonymously and exclusively for research purposes. Regarding data analysis, we used (i) visual analysis for the quantitative questions using boxplot and (ii) content analysis [Bardin 2015] of the qualitative data completed by the participants.

4. Results and discussions

The study results include the analysis of quantitative and qualitative questions assigned by computer science students regarding the data modeling process in educational games. Figure 3 illustrates a boxplot on the quantitative questions (on a Likert-5 scale).

Regarding **understanding the objective**, there was agreement among the participants: all gave a score of 3 or higher, with an average of 3.9 and a median of 4.

[1]https://drive.google.com/file/d/1ziEwdFyk57cuRdwuP52q8Z-80KE3Ehez/view?usp=sharing

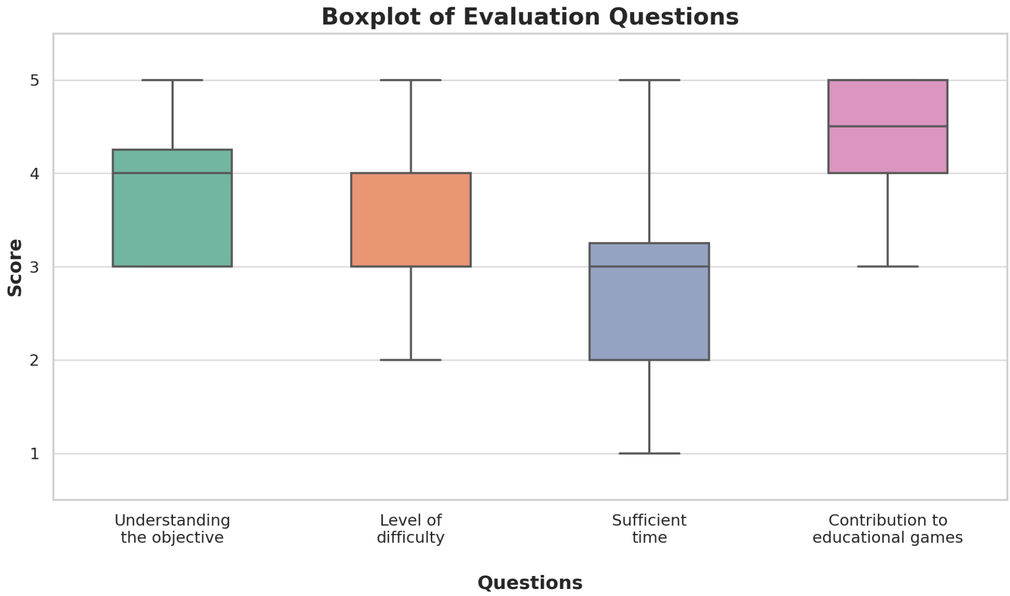

Figure 3. Boxplot with the evaluation of quantitative questions.

This demonstrates unanimity regarding the need to model data before implementing capture structures – whether in the GLBoard template or another tool. Regarding the question "What do you consider to have done well in data modeling?", the participants provided different answers: (i) without coherence with the question (for example: "no"); (ii) generic, with few details about the process (i.e.: "made very objective choices"); (iii) no understanding or with difficulty (i.e.: "could improve more"); and (iv) detailed, regarding the game information (i.e.: "identify how the player lost or won the level"). When relating these responses to the previous quantitative question, it is noted that the students understand the objective of modeling data. However, they had difficulty explaining what they did well in the process.

Regarding the **level of difficulty** of data modeling, only one student (6%) found the process "easy". The others indicated that data modeling is "medium" (50%) and "hard" (44%), corroborating that it is a complex process. This fact is further intensified when analyzing that: (i) a little more than half of the participants had already developed educational games (56.25%); (ii) the 2 participants who had experience with GLA implementation found data modeling to be a process of "medium" and "hard" difficulty; (iii) 10 participants (62.5%) reported having a good level of knowledge of programming logic; (iv) 10 were in the 6th period or higher of the course; and (v) 56.25% indicated having average and good knowledge (18.75%) of databases. When asked what they found most challenging in data modeling, most participants pointed out defining the data/variables for collection and abstraction from this process. Two responses stand out in this regard: "Thinking about all the conditions the player faces" and "without being very sure of what you are looking for, it is a bit difficult to point out variables.". These comments reflect a characteristic challenge of game learning assessment by learning designers: taking ownership of an educational game and, based on its learning objective, identifying the key points of the game that indicate whether it was or is being achieved.

