Can data reveal learning in computing? Insights from an empirical study in Game Learning Analytics

Waldecir Martins[1], Fabrizio Honda[1,2], David Fernandes[2], Fernanda Pires[1], Marcela Pessoa[1]

[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)

{wdsm.lic22,fpires,msppessoa}@uea.edu.br,

{fabrizio.honda,david}@icomp.ufam.edu.br

Abstract. *Computer Science Education has taken on a fundamental role in the Brazilian educational context, especially after its mandatory inclusion in Basic Education. However, its concepts are complex, even for students in higher education. In this scenario, educational games emerge as a promising strategy to mitigate this complexity, especially when combined with Game Learning Analytics (GLA), which enables the analysis of the player's journey to identify evidence of learning. This work consists of an empirical study in GLA, aimed at analyzing interaction data (logs) that we collected through GLBoard from students using the educational game "OptimaCorporation". The study involved eight Computer Science students, and the results indicate that interaction data can reveal signs of learning, but they require complementary tools.*

1. Introduction

In recent years, the Brazilian government has taken initiatives to make the inclusion of computing in Basic Education mandatory, such as Law No. 14,533/2023 [Brasil 2023], highlighting the importance of developing this curricular component for students' education. As a result, computing education has gained ground in both Elementary and higher education curricula, as well as in strategies to support the understanding of its concepts [Vasconcelos and Neto 2020]. However, the field is not trivial, and even higher education students face several obstacles in understanding computing content [Lima and Menezes 2024, Ramos 2009].

Among the possible causes that make learning computer science challenging, we can mention: (i) difficulty in understanding logic; (ii) abstraction of concepts; (iii) cognitive overload, considering formal notations and new content; (iv) difficulty in associating concepts with practical applications, among others [Castro and Cruz 2022, Santos et al. 2024, Silva et al. 2025]. These obstacles contribute to a lack of motivation in learning and increase dropout rates, especially in introductory programming courses [Silva et al. 2024]. Therefore, it is necessary to use tools that reduce the perceived difficulty and boost learning, such as educational games [Oliveira and Boff 2024].

Educational games can help with understanding and mitigating the difficulties associated with computer science content [Egenfeldt-Nielsen 2006, Videnovik et al. 2023, Macena et al. 2022]. These learning objects provide interactive environments that promote student motivation and engagement, while also reducing the perception of complexity by transforming computational problems into playful challenges [Plass et al. 2015, Martins et al. 2025]. Moreover, when aligned with pedagogical theories, they facilitate cognitive development by placing the student at the center of the learning process [Pires et al. 2021].

Despite the potential of educational games, the process of assessing learning is challenging: (i) there are few established evaluation models; (ii) most assessments use questionnaires that only consider the player's perception; and (iii) games are generally not used as assessment tools, disregarding relevant elements in gameplay that may express the player's progress [Silva et al. 2021, Melo et al. 2020]. Game Learning Analytics (GLA) can be an alternative: collecting, analyzing, and interpreting data from serious games [Freire et al. 2016].

In this context, this study poses the following research question (RQ): "What insights can we extract from data collected from a game about optimization problems, and how can they reveal evidence of learning in computing?". To answer the RQ, we conducted an empirical study. The main contributions of this study include: (i) the use of a generic data-capture model (GLBoard); (ii) the analysis of a topic that remains underexplored in educational games in the literature (optimization); and (iii) the use of emerging technologies (Large Language Models – LLMs) for data modeling, analysis, and interpretation of GLA data.

2. Foundations and related work

GLBoard is a generic model for capturing and analyzing data from educational games, composed of four main modules: (i) database; (ii) Unity package; (iii) API; and (iv) Dashboard. For each player, the model provides a data template in JSON format, structured into four classes: (i) $PlayerData$ – name, ID, and basic demographic information of the player; (ii) $GameData$ – gameplay data, consisting of one or more phases; (iii) $Phase$ – game levels, including names, status, and a set of sessions; (iv) $Session$ – attempts within the phases, including start/end date, completion status, performance, and the player's path ($path\_player$). The developer can freely configure the $path\_player$, adding the GLA variables corresponding to their educational game.

After implementing the data template, whether from GLBoard or another GLA tool, the tracker collects data and makes it generally available through dashboards. By analyzing these logs (records of player interaction), after the logging, processing, and transformation of events, it is possible to obtain indicators that reveal behavioral patterns and may express learning [Alonso-Fernández et al. 2019]. The following works are related to this context, involving the analysis of logs from educational games.

The work by Alonso-Fernández et al. [2019] presented the application of GLA in three serious games. The authors used a Learning Analytics Model (LAM), which connects the objectives of the games to the data to be analyzed. The research aimed to: (i) validate a game focused on raising awareness about cyberbullying; (ii) validate the design of a game aimed at improving the autonomy of users with intellectual disabilities; and (iii) enhance the assessment of a game about first aid techniques. The GLA model used was xAPI-SG [Serrano-Laguna et al. 2017], with collected data including: (i) interactions with objects; (ii) attempts per mini-game; (iii) total score, among others. The results were satisfactory, achieving the defined objectives and providing lessons learned about the contributions of GLA.

