Evaluation of a specialist agent in Game Learning Analytics by learning designers: a case study

Fabrizio Honda[1,2], Marcela Pessoa[1], Elaine Harada[2], Fernanda Pires[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)

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

Abstract. *Using Game Learning Analytics techniques, or Learning Analytics for Serious Games, provides valuable insights to stakeholders, allowing the validation of game design and the identification of learning evidence. However, learning designers consider a preliminary step complex: modeling the data to list GLA variables. Using Large Language Models is an alternative in this scenario, which motivated us to create the "GLA Specialist" – an intelligent GLA agent – in a previous work. This research consists of a case study to evaluate it with computer science students from a public university as learning designers. The results were positive, indicating good acceptance of the agent and data models compatible with the GLBoard model, along with a list of corrections.*

1. Introduction

The field of Game Learning Analytics (GLA) [Freire et al. 2016], or Learning Analytics for Serious Games, plays a key role in evaluating serious games. GLA techniques involve collecting player interaction data during gameplay in a non-intrusive way so as not to interrupt the flow. In addition, it includes the visualization and analysis of this data, commonly through tools such as Dashboards. This strategy benefits both game designers, allowing them to analyze whether the game design is adequate and identify possible inconsistencies for correction, and the players themselves, who can use GLA data to monitor and be aware of their progress [Banihashem et al. 2024]. In Educational Serious Games (or educational games), the application of GLA becomes not only necessary but fundamental [Alonso-Fernández et al. 2021, Alonso-Fernandez et al. 2017] since it allows the identification of evidence of learning through mapping the player's path: platforms jumped, objects interacted with, NPCs talked to, among others. Researchers can combine these data with data science techniques and correlate with heuristic assessments (self-assessment, for example), which can provide a deeper analysis of learning progression [Alonso-Fernández et al. 2021, Silva et al. 2021, Alonso-Fernández et al. 2022].

One of the tools that enables the implementation of these techniques is GLBoard [Silva et al. 2022], built to overcome some limitations in the area, such as implementation complexity and lack of standardization [Saveski et al. 2016, Alonso-Fernandez et al. 2017]. Its objective is to assist in implementing GLA techniques to collect, analyze, and visualize data records (logs) generated by student interaction with educational games. However, although researchers applied it in games from literature

DOI: 10.5753/sbie.2025.12894

[Macena et al. 2024, Honda et al. 2023], a step before the implementation of GLA techniques is complex: data modeling, or data selection [Hauge et al. 2014]. This process involves defining which data will be collected and the justification for why they are relevant to help map the learning path and analyze player behavior. The difficulties emerging from this process are the appropriation of the game to understand the learning objectives, abstraction to define the collection variables, dedicated time, and previous computing content that assists in modeling [Honda et al. 2025].

Generative Artificial Intelligence (GAI) is a possible alternative to this scenario, especially using Large Language Models (LLMs). These models generate text like humans and perform countless other tasks, whose researchers have been applying them in diverse domains: robotics, medicine, finance, law, and education, among others. [Kasneci et al. 2023, Naveed et al. 2023]. Given its capabilities, researchers are also applying it to Learning Analytics (LA) [Misiejuk et al. 2025] and Serious Games [Mitsea et al. 2025]. Recently, we also observed its application to the GLA area: the creation of the "GLA Specialist", an intelligent agent created in ChatGPT [Honda et al. 2024]. It consists of a customized chatbot with GLA content and the GLBoard model. It can assist in educational games' data modeling processes and insert data into a capture structure (data template) – by default, in the GLBoard JSON format. Despite receiving a positive reception from GLA experts who evaluated it, the authors had not yet applied the agent to the target audience (learning designers).

In this regard, this work aims to conduct a case study to answer the following research questions: RQ1 – "How do computer science students (learning designers) model educational game data with a GLA specialist agent, and what are the results?" and RQ2 – "What are the perceptions of these students when using the GLA specialist agent and when comparing it with the manual modeling technique?". The study's contributions include research that contemplates models based on LLMs, Serious Games, and LA, which is still little explored in the literature, and the evaluation of an AI agent focused on GLA, which learning designers considere a good alternative for modeling data and generating capture structures.