Regarding the question of **sufficient time** to carry out the data modeling process, which lasted around an hour and a half, 7 (44%) thought that the time was insufficient, and 5 (31.3%) indicated that it was average. This point also corroborates the complexity of data modeling, which requires not only knowledge but also time, which varies according to the game – robust mechanics, number of elements in the level design, and distinct phases, for example, significantly impact the time/complexity to model the data and, consequently, the implementation of GLA techniques [Macena et al. 2024]. In addition, participants were asked if "there was anything you learned during the modeling process that you did not know before." The majority (56.25%) reported that they did not learn anything new, 2 (12.5%) reported generic information, and 5 (31.25%) indicated that they did, with emphasis on (i) "the use of lists to capture some data in more detail"; (ii) "analyzing what the variables can say about the player" and; (iii) "as the process went on it became easier to identify which data I should store". Item (i) reports an essential strategy for data modeling since data is grouped into lists, which makes it easier to organize and manage data. Item (ii) is the precise purpose of data modeling: to collect evidence that can provide insights into the player's behavior and path. In turn, (iii) is directly related to the time for modeling: the more time dedicated to the activity, the more it is possible to analyze the game in detail and locate important variables for collection.

Although the difficulty level was high and there was insufficient time to model data, 14 students (94%) believe that this process contributes to developing more effective educational games. Among the reasons, participants pointed out aspects such as reduction of implementation time, definition of variables to identify signs of learning, support for data capture, more profound perception of general and specific aspects of the game, understanding of player behavior, identifying game design problems, etc. In addition, six students (38%) indicated that modeling data was similar to previous activities such as designing class diagrams, gathering requirements, database modeling, etc. At this point, 10 participants (62.5%) reported that prior knowledge of previous disciplines contributes to data modeling, such as: programming (introduction, data structures, and object orientation), systems modeling, database, and theoretical and pedagogical disciplines. The participants highlighted the following contributions: the creation of diagrams, modeling of systems and databases, analysis of entities, survey of variables, and learning theories. These points also corroborate the complexity of data modeling: knowledge of previous disciplines is fundamental for this process. However, only 58.8% of the participants had taken Database (DB) I, 47.1% Modeling and Systems Design (MSD) and 52.9% Algorithms and Data Structures II. Furthermore, when asked about the level of knowledge they believe they have in these disciplines, the majority indicated that it is medium/low, with averages of 2.7 in DB and 2.1 in MSD.

In general, the study indicated that data modeling is essential for constructing educational games in which students demonstrate an understanding of their objectives. However, the level of complexity was considered high, and the execution time of the study was insufficient. Thus, in answering the research question "What are the difficulties of computer science students in the process of data modeling for educational games?", the main difficulties are: (i) in the abstraction of the process; (ii) in appropriating the educational game to identify the capture variables; (iii) in the time dedicated to this process; and (iv) in understanding content from previous disciplines, supporting the fact that the data modeling process of educational games is complex.

Among the limitations of this study, the following stand out: (i) time, since it was applied only in one class of a discipline, which may have compromised the students' experience and the quality of the modeling performed; and (ii) perspective, considering only the students' point of view regarding the process. Item (i), despite being a limitation, was also an important finding that corroborates the complexity of data modeling, while (ii) was precisely the study's objective. Since this is an introductory study and, to date, the only one in the literature that covers the data modeling process in detail, these limitations are minor and will be solved in future research.

5. Conclusions

The GLA field significantly contributes to educational games, providing data-based evidence to analyze the player's journey and learning progression. GLBoard is a tool that enables the implementation of these techniques, providing a JSON template for data collection with generic and flexible variables, making it replicable to any educational game. However, the process before this implementation, called data modeling, is complex. In addition, GLA studies generally do not include the student as a learning designer, even though most educational games are developed in an academic environment.