The work by Alencar et al. [2020] demonstrated a data analysis applied to the game "Tricô Numérico", aiming to evaluate the level design and player behavior. To achieve this, they implemented a capture algorithm capable of recording: (i) collisions with enemies; (ii) correct answers; and (iii) errors in arithmetic operations, sending this data to Google Sheets via HTTP requests. Based on these records, the study identified dropout patterns associated with high collision rates, indicating possible design issues. The results showed that more engaged players generated more gameplay logs, highlighting the importance of capturing multiple variables to understand the learning process and assess the effectiveness of using logs.

In the work by Gil et al. [2024], GLA techniques were applied using sessions from the game "Lightbot", to understand the development of computational thinking skills. The authors proposed a process for capturing and processing data, in which the system recorded player interactions, structured them into logs, and subsequently converted them into performance indicators, such as resolution time and the number of attempts. The results provided a detailed view of the problem-solving process, contributing to the assessment of the game's effectiveness as a support resource for learning computing.

Table 1 presents a comparison between the studies found in the literature and this research. While Alonso-Fernández et al. [2019] focused on the application of GLA in serious games from various domains, and the work by Alencar et al. [2020] concentrated on log analysis to improve level design in a game, Gil et al. [2024] applied GLA to analyze the development of computational thinking. This study stands out for: (i) being the only one to employ the GLBoard model, ensuring a standardized structure for log collection; (ii) the analyzed game addresses a topic that is rarely explored in the literature and has a high level of abstraction: optimization problems; and (iii) using LLMs for data modeling, implementation of the GLA structure, and support for data analysis and interpretation by sending prompts with game logs and identifying insights from the responses.

Table 1. Comparison between this study and related works.

Study	GLA Analysis	Computing Game	GLBoard	LLMs
Alonso-Fernández et al. (2019)	X	–	–	–
Alencar et al. (2020)	X	–	–	–
Gil et al. (2024)	X	X	–	–
This study	X	X	X	X

3. Methods

In this research, we investigate the insights that can be gained by analyzing logs from the game "OptimaCorporation" and how these insights can help identify evidence of learning. Based on this, we conducted an empirical study, described below.

3.1. Objective, context and participants

Learning in computing is essential, especially in light of Law No. 14,533/2023, which made the inclusion of computing concepts mandatory in both elementary and high school education in Brazil [Brasil 2023]. Therefore, we need to understand whether computing students are correctly learning these concepts, as they are fundamental to their academic and professional development. GLA can support this understanding through a set of techniques for collecting, analyzing, and interpreting players' interaction records (logs). This process helps identify evidence of learning and motivates this research, in which we applied the game "OptimaCorporation" and analyzed the players' logs.

At State University of Amazonas (UEA), the game content, covering optimization problems such as "NQueens", "KnightTour" and "ShortestPath", is usually taught in the course Algorithm Design and Analysis (ADA), which requires a high level of abstraction. To validate whether the game is suitable for supporting the understanding of these concepts, we invited undergraduate and graduate students affiliated with UEA. These students were part of a research and development lab focused on educational games. They possessed knowledge of computing concepts and experience in educational game design.

We selected the participants based on these criteria, as well as convenience. We did not limit the study to students who had already taken the course, in order also to capture perceptions from participants with no prior contact with the topics covered in the game. In total, we had eight participants, who are referenced throughout this paper using the codes P1 to P8. The demographic and academic profiles of the participants are: (i) gender – 50% (n=4) identified as male and 50% as female; (ii) age range – between 23 and 27 years old; (iii) education level – four participants (50%) had already graduated, three with a degree in Computing Education and one in Computer Engineering, while the other four (50%) were undergraduate students in computing, two in Computing Education and two in Information Systems; (iv) academic semester – among the undergraduate students, they were enrolled in the 4th, 8th, and 10th semester. Two participants had not yet taken the ADA course, and only one had no experience with educational games.

3.2. Procedures

In this study, we carried out five procedures (Figure 1).

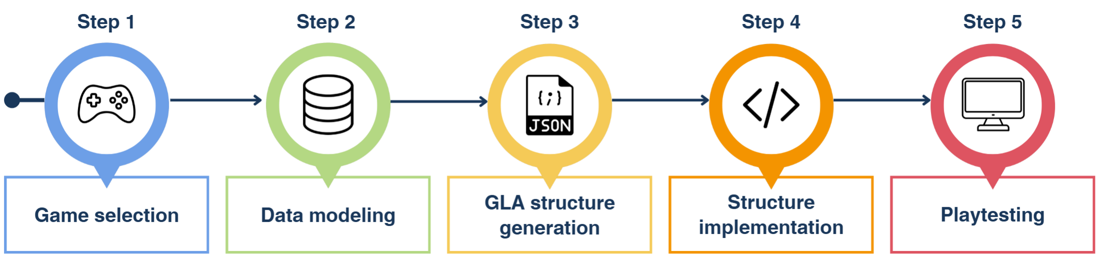


Figure 1. Steps carried out in the study.