2. Foundations and related work

GLBoard is a model focused on educational games, which provides a set of tools to implement GLA. It consists of four main components: (i) Unity package – contains the capture model, where developers can install it via the Unity game engine; (ii) Database – stores game and player data; (iii) API – manages the other modules, responsible for the main communication; and (iv) Dashboard – displays graphs generated from general metrics and includes the JSON with the collected "raw" data. After installing the package, the developer imports the library through a script, which provides a capture structure (data template) to insert the GLA variables. This template contains four main classes: $PlayerData$ (player information, such as date of birth, gender, etc.), $GameData$ (stores game data, containing an object of type Phase), $Phases$ (a detailed list of the game's phases, which records the phase name, completion status and a set of objects of type $Section$) and $Section$ (a session represents each attempt in a phase. Responsible for storing the phase start and end time, performance, $path\_player$, etc.). Unlike the other variables, which are fixed (generic to all games), $path\_player$ is flexible: the developer can insert any variables they want to capture. Its focus is on storing the player's path within the phases, which vary from game to game.

However, to fill the $path\_player$ in GLBoard and thus compose its data template – or another GLA tool, learning designers must define which data will be collected and its importance to express the player's evolution in the phases. This process is data modeling (or data selection), and studies show that it is not a trivial process, mainly due to the difficulty in abstracting the data to be collected, appropriating the educational game, time allocated for the activity, and prior computing knowledge [Honda et al. 2025]. This process must be carried out from the beginning of the game's design [Hauge et al. 2014, Alonso-Fernández et al. 2021, Kitto et al. 2020], as it directly influences the filling of the data template. When the game is already finished or partially implemented, even experts in educational games and GLA face challenges in inserting the template into their games, such as game abstraction, complex mechanics, adaptation of the structure, and navigation through the code [Macena et al. 2024].

Given its complexity, studies such as Honda et al. [2024] investigate the use of LLMs to assist in this context. The authors built an intelligent agent in ChatGPT entitled "GLA Specialist", whose focus is to guide and assist in the data modeling process, defining the GLA variables and inserting them into a specific data template – by default, from GLBoard. GLA experts evaluated it as a potential tool. However, the authors have not yet applied it to the target audience (learning designers), the focus of this study. We searched the literature to locate related works involving intelligent agents based on LLMs to assist in Learning Analytics and Serious Games, as described below.

The work of Merikko and Silvola [2024] investigates the application of LLMs in educational contexts to facilitate help-seeking behaviors by students. The study begins with a thematic analysis of 263 statements prepared by experts, aiming to build a model for classifying support needs. Then, the authors developed a chatbot prototype from the GPT-4 and WhatsApp APIs to listen to students' concerns, recognize topics about well-being and academic support, and suggest contact with the institution's professors. In addition, the system also provides control over the conversation and support resources. The results demonstrate the efficiency of the LLM in recognizing topics in interactions and facilitating help-seeking, and include a discussion by the authors on how researchers can use the data generated by the chatbot for Learning Analytics.

Wang et al. [2025] propose "GenMentor", a multi-agent framework integrated with LLMs to provide personalized and goal-oriented learning in Intelligent Tutoring Systems (ITSs). The authors implemented the agent in distinct LLMs (GPT-4o and LLaMA 3.2-3B) used individually in the experiments. Each agent initially maps students' goals to the required skills using the corresponding LLM – which was tuned and trained with a customized dataset. Then, the agent programs a learning path and adapts the content through an exploration-writing-integration mechanism to align with the student's specific needs. Evaluations indicate that "GenMentor" helps in learning guidance effectively and offers quality content. Although it does not explicitly mention the term "Learning Analytics", its practices are present in the work: collecting user data, analyzing the data to build dynamic profiles, and suggesting pedagogical interventions based on the data.

Mostachetti et al. [2025] investigates the integration of an LLM in a Serious Assessment Game (ASG) to analyze exercise data and recommend personalized rehabilitation programs for neurological patients. The study included meetings with health professionals to identify target pathologies (Parkinson's, stroke, spinal cord injury, etc.) and elicit clinical parameters. The authors integrated ASG with GroqCloud, using the mitral-8x7b-32768 model, and designed a prompt for the LLM to act as a physical therapist and serious game developer. This way, the model would evaluate performance, adjust the game difficulty in real-time, and suggest configurations for other serious games. The results of the preliminary tests were positive, indicating the model's efficiency in making adjustments in real-time. In addition, it proposed suitable configurations for the games Whac-a-Mole, Mokka Coffee, and Paint On Canvas. However, some limitations stand out, mainly in suggesting numerical values, sometimes described with text by the LLM.

