SCIENTIFIC Publishing Limited

Digital Technologies Research and Applications
http://ojs.ukscip.com/index.php/dtra

Article

Game Learning Analytics for Serious Game Scoring

Mihal Miu [1,*], Niebert Blair [2] and Alexander Ferworn [1]

[1] Computational Public Safety Lab, Department of Computer Science, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada

[2] Independent Researcher, Georgetown, Region 4, Guyana

[*] Correspondence: mihal.miu@torontomu.ca

Received: 6 October 2025; **Revised:** 1 November 2025; **Accepted:** 5 November 2025; **Published:** 27 November 2025

Abstract: Game Learning Analytics (GLA) is the collection, analysis and visualization of player interactions within serious games. This interaction provides valuable insights into learning outcomes at the expense of a massive amount of data collected and complex analysis procedures. Instead of collecting detailed player interactions within serious games, we propose to capture only the outcomes of each serious game. That is, was the game goal achieved, what was the cost of achieving the game goal and what knowledge or resources did the player need for achieving the game goal? The focus of our GLA is on evaluative indicators (time and effort in achieving the game goal) collected over multiple game sessions and multiple players. Such a collection of game outcomes would enable us to observe performance improvements of players who solve the same game in order to optimize performance and also to determine learning outcomes of single players, learning outcomes of a cohort of players and learning outcomes of how a player relates to the other players in a cohort. From serious games we want to find out who has specific knowledge and skills in order to use learning patterns. First, to improve operational performance. Second, to identify individuals who may successfully handle any emergency situations based on skills obtained from playing serious games. Third, to identify useful serious games with respect to training.

Keywords: Game Learning Analytics; Serious Games; Interactive Dashboards; Scoring Metrics; Virtual Reality; Cognitive Acuity

1. Introduction

The primary purpose of Serious Games (SGs) is educational [1,2]. SGs have clear applications in emergency management, transportation, education, health care and defense. One of the authors (N.B.) has also applied serious games in areas of engineering design of remote power systems [3]. SGs have become a standard educational tool to train individuals who work in hazardous environments. It is important to understand the learning and skills obtained by players from SGs. This may be facilitated through game learning analytics (GLA). GLA is the collection, analysis and visualization of player interactions within SGs [4].

At ACDSA 2025 (International Conference on Artificial Intelligence, Computer, Data Sciences and Applications), we introduced Serious Game Scoring (SGS), an objective evaluation that determines the performance of a player given the methodology/process used to achieve the game goal. SGS has multiple objectives [3]. The first objective is to predict players' performance based on experience obtained from playing serious games. A second objective is to evaluate players' performance using meaningful metrics. A third objective is to evaluate players' methodology to complete the game. The purpose of SGS is not just to obtain high performance within a game but to relate a player's performance to real-world tasks that require skills improved or obtained from playing relevant serious games [5].

An implementation of GLA for SGS will allow analysis with regard to the effectiveness of game design, player learning and usefulness of the specific scoring mechanism. We will implement an analytics dashboard that displays reports of serious games outcomes in a tabular format and a visual format in the form of graphs. Such a dashboard would allow for visualization of learning outcomes, allowing us to identify players with specific knowledge and skills. We baptized this dashboard as “Spectrum”. Spectrum is meant to be user-friendly with a low barrier to entry and be used by data scientists, players and examiners.

GLA may be obtained by using Experience API for Serious Games (xAPI-SG) [4,6,7]. xAPI-SG has been developed to collect general information from SGs. A consequence is that a single game may first produce large quantities of data collected (verbs, activity types and extensions) and second, the data collected may be hard to analyze since it may or may not be quantitative. An alternative would be to collect the outcome of a game represented by SGS. When using SGS for game analysis, it would enable us to understand players’ learning by using SGS’s multiple dimensions. That is, we could understand player performance, process used, experience obtained, cost of obtaining the game goal and other dimensions important to the game [5].

By collecting SGS of multiple players in a single repository, it would enable us to determine learning outcomes of single players, learning outcomes of a cohort of players and learning outcomes of how a player relates to a cohort of players. Having access to a collection of SGS for specific games will enable us to determine the knowledge and skills that players possess and also lack. If we lack the required knowledge and skills, would it be possible to obtain them by playing new SGs while using Spectrum to identify player progress [8]?

Baker and Siemens (2014) divide learning analytics into five main categories: prediction, structure discovery, relationship mining, discovery with models and visualization. Their paper is one of the classic papers that influence research and development in game learning analytics [9]. Their identification of the five main categories of learning analytics may help developers implement effective GLA. GLA developers may ask themselves if their implementation fits within any of the five categories.

Scheneider and Lemons (2020) present a Systematic Literature Review (SLR) on the use of learning analytics interactive dashboards in serious games with regard to viewing educational data. They examine the evolution of interactive dashboards used in investigative studies and the effectiveness of algorithms and data visualizations. An attempt is made to understand learning processes and relationships between instructors, players, game developers and researchers [10].