Game selection: The first step involved selecting a computing-related game based on the following criteria: (i) inclusion of one or more computing topics from the curriculum component of the UEA; (ii) simple mechanics to facilitate data modeling, log collection, and analysis; (iii) an associated publication available in open-access databases; (iv) a demonstration video available on YouTube; and (v) prior knowledge of the game by the authors. Based on these criteria, we selected "OptimaCorporation", an educational game designed to reduce the perceived complexity of optimization problems

[Martins et al. 2025]. The game features three main stages, each with three levels (nine in total): (i) N-Queens – where the player must allocate servers on a checkered grid without overlapping their movement areas; (ii) Knight's Tour – where the player must move across the board in an "L" pattern without repeating any squares; and (iii) ShortestPath – where the player must follow the fastest route from the starting point to the destination. "OptimaCorporation" is designed for Windows devices and enables players to progress through interactive mechanics tied to learning aspects, while assuming the role of the story's protagonist, thereby stimulating cognitive development. We contacted the game's creators, who provided the source code for the study. The game was developed in C# using the Unity game engine.

Data modeling: To collect player interaction records, we needed to define which data to collect and justify how they could provide evidence of learning. This process, known as data modeling, is not considered trivial by learning designers [Honda et al. 2025]. The game "OptimaCorporation" had already implemented GLA techniques in its source code using the GLBoard model. However, we identified the need to propose a new data model aligned with the objectives of this study. To address this, the authors of this work, who are GLA experts, analyzed the educational game and developed a new data model. We carried out this process with support from the "GLA Specialist," an agent based on Generative AI that guides the data modeling process through targeted questions and helps clarify concepts related to GLA and GLBoard [Honda et al. 2024]. Since the game has three stages with distinct mechanics, we modeled the data according to the specific characteristics of each one. For example, Table 2 presents the data model for the "NQueens" stage.

Table 2. Data model for the *N-Queens* challenge.

Nº	Variable name	Type	Example	Justification
1	challenge	string	"N-Queens"	Identify the challenge being solved
2	board	string	"5x5"	Indicate the size of the board
3	moves_made	integer	5	Measure the number of actions taken
4	time	string	"28s"	Show the time to complete the level
5	errors	integer	0	Count the number of mistakes made
6	undos	integer	0	Count how many actions were undone
7	stars	integer	3	Assess the player's performance
8	move_index	integer	0–4	Indicate the order of the action
9	used_backtracking	boolean	false	Detect whether backtracking was used
10	was_positioning_correct	boolean	true	Verify if the action was correct
11	server_positioned.row	integer	1 to 5	Indicate the piece's row position
12	server_positioned.column	integer	1 to 5	Indicate the piece's column position
13	timestamp	string	"12:23:27"	Record the time the action occurred

GLA structure generation: After defining the GLA variables, we incorporated them into the data template of the GLA model. In this step, we also utilized the "GLA Specialist", as one of its roles is to assist in generating data capture structures. Based on the data model we had created, we asked the agent to organize the variables within the GLBoard template. Figure 2 shows part of the structure, which includes the $finalized\_challenges$ (atomic variables calculated after gameplay ends and other data) and the data belonging to $path\_player$ (recorded during gameplay and related to the player's path through each phase). The complete structure is available via link[1].

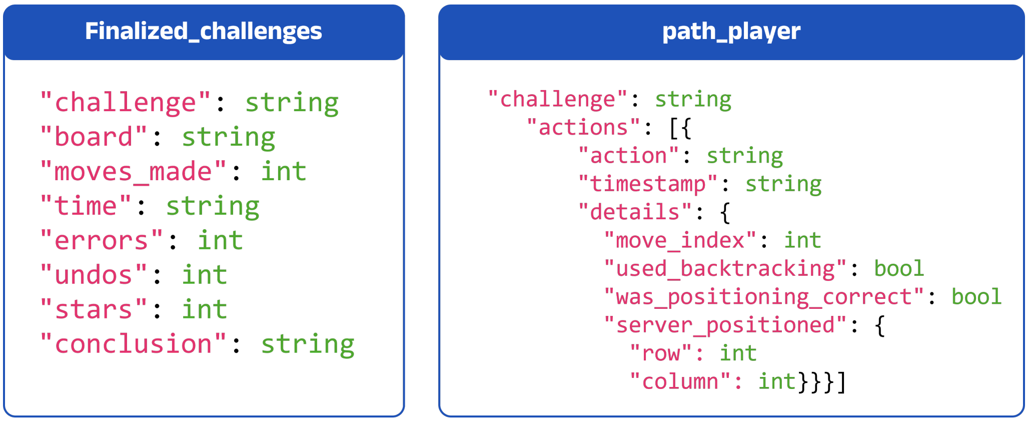


Figure 2. GLA structure from "OptimaCorporation".