As can be seen in the related works, there is a gap in the literature at the intersection between Serious Games (SG), Learning Analytics (LA), and LLMs/Intelligent Agents (LLM-based). Most meet only two criteria, as shown in Table 1. Furthermore, most of the works that consider LLMs do not involve the design of intelligent agents but rather the use of known models (ChatGPT, for example) in specific domains. This work seeks to foster research at this intersection, focusing on evaluating an expert agent based on LLM for Learning Analytics in Serious Games/Game Learning Analytics.

Work	**SG**	**LA**	**LLM-based**
Wang et al. [2025]		X	X
Merikko and Silvola [2024]		X	X
Mostachetti et al. [2025]	X		X
This research	**X**	**X**	**X**

Table 1. Comparison of related works with this paper.

3. Methods

This work consists of conducting a case study to investigate the use of a GLA specialist agent by learning designers. This approach is an empirical investigation of a case, aiming at studying a contemporary phenomenon in its real-life context through multiple methods to collect data, in which the researcher does not assume an active role [Wohlin 2021].

3.1. Goal, context and participants

This research aims to analyze (i) the experiences of computer science students (learning designers) in data modeling for educational games using "GLA Specialist" and (ii) their perceptions when comparing this process – using the agent – with the manual data modeling technique. The research context is at the intersection between GLA and LLMs, an area still little explored in the literature, which we seek to foster through a case study.

In a previously empirical study we conducted [Honda et al. 2025], computer science students at the Amazonas State University (UEA) performed manual data modeling for educational games they were developing as part of a curricular course. We considere these students learning designers for two reasons: the majority (75%) (i) attend in the Bachelor's Degree in Computer Science Education, which aims to train professionals capable of designing and applying learning objects in educational contexts; and (ii) participate in an educational technology research and development laboratory, where they work on developing learning materials. These considerations are also consistent with the literature, as, in the Brazilian context, students and educators primarily develop many educational games in academic settings [Tarouco et al. 2005, Portella et al. 2017, Cordeiro and Duarte 2020, Nascimento and Leite Bruno 2024]. Therefore, it is common for students to develop educational games in their courses, assuming multiple roles, such as developer, designer, tester, and learning designer. In this sense, computer science students are also relevant stakeholders in this process and should participate in data modeling and implementing GLA techniques. This scenario is, therefore, the contemporary phenomenon investigated in this research, in which the case analyzed focuses on a specific group of participants with little experience in GLA.

Since one of the objectives of this study is to compare manual modeling with that performed by an intelligent agent, the only criterion we adopted for participant selection was participation in the previous study. Therefore, in this case study, we selected nine people: six (66%) men and three (34%) women, aged between 19 and 24, and 100% were undergraduate students in the Computer Science Education degree program at the Amazonas State University (UEA). Regarding experience, 56% had already developed educational games, and only one person (11%) had implemented GLA techniques/modeled data for games.

3.2. Procedures

We carried out this study in four stages (Figure 1).

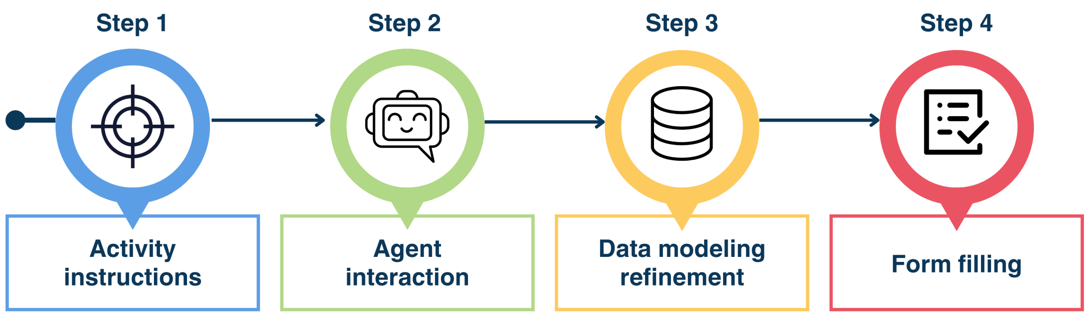

Figure 1. Case study stages.

Activity instructions: the first step consisted of communicating the study instructions to the participants. Since we conducted the research in the classroom, we carried out the study through activities with the students. Therefore, we addressed topics such as a summary of the previous research [Honda et al. 2025], activity requirements, and additional observations. We informed the participants that the activity consisted of modeling data for their educational games using the "GLA Specialist" and completing an evaluation form afterward. Among the observations we shared to participoantes were: (i) they should modeling the data until judged it to be sufficient; (ii) interaction with the agent was free; (iii) they could one used as needed or more interaction chats, as long as they register the links; and (iv) they should fill the form after modeling the data.