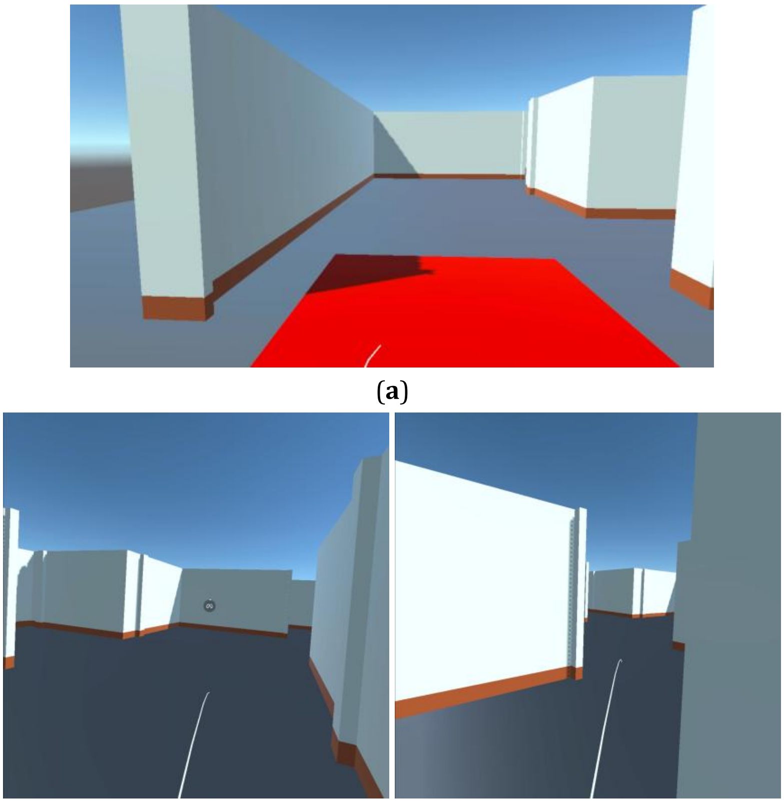
Alonso-Fernandez et al. (2021) introduce an evidence-based process used to improve the assessment of players by using their interaction data. xAPI-SG provides a systematic collection of player interactions within serious games. The collection of player interactions is used to build prediction models that will assess new players’ capabilities based on their interactions within a game [11]. There are several drawbacks. First, predictor models require large amounts of player interaction data for a game. Second, games must be validated. Player interaction needs to be standardized. Player interaction is also game-specific. A consequence of using game-specific interaction leads to the third drawback being that the predictor model is game-based.

Alonso-Fernandez et al. (2022) continued their work on GLA in order to provide a generic framework of evidence-based assessment for serious games. It is argued that capturing player interaction using a standard data format is essential for GLA. A standard data format of GLA allows for the creation of visualization tools where players’ interaction can be explored and assessed to be desirable and contribute to learning. The evidence-based assessment approach will determine if the player’s actions lead to learning outcomes [12].

Ismayilzada et al. (2025) provide a literature review of GLA in VR Environments. Their conclusions prevent generalization due to the small sample size of studies used in their literature review. Also using GLA in VR appears to be a relatively new field. VR is used mostly for entertainment; however, serious games are starting to be available for VR. The use of GLA along with the use of interdisciplinary approaches and diverse analytical tools is encouraged. Nevertheless, the use of GLA in enhancing serious games for VR is encouraged [13].

2. Materials and Methods

Our implementation of GLA using SGS will be based on evaluating a VR maze game that we refer to as “simple maze game” exhibited in **Figure 1**. The simple maze game has a single path with no cycles. The player needs to find a path through the maze only by seeing the walls in front of them. In other words, the player will never have a bird's-eye view of the entire maze, but only a field of view of the current location within the maze. It is based on this field of view that the player must move forward to the next location and determine if this decision gets the player closer to the goal or backtracks and tries again. The player may play the game multiple times with the hope of finding a path and then replaying in order to optimize game-play performance [14,15]. Our simple maze game is designed to evaluate performance based on the theory of cognitive acuity. Our simple maze game tests memory, reasoning, focus and concentration and also self-awareness [16]. Memory is tested since the player may solve the game by playing it multiple times and memorizing the path. Reasoning is used in order to determine if a partial path gets the player closer or further away from the solution. A player needs to focus on obtaining a solution. This solution may not be optimal. A player needs to concentrate and play the game multiple times in order to optimize performance. Self-awareness is tested by detecting if player bumps into walls and does not apply excessive force to game controllers or violent head movements.



(**a**)

(**b**) (**c**)

Figure 1. Simple maze game: (**a**) Start maze resolution by teleporting onto the red square; (**b**) Navigate through teleportation with the use of the VR controller; (**c**) Navigate is difficult since spaces are tighter. Expert players need to avoid collisions with walls.

The simple maze game exhibited in **Figure 1**, has commonality with emergency egress. Emergency egress refers to a safe and unobstructed exit route from a building during an emergency such as a fire [17]. Ubieto-Artur, Asion-Suner and Garcia-Hernandez promote domestic fire safety through fire drills in virtual environments where safe behavior is tested. Finding your way through a building is similar to solving a maze. We find these virtual fire drills very similar with regard to skills gained from our maze. First, a player must find a route through the maze that is equivalent to exiting a building, and second, the player must not collide with any of the walls in the maze, which is equivalent to acting safely during the virtual fire drill.