Structure implementation: Next, we proceeded with implementing the capture structure in the game to enable interaction data collection. Since "OptimaCorporation" was already using GLBoard and had a previously implemented structure, we reused some GLA variables and code snippets, making specific adjustments to reflect the new structure. One of the authors carried out this implementation under the supervision of the others, using the Unity game engine, version 2022.3.50f1. He analyzed the game's C# codes and the scene's components, and then created and updated the GLA variables according to the structure defined in Figure 2. We conducted unit tests to ensure that the log collection was functioning correctly and generating appropriate, non-empty data.

Playtesting: After successfully implementing the GLA structure, we invited participants to test the game in a controlled environment – a research lab at the UEA. Each participant carried out the test individually, under the supervision of the GLA experts (the authors of this work). We chose this strategy to provide support in case any questions arose during the test sessions. Using a computer provided by the authors, participants explored the game and ended their gameplay sessions at their discretion. There was no requirement to complete all nine levels or to play within a specific time limit. After completing, they filled out a Google Forms questionnaire with basic demographic information, including name, gender, date of birth, education level, degree program, and current academic term. We emphasize that all participants agreed to share their data anonymously and exclusively for scientific research purposes. Furthermore, we ensured compliance with research ethics, and the study was approved by the Research Ethics Committee (CEP), under the CAAE number 85800424.8.0000.5020, as indicated in approval statement no. 7.465.500.

3.3. Data collection and analysis

We recorded and stored participants' interaction data in accordance with the implemented GLA structure. The tracker generates these records in JSON format and submits them to the GLBoard platform API. We accessed individual logs for each participant via URL and stored them in PDF format for analysis[2]. We obtained more than 3,500 lines of code

[1]Full structure link: https://doi.org/10.6084/m9.figshare.30826844

[2]Researchers interested in accessing the anonymized logs may contact the authors via email.

from the eight participants, mainly comprising the variables $finalized\_challenges$ and $path\_player$. Among these logs, 712 correspond to movements performed during the stages (Table 3). All participants played the "NQueens stage", and 63% (n=5) explored the "Knight's Tour" and "ShortestPath" stages. Participant P6 showed the highest level of game usage, whereas P1 recorded the lowest number of gameplay logs.

Table 3. Consolidated movement log counts per stage.

Stage	P1	P2	P3	P4	P5	P6	P7	P8	Total
NQueens	16	24	23	22	11	21	49	22	**188**
Knight's Tour	0	0	0	63	69	156	48	52	**388**
ShortestPath	0	29	0	25	32	25	0	25	**136**
Total	**16**	**53**	**23**	**110**	**112**	**202**	**97**	**99**	**712**

To analyze the logs and extract relevant insights that could support the identification of learning evidence, we decided to use LLMs, in addition to the perspectives of GLA experts (the authors of this study). LLMs are Generative AI models capable of generating natural language text and performing a wide range of tasks in a human-like manner [Kasneci et al. 2023]. The work by Bastos et al. [2025] demonstrated that these models have strong potential to contribute to the analysis and interpretation of GLA logs, especially in player-specific analysis, identifying trial-and-error behavior, analyzing the $path\_player$ field, generating pedagogical insights, among other tasks.

Therefore, we selected Gemini[3] to analyze the logs from "OptimaCorporation", as it proved to be the most efficient tool in the work of Bastos et al. [2025], and we used the prompt provided by the authors in their article[4]. For a more detailed analysis, we used the "Fast" version (Gemini's default) and the "3 Pro" version (which employs complex reasoning). We started a conversation with the model, sent the prompt, and uploaded a file containing the logs of all eight participants. Additionally, we sent each participant's log file individually in separate conversations, as the JSON files had many lines and the LLM could potentially overlook important information. We then continued interacting with the model to extract insights, collected the responses we deemed relevant to the research, and conducted our analyses. In some cases, it was necessary to start new conversations and resend the prompt and log files, since the LLM began to "hallucinate": generating imprecise and incoherent responses [IBM 2023]. We emphasize that the use of Gemini was solely to support interpretation, under the supervision of the authors (GLA experts).

4. Results and discussion

In order to answer the first part of the RQ, "What insights can we extract from data collected from a game about optimization problems?", we organized the results into insights: (i) about the stages, (ii) about the participants, (iii) related to the archetypes (behavioral profiles), and (iv) pedagogical conclusions for the "class" and for the game designers.

Table 4 presents a comparison between the stages, indicating the number of players, success rate, and average, maximum, and minimum attempt times. We can observe

[3]Available at: https://gemini.google.com/
[4]Available at: https://drive.google.com/file/d/1mNqhLGIHLyldGn4NNS8eoeZY66b21rpq/view

that: the "KnightTour" contains the most complex set of levels, with the lowest number of players and the lowest success rate. This point relates to the very nature of the problem: the player cannot reaccess houses he has already visited. Therefore, an incorrect decision can compromise the entire journey and lead to defeat. While this complexity is associated with the problem, it suggests that the game design needs adjustments to minimize it.