Agent interaction: although ChatGPT offers free versions, using "GLA Specialist" agent requires ChatGPT Plus (paid account). Therefore, we inserted a specific research group account into the machines so the students could use the agent. Thus, the participants began the activity by accessing the "GLA Specialist" and starting their interactions with the chatbot individually. The participants generally informed the agent that they wanted to perform data modeling for their educational games. Then the agent asked for the characteristics of these games – giving examples, in some cases – so that it could understand them and assist in the process. Thus, the students began preparing simple prompts with information about their games, such as name, target audience, theme, content, related BNCC skills, gameplay, story, learning mechanics, etc. Since the agent does not require complex prompts, as it is a GLA specialist, students did not need to worry about detailed instructions to the model; they asked it to model data and provided information about its games. The students then sent the simple prompts to the agent, who returned a table with the data modeling of the games to the participants, including variable number, name of the GLA variable, collection example, and justification for capture.

Data modeling refinement: after the agent generated the data modeling tables, the participants continued to refine them. The refinements included explaining the mechanics of other phases, requesting explanations about defined variables, indicating improvements and inconsistencies in some data, excluding or modifying some fields, and adding more details about specific aspects. Considering this and the previous phase, the total interaction with the agent lasted approximately one hour and thirty minutes.

Form filling: after completing their modeling, the students filled out the form using Google Forms. They recorded the links of their interactions with the intelligent agent, described their experience using it, and described their perception when comparing it with the manual data modeling technique in the previous study. This formulary is available at, including its questions and options[1].

3.3. Data collection and analysis

In order to be considered a case study, the research must involve multiple data collection methods [Runeson et al. 2012]. In this regard, the data collected in this study include (i) empirical data – the students' interactions with the "GLA Specialist"; (ii) objective measures – quantitative assessments of the specialist's use of data modeling and its comparison with the manual technique; and (iii) subjective perspectives – qualitative assessments of the students with justifications for the quantitative data. In addition, the research also aggregates data from the previous study [omitted for review], including (iv) documentary evidence – the data models manually prepared by the students, in spreadsheet format. Therefore, the aim is to triangulate these data to increase the study's validity. Regarding the evaluation instrument, the students formally consented via the form to make their data available anonymously for research purposes only. Thus, this study respects the ethical principles of human subjects.

Regarding data analysis, we used (i) a stacked bar chart about the proposed variables and their belonging to the $path\_player$, about the modeling by the agent and the students, as well as a frequency table with the GLA variables most defined by the agent; (ii) a boxplot chart for the quantitative questions about the perception of the use of the

[1] https://drive.google.com/file/d/1sPCoLZwDt5Pxa-NP3kYFKCl1aotwpge8/view?usp=sharing

agent and comparison with the manual technique; and (iii) word cloud and content analysis [Bardin 2015] for the qualitative questions – which we applied by reviewing the participants' answers to identify recurring ideas. Based on this review, we defined thematic categories to group similar responses. We then assigned each response to a category and examined them to highlight common patterns and insights specific to each group.

4. Results and discussions

The results of this case study include (i) the analysis of the students' interactions with the expert agent, and (ii) the data models that students generated with the help of the agent – both we used as a basis for answering RQ1; (iii) the analysis of the students' perception regarding the experience in using the agent, and (iv) a comparative evaluation of the perspective of use between the expert agent and the manual modeling technique – both we employed to answer RQ2.

Regarding the first part of RQ1, students model data with "GLA Specialist" in different ways: the beginning of the conversation is similar, informing the agent that they want to model data, but the interactions differed – both on the part of the students and the agent. Most students presented only three interactions[2] with the chat. P6 and P8 had seven interactions with the agent, which proposed the most significant number of variables (20). This aspect can suggest that the greater the number of interactions, the greater the number of variables proposed. Although P5 was the participant with the most significant interaction (13), it was necessary to use two chats due to the limitation of ChatGPT, slightly compromising the modeling process. As for the agent, it requested different amounts of game elements from the participants: 10 data in two cases and 8, 7, and 6 data in other interactions. We also observed an atypical behavior: on one occasion, the agent did not request game elements – although they are present in its self-description – but recommended steps for modeling. In another case, it did not explicitly request data but indicated general variables for collection.