Analytics capabilities are designed into the game from the very beginning [18]. When the player solves the maze, the score (or SGS) is calculated and automatically uploaded to a server. As the player plays the game multiple times, it is hoped that learning occurs and the player's score becomes optimal. From all scores belonging to a player, a learning model may be obtained by analyzing the KEPT score and CEKPT score (optimal score). KEPT and CEKPT scores are SGS concepts that we presented in our paper at ACDSA 2025. The KEPT score is more complex when examined in detail. First, the KEPT score may be divided into two categories: i-KEPT or initial KEPT score and b-KEPT or baseline KEPT score. This is necessary due to the following fact. When you play a game for the first time, you achieve the game goal, but you may violate some of the game rules. This is your i-KEPT score. The b-KEPT score is the first score that you achieve without violating the game rules. This means that learning occurred (you learned how to play the game without breaking the rules) between the i-KEPT score and the b-KEPT score. Further learning occurs between the b-KEPT score and CEKPT score. This learning represents an optimization that occurred with regard to the game solution.

Our implementation of SGS for the simple maze game will have four dimensions: cost, effort, safety and experience. The metrics used are intrinsic to the outcome of the game. Details are described in **Table 1**.

Table 1. SGS dimensions implemented in a simple maze game.

SGS Dimension	Measurable Dimension	Unit
Cost	Time	Seconds
Effort	Teleportations	An integer representing the total teleportation count
Safety	Collisions	An integer representing total collisions with walls
Experience	A ratio composed of historical SGS dimensions	Ratio of the total number of maze resolutions over the total cost of maze resolutions

Since we are not interested in game details but the final outcome of a game, our LAM should not be overwhelming nor hard to analyze. Our analytics results are based on KEPT which is well defined (knowledge, experience, processes and tools used in solving a game). This also results in analytics results that are easy to understand [19]. For a single player playing a game, we can display game outcomes in a tabular format sorted by time. A simple examination will allow the determination of the b-KEPT (baseline) score and CEKPT (optimal) score.

Our infrastructure is designed to be secure and available 24/7 to gamers, examiners and data scientists. Spectrum is designed as a service and built using Python and Flask. It has a secure user interface where players, examiners and data scientists may log into, search for game results and generate reports [20]. In other words, Spectrum provides a dashboard web application that provides results in a tabular format or visually through various graphs/charts. You can find details with regards to accessing Spectrum in **Appendix A**.

Spectrum may be used by any game (in our case, the simple maze game) over the internet using a RESTful API. The maze game runs on a Meta Quest 2/3 headset. The maze game connects to the internet and may consume Spectrum's RESTful API to upload game results. Once game results are uploaded to Spectrum, they are automatically inserted into a relational database. The database server is secure and is not accessible to the outside world. Only Spectrum has access to the database server. In turn, the application server that runs the Spectrum is protected from the internet by a firewall. This is exhibited in **Figure 2**.

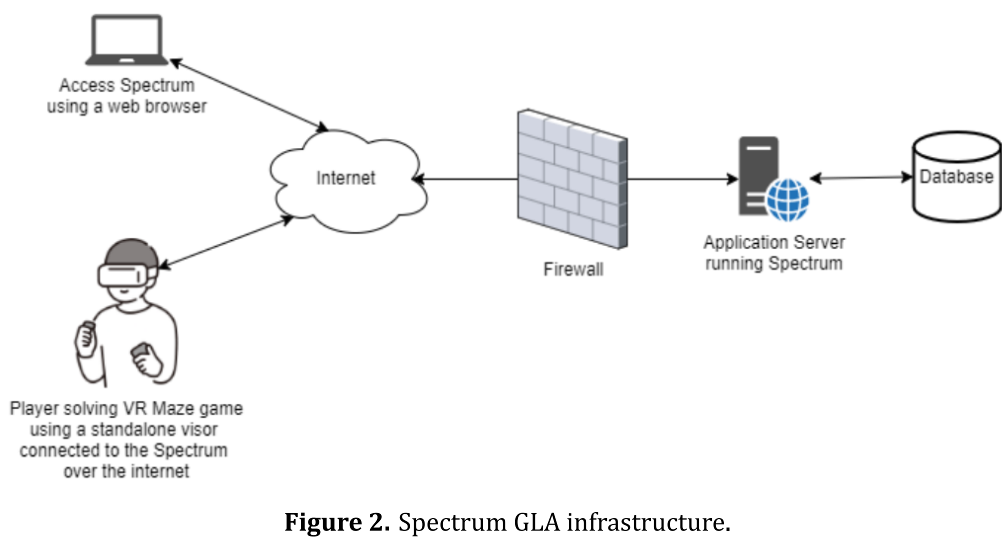


Figure 2. Spectrum GLA infrastructure.

Game scores are stored in a relational database (in our case, PostgreSQL). Spectrum uses SQLAlchemy to retrieve data from the database and operates upon it by storing it in memory using instances of dataclasses. Dataclasses reduce boilerplate code and are useful for modeling data. Instances of dataclasses may be used by Pandas through DataFrames. The LAM is created by our analytics engine built with Pandas and SciPy. Visualization of analytics results is provided by tables and Matplotlib. Tables help understand game outcome details, while visualization gives an overall representation of results. Matplotlib visualization makes it easy to analyze and understand data trends and relationships. SciPy is used to support regression analysis and its result is visualized in a Matplotlib chart. This is exhibited in **Figure 3**.