In contrast, the "ShortestPath" shows 100% success in all levels and the shortest attempt times. In this case, the levels may be too simple for players, lacking the challenges needed to keep them engaged. The "NQueens" stage appears to be the most balanced, having been tested by most players and showing varied success rates. However, the high gameplay time may indicate that solving these levels is not trivial.

Table 4. Performance per stage and difficulty level.

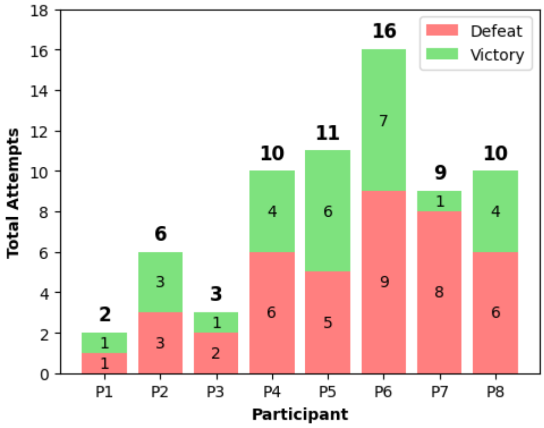
Stage	Level	Players	Success (%)	Avg. Time	Max. Time	Min. Time
NQueens	1	8	50.0%	97.2 s	244 s	30 s
	2	6	10.0%	155.3 s	336 s	74 s
	3	1	100.0%	35 s	35 s	35 s
Knight's Tour	4	5	6.2%	32.4 s	84 s	15 s
	5	1	0.0%	48.3 s	85 s	19 s
	6	0	–	–	–	–
ShortestPath	7	5	100.0%	9.5 s	22 s	5 s
	8	5	100.0%	13.2 s	34 s	5 s
	9	5	100.0%	6.8 s	13 s	4 s

Regarding the participants, Figure 3 displays, respectively, the stacked bar charts: (i) showing the number of attempts; (ii) showing the averages and correct moves in the "NQueens" stage[5]; and (iii) showing the time spent in each stage (in seconds).

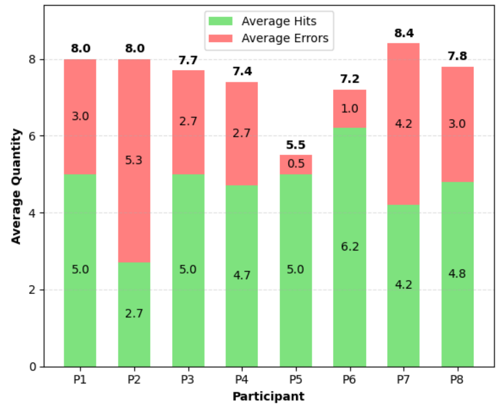
By cross-analyzing the graphs in Figures 3(a) and 3(c), we observed that P1 remained in gameplay for almost 10 minutes, making only two attempts and exploring only the "NQueens" stage. This point may indicate cognitive overload or demotivation due to the complexity of the levels. P2 played two stages, focusing on "NQueens", where they spent more than 6 minutes, suggesting they needed more time to devise strategies. P3 was one of the participants who played the least (under 4 minutes and only three attempts in "NQueens"), which may indicate a lack of interest in the game. This point could be related to their academic background, as they had not yet taken the course that comprises the game's content. P4 and P8 showed a good gameplay rhythm, with a balanced number of attempts and wins/losses, and played for over 8 minutes. P6 was the most active player and one of those who spent the most time playing, suggesting they were motivated to complete all the game's stages. P7 also had one of the longest playtimes: nine attempts across the stages and over 12 minutes of gameplay. However, they did not play the "ShortestPath" stage.

Regarding the "NQueens" levels, most participants had a higher average of correct moves than mistakes, indicating an understanding of how to solve the problem. However,

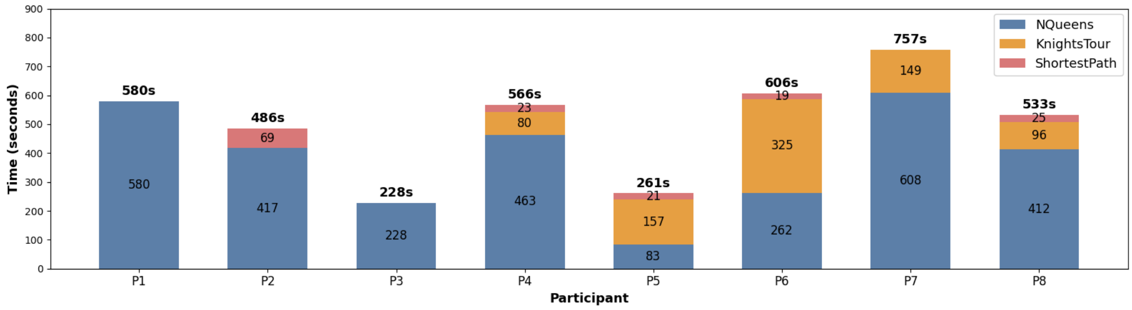
[5]Only the "NQueens" levels allow per-move error tracking. In "KnightTour" and "ShortestPath", we identify errors from the final solution.