In this sense, this work sought to conduct an empirical study with the following research question: "What are the difficulties of computer science students in data modeling for educational games?". To this end, a series of steps were conducted: (i) selection of participants – including computer science students and participants from a research and development laboratory for educational technologies; (ii) data modeling class – to present the content to the students; (iii) elaboration of data models – where the students modeled the data corresponding to the educational games they were building or that they selected; and (iv) data collection and analysis, using Google Forms to collect quantitative and qualitative questions about the students' perception of data modeling. This data generated a boxplot graph to help understand the results and content analysis was used to answer the qualitative questions.

As a result, the students indicated that data modeling contributes to creating educational games and reported understanding the objective. However, the process was considered complex, and the time to complete the study was assessed as insufficient - lasting approximately one hour and thirty minutes. The main difficulties reported were related to the abstraction of the process, the appropriation of the game to define the capture variables, the time required for this analysis and execution of the activity, and the understanding of content from previous courses, which assist in this process. From the perspective of computer science students, the study provided a more detailed understanding of the data modeling process, which is only mentioned in GLA papers and generally does not consider the student as a stakeholder. With this, it is possible to investigate and make viable proposals to minimize these difficulties and assist students in this fundamental process for educational games.

Future work includes holding new classes with more participants, adding more dynamic elements, providing more time for the execution of the activity – possibly dividing it into more than one application moment, inserting a brief explanation about Database, Data Structures, and Systems Modeling content, adding more data modeling examples, etc.

6. Acknowledgment

In this study, Generative AI (GenAI) was used through Chat-GPT 4o from OpenAI to generate the codes for graphs in Overleaf, aiming to help minimize time and effort in constructing these representations.

This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Finance Code 001 and Conselho Nacional de Desenvolvimento Cientifico e Tecnologico - Brasil - CNPq (Process 303443/2023-5). This work was partially funded by the Amazonas State Research Support Foundation – FAPEAM – through the PDPG/CAPES/FAPEAM project.

References

Almeida, L. H. G. d. and Brennand, E. G. d. G. (2016). Learning analytics em ambiente virtual de aprendizagem moodle: um estudo de caso em componentes curriculares para cursos semipresenciais. *Gestão & Aprendizagem*, pages 76–93.

Alonso-Fernandez, C., Calvo, A., Freire, M., Martinez-Ortiz, I., and Fernandez-Manjon, B. (2017). Systematizing game learning analytics for serious games. In *2017 IEEE global engineering education conference (EDUCON)*, pages 1111–1118. IEEE.

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., and Fernández-Manjón, B. (2019a). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141:103612.

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., and Fernández-Manjón, B. (2022). Game learning analytics:: Blending visual and data mining techniques to improve serious games and to better understand player learning. *Journal of Learning Analytics*, 9(3):32–49.

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., and Manjón, B. F. (2021). Data science meets standardized game learning analytics. In *2021 IEEE Global Engineering Education Conference (EDUCON)*, pages 1546–1552. IEEE.

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., and Fernández-Manjón, B. (2019b). Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior*, 99:301–309.

Banihashem, S. K., Dehghanzadeh, H., Clark, D., Noroozi, O., and Biemans, H. J. (2024). Learning analytics for online game-based learning: A systematic literature review. *Behaviour & Information Technology*, 43(12):2689–2716.

Bardin, L. (2015). Análise de conteúdo (la reto & a. pinheiro, tradução)(6ª edição). *Lisboa, Portugal: Edições*, 70.

Cordeiro, E. A. and Duarte, E. M. (2020). Jogos educacionais digitais: estado da arte em trabalhos de conclusão de curso. *Revista Sítio Novo*, 4(1):125–133.

Emerson, A., Cloude, E. B., Azevedo, R., and Lester, J. (2020). Multimodal learning analytics for game-based learning. *British journal of educational technology*, 51(5):1505–1526.

Freire, M., Serrano-Laguna, Á., Manero, B., Martínez-Ortiz, I., Moreno-Ger, P., and Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. In *Learning, design, and technology*, pages 1–29. Springer Nature Switzerland AG.

Hauge, J. B., Berta, R., Fiucci, G., Manjón, B. F., Padrón-Nápoles, C., Westra, W., and Nadolski, R. (2014). Implications of learning analytics for serious game design. In *2014 IEEE 14th international conference on advanced learning technologies*, pages 230–232. IEEE.