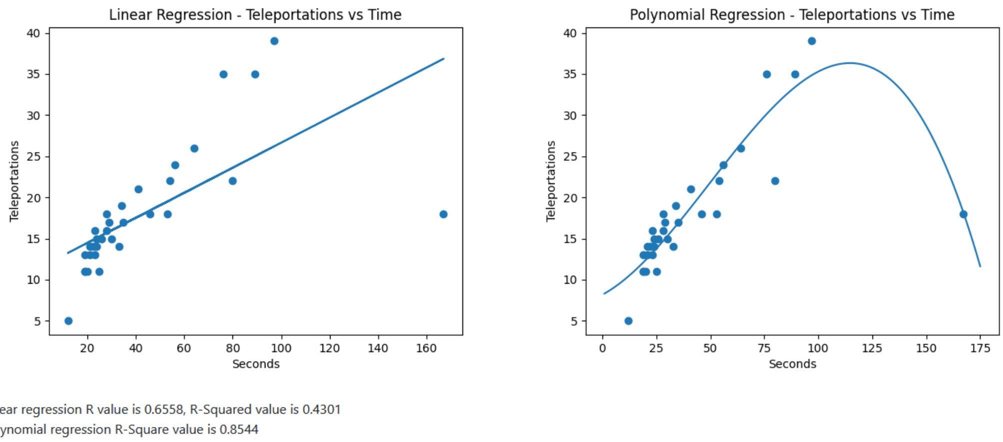


Figure 3. Regression plots for all players for the simple maze game.

3. Results

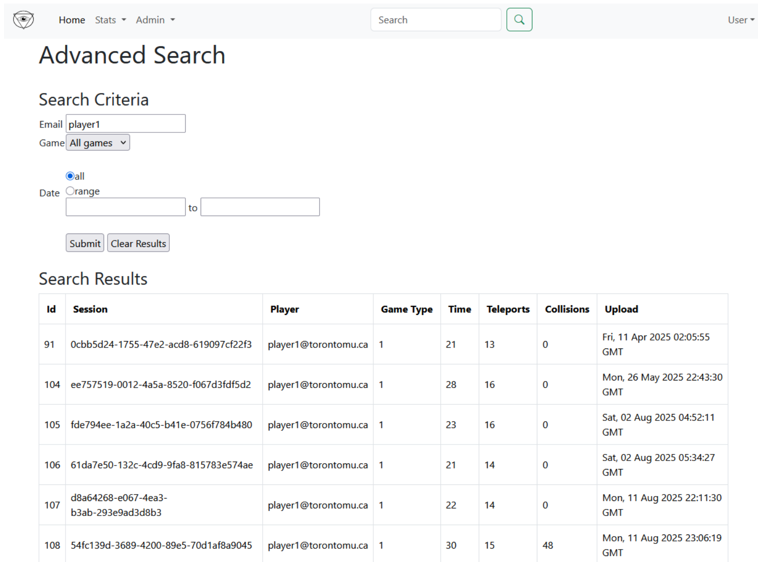
Table 2 presents a summary of key findings with regard to implementing GLA through the use of SGS. These key findings will be elaborated in this section and the discussion section.

Table 2. Summary of key findings with regard to GLA and SGS.

Finding	Summary Explanation
Player history	A player's history may be viewed, exhibited in **Figure 4**.
Game History	Players who successfully completed the game may be observed.
Cost of learning per player	A player's maze resolution capabilities may be viewed on multiple metrics: cost, effort, and safety experience.
Usage of serious games per player	Can observe what games were played by which players.
Analysis of a single player or a cohort of players	Important milestones such as i-KEPT, b-KEPT and CEKPT scores may be viewed without looking at the overall history (see **Figure 5**).
Validate serious games	GLA can validate if a game facilitates learning.
Validate scoring mechanism	SGS implementation may be validated to determine if they are appropriate for evaluating serious games and player learning.

Spectrum provides a player's history of game sessions. This history may be used in two ways. First, we can identify players that learn with a low cost, players that learn with a reasonable cost and players that learn at higher costs. Second, we can use a player's history to evaluate the usefulness of a serious game. This can be achieved by identifying expert players in one type of serious game. We can use these expert players to test the new serious game. If these players exhibit learning on the new serious game, then the game is valid. Furthermore, the rate of learning or cost of learning can be examined in order to identify deficiencies and recommend improvements for the new serious game.

Analytics in Spectrum go beyond the simple reporting of scores and completion times. By capturing KEPT and CEKPT metrics across multiple play sessions, Spectrum allows us to identify how learning occurs in stages and whether improvements are the result of genuine cognitive gains or repeated practice. For example, if a player demonstrates rapid improvement between i-KEPT and b-KEPT but plateaus when attempting to reach CEKPT, this may indicate that basic rules were quickly understood but optimization strategies were not fully developed. Such patterns become valuable in differentiating between procedural knowledge acquisition and higher-order skill mastery.