(a) Attempts per participant.



(b) Average number of correct moves and errors (“NQueens”).



(c) Time spent in the game per stage.

Figure 3. Graphical summary of participants’ performance.

this was the set of levels with the longest gameplay time (average $>$ 6 minutes), especially for P1 ($>$ 9 min), P7 ($>$ 10 min), and P4 ($>$ 7 min). This point may suggest that, although they understood the rules of the stage, the resolution itself is not trivial. P5 and P6 stood out positively: they had the highest averages of correct moves, lowest averages of mistakes, and relatively short completion times ($<$ 2 min and $<$ 5 min, respectively). P7 showed balanced averages, but took the longest time to progress through the levels, displaying a slower but more cautious gameplay rhythm. On the other hand, P2 is a case of concern: they had a higher average of mistakes than correct moves. They took nearly 7 minutes to complete the stages, suggesting difficulty in understanding them.

Regarding the participants’ behavioral profiles, we identified four main archetypes: (i) **Strategist** – shows efficiency in simpler challenges and high engagement (in terms of time and attempts) in more complex stages. P6 fits this profile, being the only participant to complete the “KnightTour” (in 37s) and Level 3 of “NQueens” without errors. Additionally, they made multiple attempts on the “KnightTour” levels; (ii) **Reflective** – spends a significant amount of time analyzing the problem before taking action, with a low click frequency. P1 fits this archetype, with long gameplay times in Levels 1 and 2 (“NQueens”): 4 minutes and over 5 minutes, respectively. While other participants also had long total times, theirs were spread across multiple short attempts, whereas P1 remained continuous in fewer sessions; (iii) **Intuitive/Visual** – shows greater ease in stages involving direct visual perception, such as “ShortestPath”. P4 and P5 are examples, solving those stages in under 10s. P2 also belongs to this group, completing the “ShortestPath” stages successfully (with a record time of 13s in Level 9). We can also include P3, solving the “NQueens” (5x5) problem in 47 seconds. However, they struggled with the 6x6 level, suggesting that visual intuition alone was not enough; and (iv) **Trial and Error** – reflects learning through repetition, where interaction with the stages did not necessarily involve a defined strategy. P4 and P5 are clear examples, with high click frequency. P8 and P7 also fit this profile, due to their persistence through multiple attempts with high error rates, as they aim to complete the same stage.

Although the participants did not belong to the same class, we can draw pedagogical conclusions from the group to support classroom improvement. For example, the participant group was heterogeneous, with four identified archetypes. Adaptive strategies based on behavioral profiles may include: (i) for reflective students (P1), sets of gradually timed exercises starting with simpler tasks; (ii) for intuitive learners (P2, P3), visual questions to help solidify presented concepts; (iii) for the trial-and-error group (P4, P5, P7, P8), activities that require prior planning to avoid impulsive actions; and (iv) for strategists, more complex challenges to maintain engagement through balanced difficulty. Regarding the game designer, we observed that the “KnightTour” levels led to demotivation and abandonment, indicating that the difficulty level was too high. This point may suggest a need to adjust the mechanics or difficulty to balance the challenge. On the other hand, in the “ShortestPath” levels, students spent little time. They were able to complete the stages quickly, which suggests that these levels were too simple and may require an increase in difficulty to foster greater engagement and challenge.

For the second part of the research question, regarding “how can the insights reveal evidence of learning in computing?”, we constructed Table 5. Based on the log analysis and the previously insights, we can identify evidence of learning related to difficulties in understanding the stages and content, as well as progression during gameplay. This point enables us to understand the learning curve of each student and provide adaptive interventions accordingly.

The insights help to “shed light” on whether students are learning the content, in line with the game’s learning objectives. Although they provide valuable information about the levels, participants, archetypes, and pedagogical conclusions, as well as contributions to game designers, they should not be the only strategy for assessing learning. Without complementary tools such as interviews, observations, learning perception questionnaires, and data science techniques, the conclusions, although relevant, are not sufficient. For example, in this study, (i) P1 was inactive for several seconds. Was this because they were thinking of a strategy, got distracted, or took a break during testing? (ii) When participants undid actions in the levels, does that mean they consciously reversed an incorrect move, or were they testing the feature? (iii) We categorized P7 under the “trial and error” archetype, despite being a graduate. That is, they possess the necessary knowledge. We cannot conclude whether their strategies resulted from exhaustion on the test day, an intent to identify bugs, or other factors. These points highlight the importance of using additional assessment strategies to obtain more accurate insights [Alonso-Fernández et al. 2021, Silva et al. 2021].

Despite the need for complementary tools to support LLM analysis, this study’s

Table 5. Table with the learning difficulties and evidence for each participant.