In the second part of RQ1, regarding the results of the interactions, all instances of "GLA Specialist" were able to build the data modeling table for the participants. The agent defined general variables (total and aggregated, calculated after a session's end) and variables belonging to the $path\_player$ (referring to the player's path through the phases). To aid this analysis, we designed the stacked bar chart in Figure 2. The chart compares the data modeling the students performed manually in the previous study (green) with the modeling they carried out using the "GLA Specialist" (blue). The Y-axis contains the number of GLA variables the agent propose and the X-axis represents each participant – note that we removed P3 and P4, as the links to access the interactions presented errors. The internal bars represent the number of variables corresponding to the $path\_player$.

In all cases, the "GLA Specialist" agent proposed more variables than the students – averages of 16.3 and 5.4, respectively. In manual modeling, the students defined at most nine GLA variables, while most defined only four or five. We also note that the "GLA Specialist" faced challenges when proposing variables corresponding to the $path\_player$ – e.g., in P5 interactions, which proposes two GLA variables to store the same data. Both results confirm that data modeling is complex, revealing difficulties for both the students and the specialist agent. However, they also highlight the agent's potential in this process:

[2]Each message sent by the student to the model is considered an interaction.

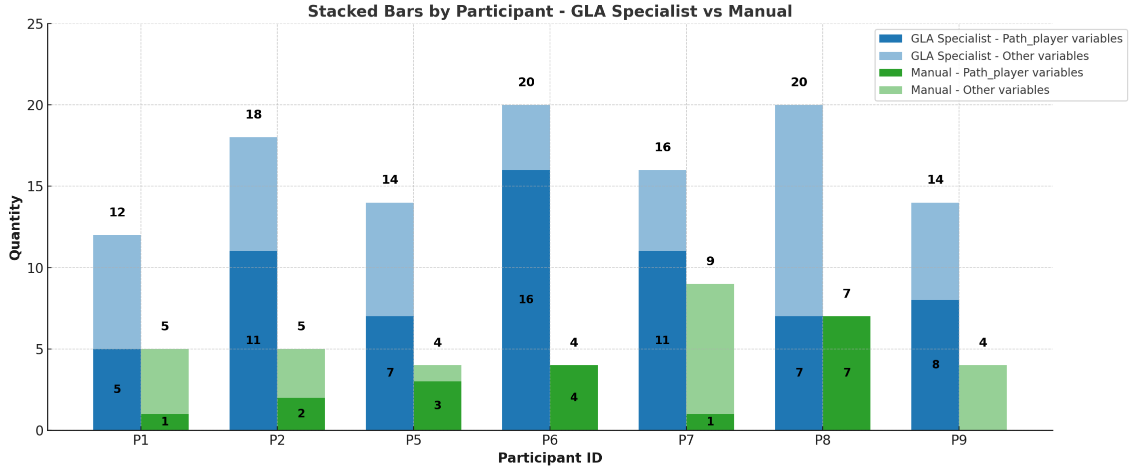

Figure 2. Comparison of models regarding the proposed GLA variables.

it analyzed nine games and generated corresponding variables, demonstrating flexibility and replication. We also observe standard variables in the models the agent generated for nine games, especially generic ones and those external to the $path\_player$. Table 2 corroborates this information by showing the most frequent variables. The agent achieved a positive result by proposing relevant variables for analyzing the player's learning. On the other hand, the agent proposed most variables in only 30% of the games, which indicates a lack of consistency in its models. Despite the distinctions between the games, learning designers can apply 89% of these variables to all of them.

Table 2. Most frequent variables in the agent's data modeling.

No.	General Name	Freq.	Original Names	Capture Justification
1	total_time	5	time_spent, total_game_time, completion_time, preparation_time	Capture the total time that players spend in stages or specific activities
2	current_stage	3	current_stage, phase, phase_name	Identify the player's current stage and track their progress
3	feedback	3	player_feedback, received_feedback, game_feedback	Record the responses that the game provides to the player
4	score	3	stage_score, points_earned, performance_score	Evaluate the player's performance in tasks or specific stages
5	player_actions	3	actions_performed, decisions_made, learning_actions	Track the decisions that the player makes to analyze behavior and progression
6	collected_items	3	selected_ingredients, collected_items, gathered_resources	Register the items that the player acquires and their quantities to evaluate game completion
7	errors	3	wrong_commands, conditional_error, repetition_error, errors	Detect when the player commits errors to identify inconsistencies or learning difficulties
8	coins	3	coins_earned, coins_received, bonus_coins, lost_coins	Count the coins that the player gains or loses to reflect their performance
9	game_interactions	3	environment_interactions, npc_interaction, npc_interactions	Log the interactions that the player performs with game elements to track the journey more accurately

Regarding RQ2 – students' perception of the agent and its comparison with the manual modeling technique, we created the boxplot in Figure 3. The graph gathers the participants' evaluations of the agent, with the Y-axis representing the score (on a Likert-

5 scale) and the X-axis representing the evaluation criteria. We described the analyses below.