Home Stats Admin Search User

Advanced Search

Search Criteria

Email player1

Game All games

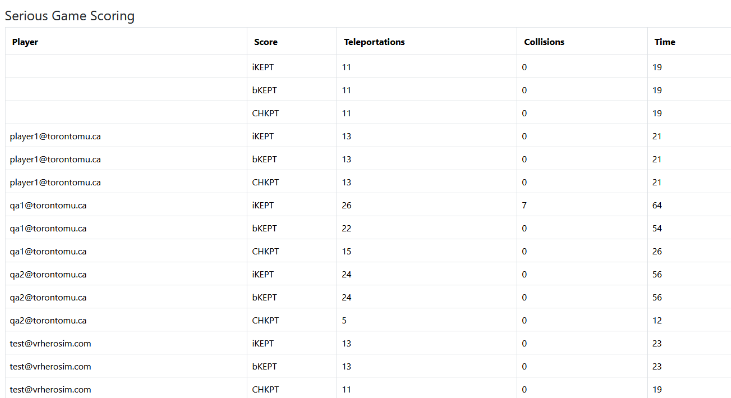
Date all range to

Submit Clear Results

Search Results

Id	Session	Player	Game Type	Time	Teleports	Collisions	Upload
91	0cbb5d24-1755-47e2-acd8-619097cf22f3	player1@torontomu.ca	1	21	13	0	Fri, 11 Apr 2025 02:05:55 GMT
104	ee757519-0012-4a5a-8520-f067d3fdf5d2	player1@torontomu.ca	1	28	16	0	Mon, 26 May 2025 22:43:30 GMT
105	fde794ee-1a2a-40c5-b41e-0756f784b480	player1@torontomu.ca	1	23	16	0	Sat, 02 Aug 2025 04:52:11 GMT
106	61da7e50-132c-4cd9-9fa8-815783e574ae	player1@torontomu.ca	1	21	14	0	Sat, 02 Aug 2025 05:34:27 GMT
107	d8a64268-e067-4ea3-b3ab-293e9ad3d8b3	player1@torontomu.ca	1	22	14	0	Mon, 11 Aug 2025 22:11:30 GMT
108	54fc139d-3689-4200-89e5-70d1af8a9045	player1@torontomu.ca	1	30	15	48	Mon, 11 Aug 2025 23:06:19 GMT

Figure 4. Search results for the simple maze game of a single player, 'player1'.



Serious Game Scoring

Player	Score	Teleportations	Collisions	Time
	iKEPT	11	0	19
	bKEPT	11	0	19
	CHKPT	11	0	19
player1@torontomu.ca	iKEPT	13	0	21
player1@torontomu.ca	bKEPT	13	0	21
player1@torontomu.ca	CHKPT	13	0	21
qa1@torontomu.ca	iKEPT	26	7	64
qa1@torontomu.ca	bKEPT	22	0	54
qa1@torontomu.ca	CHKPT	15	0	26
qa2@torontomu.ca	iKEPT	24	0	56
qa2@torontomu.ca	bKEPT	24	0	56
qa2@torontomu.ca	CHKPT	5	0	12
test@vrherosim.com	iKEPT	13	0	23
test@vrherosim.com	bKEPT	13	0	23
test@vrherosim.com	CHKPT	11	0	19

Figure 5. SGS results for the simple maze game of all players in a tabular format.

Another critical feature of analytics in Spectrum is its ability to classify learning costs. Players who achieve baseline and optimal performance with fewer attempts are categorized as low-cost learners, whereas those requiring many repetitions are high-cost learners. This classification provides a nuanced understanding of efficiency in learning processes. Low-cost learners may be ideal candidates for advanced training, while high-cost learners can be studied to identify barriers that impede progress. By analyzing collision frequency, maze resolution time, and score evolution simultaneously, Spectrum presents a comprehensive picture of performance rather than isolated data points.

Furthermore, analytics can be aggregated at the cohort level. By comparing learning curves across groups, educators can validate whether a serious game itself promotes consistent knowledge transfer. If a majority of players display a similar trajectory of improvement, the game can be considered effective in teaching its intended outcomes. On the other hand, if performance varies widely, design elements may need refinement to ensure inclusivity and fairness. Spectrum thus transforms analytics into a diagnostic tool: it not only assesses individuals but also highlights the effectiveness of game mechanics themselves. This dual capacity is delivered in evaluating both learner and learning environment which reinforces the role of analytics as a cornerstone of Serious Game Scoring.

GLA should be useful. For example, our regression plots represent an oversimplification of our game where an independent variable is analyzed to a dependent variable. A game may be influenced by multiple independent variables. This is where analysis may start and continue with SGS visualizations. Another issue is that the game trials should express confidence. At the bottom of **Figure 6,** we have the experience plot where we total up all game metrics such as teleportations, maze resolution time, collisions and games played. What we want is a high game session total (Games Played) with a low collisions (Total Collisions). This may be used for solving mazes where we optimize resolution time (lower time is better) or teleportations used (fewer teleportations are better) in maze resolution. **Appendix B** offers a brief discussion on useful and useless metrics with regards to GLA.