No.	Identified Difficulty	Gameplay Evolution
P1	In Stage 2 ("NQueens"), 91s of inactivity followed by the exact repetition of a mistake (R6, C4) suggest difficulty in sustaining logical reasoning on the larger board (6x6)	Won Stage 1 using the "Undo" function to fix mistakes and began Stage 2 with a sequence of 5 correct moves in 47s, showing understanding of how to solve the problem
P2	In "NQueens" (Stage 1), had three consecutive defeats. The last one showed a higher error rate, suggesting a reliance on trial and error ("guessing")	Progress in the "ShortestPath" stage: although the solution in Stage 7 was inefficient, it was optimized after two sessions, achieving top performance in Stage 9 (3 stars) with a record time of 13s
P3	In Stage 2 (6x6), made placement errors even when using backtracking. Also had 40s of inactivity before a sequence of mistakes, indicating cognitive overload	After a failed attempt in Stage 1, showed resilience by optimizing the next solution, securing a win and reducing the resolution time from 89s (loss) to 47s
P4	In "Knight's Tour", after a sequence of failures, increased pace through trial and error. In "NQueens" (6x6), persisted for over 4 minutes and gave up, suggesting lack of motivation	Showed resilience in "NQueens" (5x5): achieved a perfect win after 6 consecutive defeats. Solved "ShortestPath" stages in under 10s, demonstrating problem comprehension
P5	In "Knight's Tour", was defeated in five consecutive attempts. The "fast-click" pattern (moves under 1s) suggests trial and error without a logical strategy	In "NQueens", corrected previous mistakes and reduced completion time by 43%. In "ShortestPath", applied optimal solutions across stages, indicating content mastery
P6	In "Knight's Tour", could not proceed on the 6x6 board, accumulating seven failed attempts before giving up. The problem's complexity may have caused demotivation	Showed improvement from Stage 2 (6x6), solving it in 2 minutes, to under 35s in Stage 3 (7x7). In "ShortestPath", achieved optimal solutions across stages
P7	In "NQueens" (5x5), made diagonal placement errors. Had a low success rate (16% in six attempts) and 0% success in "Knight's Tour", indicating more mechanical than logical actions	Persisted after six losses in "NQueens", achieving a win and reducing completion time from 112s to 49s. In "Knight's Tour", although not completing the challenge, showed precision in movements
P8	In "NQueens", the efficient strategy used in 5x5 did not transfer to 6x6: two consecutive defeats (one lasting nearly 3 minutes) with multiple errors, suggesting cognitive overload	Showed a strong learning curve after an initial defeat with five errors and 94s in Stage 1 (5x5), later achieving a perfect win in 65s

findings converge with those of other works in the literature. For example, the results of the research by Alencar et al. [2020] and Gil et al. [2024] indicate that log analysis can identify problems in level design, error patterns, and player difficulties. We identified similar evidence, such as the need for adjustments at the "Knight's Tour" level, the tendency toward longer resolution times in the second level of "N-Queens", and difficulties that may indicate cognitive overload. In Alonso-Fernández et al. [2019], the authors map the use of data science in GLA, highlighting data visualization techniques (such as dashboards) to support the analyses. The research by Bastos et al. [2025] investigates the use of LLMs to assist in log analysis and generate pedagogical insights. Similarly, our results may suggest that LLMs contribute to the interpretation of educational game data, just as data visualization can support more robust analyses.

We present as limitations: (i) the data modeling, which may not have been the most appropriate, although experts in the field carried it out with support from the "GLA Specialist"; (ii) the log analyses supported by Gemini, which, even under expert supervision, may not have been the most accurate; and (iii) the small sample size, consisting of only eight participants. Although the study focuses on obtaining insights from GLA data supported by LLM analysis, this limitation may affect the generalizability of the findings.

5. Conclusions

This study posed as a research question (RQ): "What insights can we extract from data collected from a game about optimization problems, and how can they reveal evidence of learning in computing?" We conducted an empirical study to answer it, involving selecting the game, data modeling, creating and implementing the GLA structure, and playtesting with computing students (undergraduate and graduate). Subsequently, we analyzed the logs with the support of "Fast" and "o3 Pro" versions of Gemini.

In response to the research question, the results include insights: (i) regarding the stages, where we observed that the "KnightTour" levels are complex, while the "ShortestPath" levels are simple; (ii) regarding the participants, who showed varying numbers of attempts and gameplay durations; (iii) regarding the archetypes, identifying the profiles of "Strategist", "Reflective", "Intuitive/Visual", and "Trial and Error"; (iv) regarding pedagogical conclusions for educators, with adaptive strategies based on the archetypes, as well as contributions to game designers concerning level balancing. These insights enabled the identification of distinct learning evidence for each participant, related to the challenges of the levels and the students' progression during gameplay. We also conclude that, although the insights were relevant, complementary instruments (such as learning perception questionnaires, observations, and interviews) are necessary for a more accurate assessment of learning.