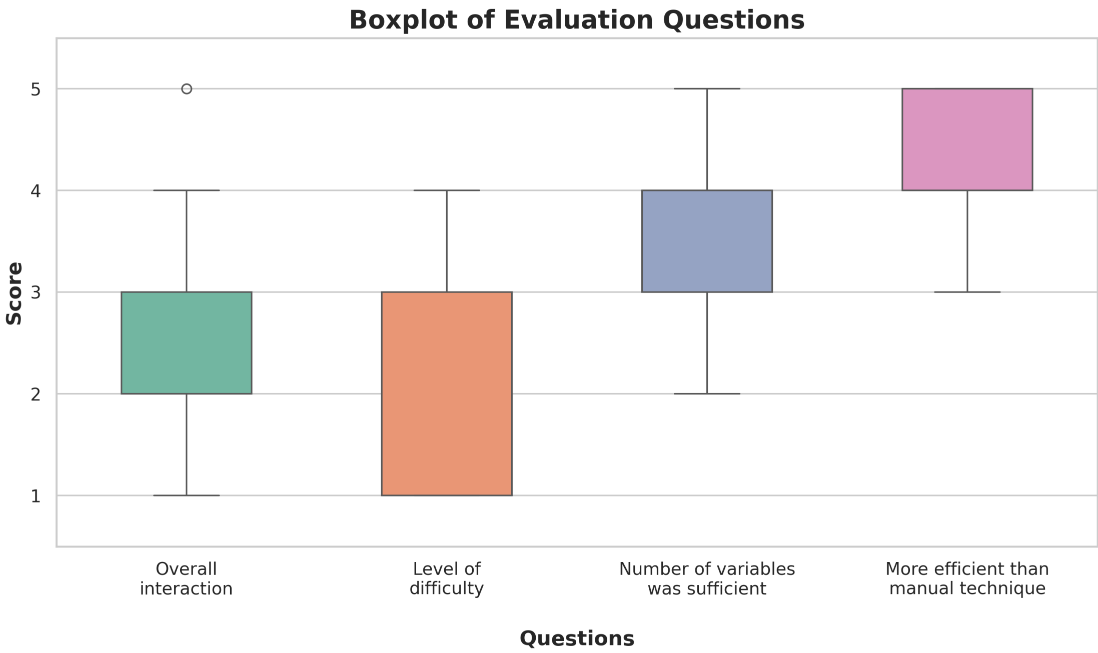

Figure 3. Boxplot of the learning designers' evaluations.

Overall interaction: regarding the level of interaction with the agent, the scale ranges from "very easy" (grade 1) to "very difficult" (grade 5). We found that more than half of the students (56%) considered it easy/very easy, while 22.2% found it "average". The analysis of the participants' justifications indicates positive points: recognition of primary data for collection, ease of use without requiring a specific prompt, precision (comparing it to conventional ChatGPT), assistance in understanding the data modeling process, speed, etc. On the other hand, the following are negative points: limitation of use of ChatGPT, unexpected responses, repetitive texts, presence of unnecessary information, amount of information requested, among others.

Level of difficulty: we noted that six students (67%) found it "easy", two remained neutral (22%), and only one (P8) found it "difficult". However, when we compared P8's evaluation with the justification for the answer and the other questions on the form, we noticed incoherence, which suggests that the student may have confused the scales. Given this, we estimate that 78% found the agent easy to use, and no participant considered it difficult. Regarding the justification for the answers, most students did not point out any difficulty; others indicated that sending the game information to the agent was the challenge, and P9 pointed out that the expert provides less autonomy than the manual technique.

Number of variables was sufficient: this criterion refers to the student's perception regarding the number of variables proposed by the agent in data modeling. 67% of the students considered the number sufficient, and 22% found it insufficient, reporting that they missed specific variables (i.e., queries to the algorithms) and generic variables (i.e., time to act). Regarding the justification for the proposed variables, 78% reported that they were good, and they indicated that the agent helped identify data they had not considered, such as difficulty level, time, attempts per phase, buttons clicked, errors, decisions, and other game-specific variables.