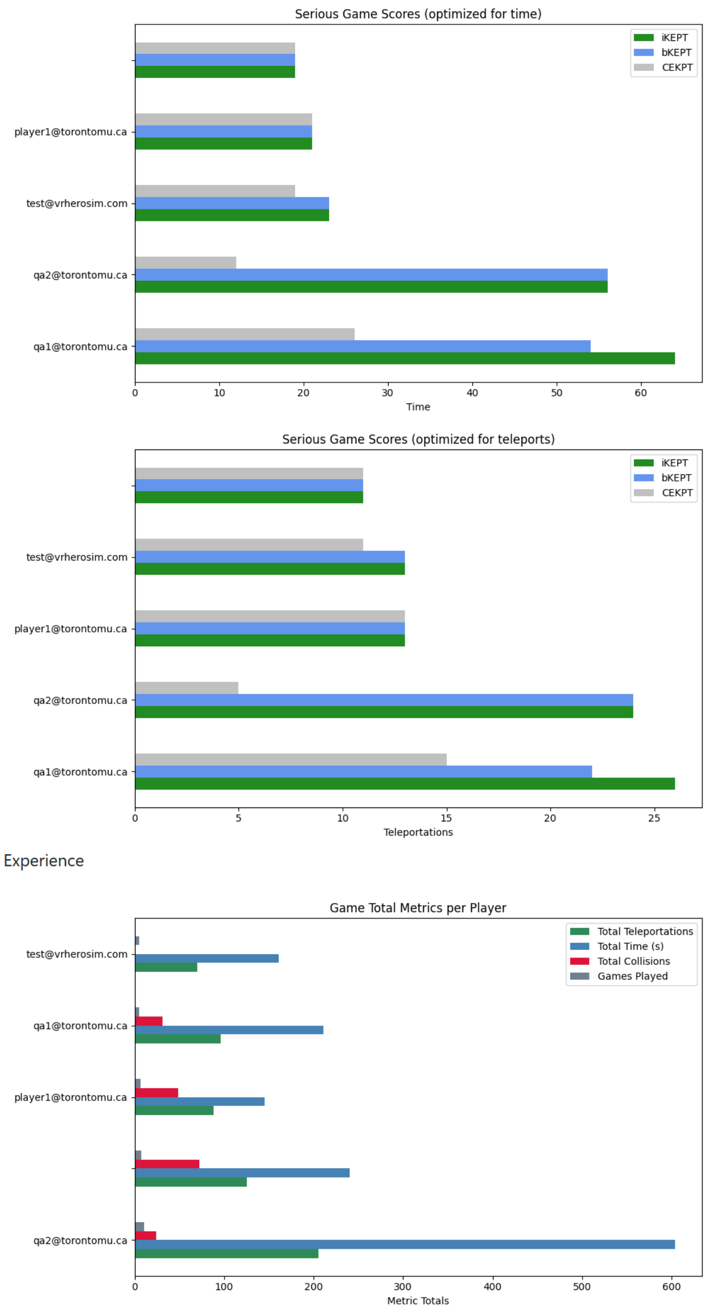


Figure 6. Serious Game Scores for all players for the simple maze game.

4. Discussion

When examining analytics for a single player, Spectrum provides granular insight into that individual's trajectory of learning and skill acquisition. By tracing i-KEPT, b-KEPT, and CEKPT scores across multiple sessions, it becomes possible to evaluate how efficiently a player internalizes game rules, reduces errors such as collisions, and optimizes their navigation strategy. Over time, such data can reveal whether improvements are the result of memorization, strategy adaptation, or enhanced dexterity. However, the true strength of the analytics framework emerges when this single-player history is compared within a cohort. By comparing an individual's progress relative to peers, we can distinguish between personal learning styles and generalizable patterns of performance. For instance, if a player consistently lags behind a cohort despite repeated attempts, this may indicate the need for targeted feedback or adaptive difficulty adjustments. Conversely, a player who outperforms the cohort may serve as a benchmark for validating the effectiveness of new game mechanics or instructional design. Thus, analytics allows us to understand not only isolated performance but also how individuals interact and align within a broader learning community.

The effectiveness of a serious game is best measured through demonstrable improvements in learning outcomes [21]. In the case of the simple maze game, learning can be demonstrated by reductions in maze resolution time, fewer collisions, and a gap between baseline and optimal scores. Wider gaps are preferred over narrower gaps since gap magnitude may be linked to optimization magnitude. These performance markers suggest that a player has not only memorized solutions but also used cognitive skills such as reasoning, spatial awareness, and concentration. Spectrum's analytics allow educators and researchers to distinguish between superficial success (e.g., memorization without transferable skills) and deeper learning (e.g., problem-solving that can extend to new contexts). For training domains such as emergency response or healthcare, this distinction is critical. If serious games are to be more than digital drills, their design must lead to skills that persist beyond the game itself. Therefore, SGS provides a framework for verifying if serious games foster real knowledge acquisition, adaptability, and resilience. These are qualities essential for professional application in high-stakes environments.

Serious Game Scoring (SGS) contributes significantly to validating not just the player but also the game itself [21]. A well-designed SGS framework does more than track points; it operationalizes performance into measurable dimensions such as cost of learning, efficiency of methodology, and adaptability under constraints. This multidimensionality ensures that SGS captures the nuances of both learning and gameplay quality. For example, if Spectrum consistently shows that even expert players plateau early in a game, it may indicate design flaws that limit learning progression. On the other hand, a smooth gradient of improvement across i-KEPT, b-KEPT, and CEKPT suggests that the game scaffolds learning effectively. Furthermore, SGS enables cross-comparison across different games, allowing developers to assess which game structures yield more meaningful learning. Ultimately, SGS provides a dual validation mechanism: it certifies the competence of a player while simultaneously ensuring the instructional integrity of the serious game. In this way, SGS evolves from a scoring tool into an evaluation standard for serious game design.

Spectrum and SGS are limited in scope by design. This is due to the availability of resources (financing and time, and development skills). Spectrum and SGS facilitate evaluations of a single-player game, specifically the VR maze game (exhibited in **Figure 1**). Furthermore, the maze game is simplified through its design, a maze with a single path and no cycles. A consequence of game simplification is that SGS may evaluate player performance using a smaller number of measurable dimensions.