This work presents the following contributions: (i) the relevance of the GLA field, with a study involving log analysis; (ii) learning analyses of a game focused on optimization problems, tested by individuals from the field; (iii) data modeling and log analysis supported by LLMs, which enhance the quality of activities when used effectively; and (iv) GLBoard, which proved to be a simple and flexible tool for GLA tasks. Future studies include: (i) applying the game in an Algorithm Design and Analysis (ADA) class, conducting analyses and sharing the results with the educators involved; (ii) cross-referencing logs with learning perception questionnaires to improve insights; and (iii) combining data science techniques for more accurate analyses.

6. Acknowlegment

The authors express their gratitude to the State University of Amazonas (UEA) for the institutional support. They are also grateful to their colleagues from the ThinkTEd Lab for the contributions and discussions that enriched this work. The authors further acknowledge the support of the Nexus research center, which provided essential material resources, including physical space, computers, and related infrastructure.

Use of Generative Artificial Intelligence (GAI)

In this study, we used ChatGPT from OpenAI to generate the codes for graphs in Overleaf, aiming to help minimize time and effort in constructing these representations.

References

Alencar, L., Melo, R., Pires, F., Pessoa, M., and Oliveira, E. (2020). Uma proposta de análise de dados exploratória para um jogo educacional de matemática. In *XIX SBGames (Brazilian Symposium on Computer Games and Digital Entertainment)*, pages 752–758.

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

Bastos, M., Honda, F., Lima, M., Pessoa, M., and Pires, F. (2025). How do llms analyze and interpret data from educational games? a study with gla experts. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1317–1330. SBC.

Brasil (2023). Lei nº 14.533, de 11 de janeiro de 2023. Diário Oficial da União. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/lei/l14533.htm. Acessed: 2025-12-01.

Castro, F. R. and Cruz, F. M. d. (2022). Ensino de lógica de programação através de problemas abstratos com conteúdo de geometria plana. *Research, Society and Development*, 11(9):e58811929683–e58811929683.

Egenfeldt-Nielsen, S. (2006). Overview of research on the educational use of video games. *Nordic journal of digital literacy*, 1(3):184–214.

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., 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.

Gil, A., Figueira, T., and Netto, J. F. M. (2024). Extracting learning analytics from lightbot gameplaysessions. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*, pages 1138–1150. SBC.

Honda, F., Pires, F., Pessoa, M., and Oliveira, E. H. (2024). Building a specialist agent to assist in the implementation of game learning analytics techniques. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 2856–2865. SBC.

Honda, F., Pires, F., Pessoa, M., and Oliveira, E. H. (2025). Challenges in educational game data modeling from the perspective of computing students: an empirical study. In *Workshop sobre Educação em Computação (WEI)*, pages 1251–1262. SBC.

IBM (2023). Ai hallucinations. https://www.ibm.com/topics/ai-hallucinations. Accessed: 2025-04-30.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274.

Lima, J. R. d. and Menezes, C. S. d. (2024). As dificuldades enfrentadas pelos estudantes na aprendizagem de programação de computadores: Uma revisão sistemática da literatura. *Revista Novas Tecnologias na Educação*, 22(1):130–140.

Macena, J., Pires, F., Pessoa, M., and Melo, R. (2022). Hello food: um jogo para praticar conceitos de algoritmos para iniciantes na computaçao. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*, pages 1066–1075. SBC.

Martins, W., Honda, F., Pires, F., and Pessoa, M. (2025). Optimacorporation: reduzindo a percepção de complexidade em problemas de otimização por meio de um jogo educacional. In *Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames)*, pages 486–491. 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.

Oliveira, G. C. d. and Boff, E. (2024). Aprendizagem pelos jogos: uma abordagem para o ensino de programação. *Scientia cum Industria*, 13(2):e241318–e241318.

Pires, F. G. d. S. et al. (2021). Thinkted lab, um caso de aprendizagem criativa em computação no nível superior.

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

Ramos, M. V. M. (2009). Ensino de linguagens formais e autômatos em cursos superiores de computação. *Revista de Computação e Tecnologia (ReCeT)*, 1(1):22–34.

Santos, A. J. N. d., Santos, J. C. C., OLIVEIRA, K. C. d., PAULA, L. O. d., Domingues, M. A., and Salomão, T. L. (2024). Educação digital: os desafios do ensino de programação.

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.

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

Silva, F. X. d., Menezes, C. S. d., and Junior, A. N. d. C. (2025). Dificuldades dos estudantes na aprendizagem de programação: Uma revisão sistemática da literatura. *Cadernos de Educação Tecnologia e Sociedade*, 18(3):994–1009.

Silva, M. B. d., Furtado, S. D. d. F., and Furtado, J. C. C. (2024). A evasão dos estudantes em cursos de computação: um estudo sobre a universidade federal do amapá. *Caderno Pedagógico*, 21(9):e7820–e7820.

Vasconcelos, R. C. d. S. and Neto, A. J. M. (2020). A computação no currículo da educação básica. *Revista Eixo*, 9(2):68–76.

Videnovik, M., Vold, T., Kiønig, L., Madevska Bogdanova, A., and Trajkovik, V. (2023). Game-based learning in computer science education: a scoping literature review. *International Journal of STEM Education*, 10(1):54.