More efficient than manual technique: refers to the comparison of data modeling generated by the expert agent (this study) with that prepared manually by students (previous research [Honda et al. 2025]). From the participants' perspective, 89% believed that the agent was more efficient and preferred it as a method for modeling data, for practicality and speed of responses, identifying data that went unnoticed, agility in data modeling, and generating more ideas. Only one participant (P7) stated that he did not have a preference for the data modeling method, pointing out that it depends on the need. Regarding confidence in the final result of the modeling, the majority of participants (89%) felt more confident when using the agent because the agent indicated more variables, provided a broader view and a more general analysis of the game, and delivered more precise data on how to implement it. One response that stood out in this context was that of P5, who reported that "with the use of the expert, it is easier for me to make mistakes than an expert.". This point reveals an unrealistic perspective since the connotation of error is precisely what enables an improvement in data modeling, and the agent is LLM-based, which sometimes tends to hallucinate and generate incoherent responses [IBM 2023] – and should be used with responsibility and caution. P7 reinforces this first aspect, who felt more confident with the manual technique due to the "autonomy of thinking about variables". Regarding efficiency, 88.8% indicate that the agent is more efficient than the manual technique. Regarding difficulty, 78% find the manual technique more difficult due to the lack of experience in modeling data, appropriating the game, the complex nature of manual work, limitations on human creativity – an unrealistic point already discussed previously, etc. Two participants (22%) find both techniques challenging due to the difficulty in thinking about/identifying the collection variables.

In general, students also indicated a good acceptance of the intelligent agent in GLA, as did GLA experts in Honda et al. [2024]. 100% of students agreed that the agent is a good option to help model data for educational games. Regarding the positive and negative aspects of the agent, Figure 4 presents a word cloud we generated from students' opinions. The agent's "good" points include accuracy of responses, analysis capacity, guidance in data modeling, ease of use, data organization, assistance with collection ideas, etc. The "bad" points identified include the amount of information requested at once, expiration of use (ChatGPT limitation), repetition of information, failure to display all variables required by the student, and lack of autonomy.

The insights found in this study include: (i) the agent is a good alternative for the data modeling process but still needs refinements; (ii) despite the significant contributions to data modeling, the intelligent agent presented difficulties in the process, reinforcing the fact that data modeling is complex; (iii) despite the ease of use, closed LLMs have limitations such as interaction restrictions, instability and more precise customization of the model (fine-tuning, for example); and (iv) despite not requiring complex prompts, sending information to the agent is still laborious. Furthermore, although the agent facilitates the data modeling process, its purpose is not to replace manual modeling, because, paraphrasing P7, the "autonomy of thinking about variables" is fundamental, and the use

Figure 4. Word cloud of what students think of the agent.

of techniques is “a matter of necessity” since “each method has advantages and disadvantages”. Thus, the protagonist role of the learning designer is essential, who must comprehend their educational game to propose GLA variables. This process is complex, and the intelligent agent’s goal is to minimize such complexity by combining students’ knowledge with an LLM. The connotation of error is necessary in this context, as it allows the learning designer to think of other strategies to overcome it. One of the agent’s roles is to act on the aspect, stimulating the learning designer’s learning from doubts and errors.

5. Conclusions

Applying GLA techniques significantly contributes to evaluating serious games, especially educational ones, allowing the player’s progress to be tracked through the phases and identifying possible evidence of learning. GLBoard enables the implementation of these techniques, providing a generic and flexible capture structure for any educational game. However, data modeling is a complex process involving defining the GLA variables and the justification for collection. LLMs can be a good alternative to minimize this complexity, which motivated creating the “GLA Specialist”: an intelligent agent in ChatGPT to help model data and implement GLA techniques. Positive results with GLA specialists indicated good receptivity, but the target audience (learning designers) has not yet evaluated the agent. In this context, we sought to conduct a case study to answer two research questions: RQ1 – “How do computer science students (learning designers) model educational game data with a GLA specialist agent, and what are the results?” and RQ2 – “What are the perceptions of these students when using the GLA specialist agent and comparing it with the manual modeling technique?”.

For the study, we defined the objective, context, and participants - the selection criterion was having participated in a previous study that involved modeling data manually (without LLM). The study began in the classroom, where we passed the activity instructions to the students. Then, the students started interacting with the intelligent agent to generate and refine the data models. After considering the models sufficient, the students ended the interaction and completed an evaluation form. Data collection included quantitative and qualitative evaluations, as well as the models from this study and the previous one. We used graphs (stacked bars and boxplots), frequency tables, content analysis, and word clouds to analyze the data.