Our argument for Spectrum (our GLA implementation) is performed on procedure and not official game trials. Spectrum's implementation has been aided by best practices, especially by Baker and Siemens. As stated in the introduction section, Baker and Siemens divided learning analytics into five main categories. Spectrum implements four of these categories: structure discovery, discovery with models, relationship mining (it is a proof of concept as of the time of writing this article), and visualization. Spectrum does not compete but complements xAPI-SG. Spectrum is an analytics platform for SGS. Spectrum does not capture player interactivity (when a player makes a turn or when a player backtracks) while playing the game. This may be performed by xAPI-SG.

Over the next three months, we will begin our official game trials. The official game trial will involve solving three mazes. The simplest maze has been referred to as a VR maze game exhibited in **Figure 1**. There will be a maze of medium difficulty that has a single path with no cycles. The medium difficulty maze will be twice as large as the simplest maze. The medium difficulty maze may be solved by memory alone; however, this maze will begin to stress test your cognitive acuity. There will be a difficult maze that has a path, but there will be cycles. This maze test will continue to test your cognitive acuity further than the medium difficulty maze. Memory alone will not be enough to solve this maze. To solve this maze, ideally, an algorithm will need to be implemented by the player. Each player will have a maximum of 15 attempts to solve each maze. If a player solves all three mazes, then their cognitive acuity is good. Learning cost can be obtained by comparing the number of successful maze resolutions to the total game time.

Spectrum brings us a step closer to creating smart serious games. Suppose an agent could examine a player's score, identify deficiencies and suggest in what areas the player may improve. Suggestions may be procedures or strategies to be used for game play. Then such an agent would turn SGS and Spectrum into a smart serious game.

Currently, maze games can be scored using SGS using Spectrum. It would be desirable to score serious games with practical scenarios representing emergency management, transportation, education, health care and defense. In the near future we will ask why a player has a better score than other players. Based on our current infrastructure, we do not have the data to answer this question. If we were to use xAPI-SG, then we would have a player's game details. These details may be compared with other players in order to determine the difference. It is this difference that may explain why a player has a better score than other players.

5. Conclusions

Spectrum allows us to see the results of a player's game session. This history may be summarized by simply providing a player's i-KEPT, b-KEPT and CEKPT score. Based on this score we can categorize a player as a beginner, proficient or expert for a specific serious game [22]. Spectrum has also been effective in refining our scoring strategy (SGS). After ACDSA 2025, we examined M.M. (denoted by player1 in **Figures 4–6**) game game-playing history using Spectrum. We noticed that KEPT represented two components: an initial score referred to as i-KEPT and a baseline score referred to as b-KEPT.

At the same time, Spectrum may be used to validate the effectiveness of the game. If a player's performance does not improve after multiple game sessions, is the player to blame or is the game problematic? The authors of this paper used Spectrum as a proof of concept and also to validate the simple maze game. Note, for the upcoming official game trials, there will be other maze games of greater difficulty. Spectrum will be used to validate these games using the procedure that we used in the simple VR maze game.

Spectrum will be invaluable when our official game trials begin later this year. It will allow us to analyze individual players, a cohort of players and the relation of an individual player's performance within their cohort. Since analysis is automated, Spectrum may be used at any step of our official game trials in order to validate any procedures.

Author Contributions

Conceptualization, M.M.; methodology, M.M. and N.B.; software, M.M.; validation, M.M. and N.B.; formal analysis, M.M. and N.B.; investigation, M.M.; resources, M.M.; writing—original draft preparation, M.M. and N.B.; writing—review and editing, M.M. and N.B.; visualization, M.M. and N.B.; supervision, A.F.; project administration, M.M.; All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

Spectrum is a hosted service and may be accessed over the internet using a web browser. Demos are available by contacting the corresponding author (M.M.). If you are interested in solving the simple maze game, you will have to obtain it from the corresponding author (M.M.) and side-load it on your Meta Quest 2/3 device. You can solve the maze anonymously; no personal identifying information is loaded to the Spectrum service. All anonymous uses are treated as a single cohort. Alternatively, you can enter your email before the start of each maze resolution (see **Figure A1**). In this case, Spectrum will use your email to calculate an SGS score for you and you alone.

Figure A1. Player is entering the email before starting maze resolution.

Appendix B

GLA should represent useful metrics. Statistical analysis, such as mean, median, minimum and maximum values, is a common procedure with respect to analysis. However, such metrics may not make sense in our situation. If we compare the averages from **Figure A2,** we see that these results do not relate to any SGS results in **Figure 6.** This demonstrates that the GLA procedure should produce useful metrics and not metrics for the sake of metrics. **Figure A2** exhibits metrics for the sake of metrics and should not be used in GLA.

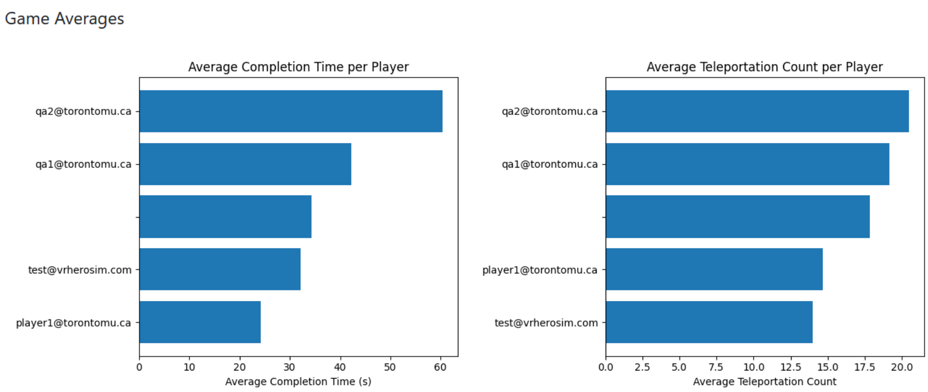


Figure A2. Game averages for all players for simple maze game.