Honda, F., Macena, J., Duarte, J. C., Pires, F., Pessoa, M., and Oliveira, E. H. (2023). Um estudo de caso para a implementação de game learning analytics (gla) no desenvolvimento de jogos educacionais. In *Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (WAPLA)*, pages 138–146. SBC.

İnan, E. and Ebner, M. (2020). Learning analytics and moocs. In *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences: 7th International Conference, LCT 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*, pages 241–254. Springer.

Kitto, K., Whitmer, J., Silvers, A., and Webb, M. (2020). Creating data for learning analytics ecosystems.

Loh, C. S., Sheng, Y., and Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*, pages 3–29.

Macena, J., Honda, F., Melo, D., Pires, F., Oliveira, E. H. T., Fernandes, D., and Pessoa, M. (2024). Desafios na implementação de técnicas de gla em um jogo educacional de algoritmos: um estudo de caso. In *Anais Estendidos do XXIII Simpósio Brasileiro de Jogos e Entretenimento Digital*. SBC.

Melo, D., Melo, R., Bernardo, J. R. S., Pessoa, M., Rodríguez, L. C., and Pires, F. (2020). Uma estratégia de game learning analytics para avaliar level design em um jogo educacional. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 622–631. SBC.

Nascimento, A. M. d. S. and Leite Bruno, S. (2024). Jogos digitais educacionais: o que as dissertações e teses dizem? *International Journal Education and Teaching (PDVL)*, 7(2):62–87.

Nascimento, L. T., Honda, F., Melo, D., Pessoa, M., Oliveira, E. H., Fernandes, D., and Pires, F. G. (2023). My name: desenvolvimento de um conjunto de mecânicas para abordar o problema da mochila em um jogo educacional. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 888–899. SBC.

Oliveira, E. H., de Oliveira, D. B., Carvalho, L. S., and Pereira, F. D. (2022). Learning analytics no ensino de introdução à programação de computadores. In *Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil (WAPLA)*, pages 50–59. SBC.

Perez-Colado, I., Alonso-Fernandez, C., Freire, M., Martinez-Ortiz, I., and Fernandez-Manjon, B. (2018). Game learning analytics is not informagic! In *2018 IEEE Global Engineering Education Conference (EDUCON)*, pages 1729–1737. IEEE.

Plass, J. L., Homer, B. D., and Kinzer, C. K. (2015). Foundations of game-based learning. *Educational psychologist*, 50(4):258–283.

Portella, F. F., Tubelo, R. A., Zanatta, E. J., and Pinto, M. E. B. (2017). Experiência da una-sus/ufcspa no desenvolvimento de jogos educacionais. *Experiências exitosas da Rede UNA-SUS*, page 196.

Saveski, G. L., Westera, W., Yuan, L., Hollins, P., Manjón, B. F., Ger, P. M., and Stefanov, K. (2016). What serious game studios want from ict research: identifying developers' needs. In *Games and Learning Alliance: 4th International Conference, GALA 2015, Rome, Italy, December 9-11, 2015, Revised Selected Papers 4*, pages 32–41. Springer.

Sclater, N. (2017). *Learning analytics explained*. Routledge.

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

Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5):30.

Silva, D., Melo, R., Pires, F., and Pessoa, M. (2021). Avaliacão de objetos digitais de aprendizagem: como os licenciados em computação analisam jogos educacionais? *Revista Novas Tecnologias na Educação*, 19(2):111–121.

Silva, D., Pires, F., Melo, R., and Pessoa, M. (2022). Glboard: um sistema para auxiliar na captura e análise de dados em jogos educacionais. In *Anais Estendidos do XXI Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 959–968. SBC.

Tarouco, L. M. R., Konrath, M. L. P., and da Silva Grando, A. R. (2005). O aluno como co-construtor e desenvolvedor de jogos educacionais. *Revista Novas Tecnologias na Educação*, 3(2).

Wallner, G. and Kriglstein, S. (2013). Visualization-based analysis of gameplay data–a review of literature. *Entertainment Computing*, 4(3):143–155.