The results for RQ1 indicate that students model in different ways: the process starts similarly, but the interactions gradually diverge – both by the students (providing different amounts of information) and by the agent (requesting game elements and proposing different amounts of variables). The resulting models are consistent with the GLBoard standard but differ in the quantity and variables proposed. As for RQ2, the student's perception of "GLA Specialist" was positive: 100% agreed that it is a good option for modeling data, 56% considered the interaction easy, 78% thought it was not difficult to use, 67% indicated that the amount of data proposed was sufficient, and 78% reported that the justification for the variables was reasonable. Regarding the comparison with the manual technique, 89% believe that the agent was more efficient than the manual technique and felt more confident using it, 78% found the manual technique more difficult, and 89% preferred to use the agent. One evaluation stands out in this aspect, which considers the manual technique more appropriate due to the autonomy it provides the student. This point corroborates one of the agent's proposals: not to take the leading role of the learning designer but rather to guide and direct them to perform data modeling more efficiently, bringing the connotation of error/doubt as a positive point to improve the process.

The research limitations include (i) access to interactions of participants P3 and P4 with the agent, whose links have expired, and (ii) the restriction on the use of ChatGPT, which influenced the execution of the study. Future research aims to minimize these limitations. As contributions, this study presents research contemplating serious games, Learning Analytics, and LLM-based models – an intersection still scarce in the literature – and a good alternative to help model data for educational games. However, we identified reservations of the agent and limitations of closed LLMs. Given this, future work includes (i) building the agent in an open-source LLM, enabling customization by fine-tuning; (ii) instructing the model to avoid requesting information all at once so as not to overload the user; (iii) emphasizing the guidance/support process so that the agent does not do all the modeling on its own, encouraging the learning designer to take ownership of their educational game; (iv) compare the expert versions with closed LLM versus open LLM; and (v) apply the GLA structures in educational games to verify whether they help in identifying evidence of players' learning.

6. Acknowledgment

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

We carried out this study with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (AUXPE-CAPES-PROEX) – Finance Code 001. Additionally, this work was partially funded by the Amazonas State Research Support Foundation – FAPEAM – through the PDPG-CAPES project. It also received support from the National Council for Scientific and Technological Development – CNPq (Process 303443/2023-5).

The authors express their gratitude to the State University of Amazonas (UEA) for the institutional support and to PROPESP-UEA for the financial support. They are also grateful to their colleagues from the ThinkTEd Lab for the contributions and discussions that enriched this work.

References

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. (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.

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.

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.

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. T. (2025). Challenges in educational game data modeling from the perspective of computing students: an empirical study. In *Workshop sobre Educação em Computação (WEI)*. SBC.

IBM (2023). Ai hallucinations. https://www.ibm.com/topics/ai-hallucinations. Accessed: 2025-05-21.

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.

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

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.

Merikko, J. and Silvola, A. (2024). An ai agent facilitating student help-seeking: Producing data on student support needs. In *LAK 2024 Workshops: co-located with 14th International Conference on Learning Analytics and Knowledge (LAK 2024)*, pages 185–194. CEUR-WS. org.

Misiejuk, K., López-Pernas, S., Kaliisa, R., and Saqr, M. (2025). Mapping the landscape of generative artificial intelligence in learning analytics: A systematic literature review. *Journal of Learning Analytics*, pages 1–20.

Mitsea, E., Drigas, A., and Skianis, C. (2025). A systematic review of serious games in the era of artificial intelligence, immersive technologies, the metaverse, and neurotechnologies: Transformation through meta-skills training. *Electronics*, 14(4):649.

Mostachetti, I., Vitali, A., Regazzoni, D., Rizzi, C., and Salvi, G. P. (2025). Llm-driven adjustments in serious games: A feasibility analysis. *Studies in health technology and informatics*, 324:164–169.

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.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., and Mian, A. (2023). A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*.

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.

Runeson, P., Host, M., Rainer, A., and Regnell, B. (2012). *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons.

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.

Silva, D., Melo, R., Pires, F., and Pessoa, M. (2021). Avaliação 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).

Wang, T., Zhan, Y., Lian, J., Hu, Z., Yuan, N. J., Zhang, Q., Xie, X., and Xiong, H. (2025). Llm-powered multi-agent framework for goal-oriented learning in intelligent tutoring system. *arXiv preprint arXiv:2501.15749*.

Wohlin, C. (2021). Case study research in software engineering—it is a case, and it is a study, but is it a case study? *Information and Software Technology*, 133:106514.