References

1. Gürer, S.; Surer, E.; Erkayaoğlu, M. Mining-Virtual: A comprehensive virtual reality-based serious game for occupational health and safety training in underground mines. *Saf. Sci.* **2023**, *166*, 106226.
2. Martínez-Durá, R.J.; Arevalillo-Herráez, M.; García-Fernández, I.; et al. Serious games for health and safety training. In *Serious Games and Edutainment Applications*; Springer: London, UK, 2011; pp. 107–124.
3. Blair, N.; Pons, D.J.; Krumdieck, S.P. Electrification in Remote Communities: Assessing the Value of Electricity Using a Community Action Research Approach in Kabakaburi, Guyana. *Sustainability* **2019**, *11*, 2566.
4. Alonso-Fernández, C.; Calvo-Morata, A.; Freire, M.; et al. Data science meets standardized game learning analytics. In Proceedings of the IEEE Global Engineering Education Conference, Vienna, Austria, 21–23 April 2021; pp. 1546–1552.
5. Miu, M.; Ferworn, A. Serious Game Scoring for Maze Resolution. In Proceedings of the International Conference on Artificial Intelligence, Computer, Data Sciences and Applications, Antalya, Turkiye, 7–9 August 2025; pp. 1–5.
6. Serrano-Laguna, Á.; Manero, B.; Freire, M. A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes. *Multimed. Tools Appl.* **2018**, *77*, 2849–2871.
7. Pérez-Colado, V.M.; Pérez-Colado, I.J.; Martínez-Ortiz, I.; et al. Democratizing Game Learning Analytics for Serious Games. In *Games and Learning Alliance*; de Rosa, F., Marfisi Schottman, I., Baalsrud Hauge, J., et al., Eds.; Springer: Cham, Switzerland, 2021; 13134, pp. 164–173.
8. Volkmar, G.; Pfau, J.; Teise, R.; et al. Player Types and Achievements – Using Adaptive Game Design to Foster Intrinsic Motivation. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 747–754.
9. Baker, R.; Siemens, G. Educational Data Mining and Learning Analytics. In *The Cambridge Handbook of the Learning Sciences*; Sawyer, R.K., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 253–272.
10. Scheneider, T.; Lemos, R. The Use of Learning Analytics Interactive Dashboards in Serious Games: A Review of the Literature. *Int. J. Innov. Educ. Res.* **2020**, *8*, 150–174.
11. Alonso-Fernández, C.; Freire, M.; Martínez-Ortiz, I.; et al. Improving evidence-based assessment of players using serious games. *Telematics Inf.* **2021**, *60*, 101583.
12. Alonso-Fernández, C.; Calvo-Morata, A.; Freire, M.; et al. Game Learning Analytics: Blending Visual and Data Mining Techniques to Improve Serious Games and to Better Understand Player Learning. *J. Learn. Anal.* **2022**, *9*, 32–49.
13. Ismayilzada, A.; Karimov, A.; Saarela, M. Serious Games Analytics in VR Environments: A Two-Stage Systematic Literature Review. *J. Interact. Learn. Res.* **2025**, *36*, 57–69.
14. Tlili, A.; Chang, M.; Moon, J.; et al. A Systematic Literature Review of Empirical Studies on Learning Analytics in Educational Games. *Int. J. Interact. Multimedia Artif. Intell.* **2021**, *7*, 250–261.
15. Crowe, D.; Averbeck, B.; Chafee, M.; et al. Mental Maze Solving. *J. Cogn. Neurosci.* **2000**, *12*, 813–827.
16. Leeds, J. P. The Theory of Cognitive Acuity: Extending Psychophysics to the Measurement of Situational Judgment. *J. Neurosci. Psychol. Econ.* **2012**, *5*, 166–181.
17. Ubieto-Artur, P.; Asión-Suñer, L.; García-Hernández, C. Promoting Domestic Fire-Safety: Virtual Drills as a Training Tool for Citizens. *Fire* **2025**, *8*, 286.
18. Emerson, A.; Cloude, E.B.; Azevedo, R.; et al. Multimodal learning analytics for game-based learning. *Br. J. Educ. Technol.* **2020**, *51*, 1505–1526.
19. Banihashem, S. K.; Dehghanzadeh, H.; Clark, D.; et al. Learning analytics for online game-Based learning: a systematic literature review. *Behav. Inf. Technol.* **2023**, *43*, 2689–2716.
20. Oliveira, B.; Oliveira, Ó.; Belo, O. Overcoming traditional ETL systems architectural problems using a service-oriented approach. *Expert Syst.* **2023**, *40*, e13442.
21. Daoudi, I. Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. *Educ. Inf. Technol.* **2022**, *27*, 11237–11266.
22. Alonso-Fernández, C.; Martínez-Ortiz, I.; Caballero, R.; et al. Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *J. Comput. Assist. Learn.* **2020**, *36*, 350–358.