Ruipérez-Valiente, J. A. Unveiling the Potential of Learning Analytics in Game-based Learning: Case Studies with a Geometry Game. In Handbook of Research on Promoting Economic and Social Development Through Serious Games. (pp. 524-544). IGI Global. ISBN 9781799897323

DOI: 10.4018/978-1-7998-9732-3.ch023

https://doi.org/10.4018/978-1-7998-9732-3.ch023

Unveiling the Potential of Learning Analytics in Game-based Learning

Case Studies with a Geometry Game

José A. Ruipérez-Valiente
University of Murcia

ABSTRACT

In recent years, due to the technological advance, we see a huge change in the way of teaching and learning. Game-based learning (GBL) is applied in order to motivate students, improve their knowledge and make the process of education more enjoyable. This chapter will focus on explaining the process of Learning Analytics (LA) in the context of GBL, which can play a meaningful role in transforming learning pathways in games into interpretable information for teachers. Respectively, the overarching aim of this work is to verify the potential of GBL and LA applied to the educational process through four case studies, each of which presents an important metric in a geometry game Shadowspect developed in order to train geometry and spatial reasoning skills by solving a series of geometry puzzles. The case studies will be focused on data-driven game design, learner modelling & adaptive learning, game-based assessment (GBA) of 21st-century skills, and teacher-oriented visualization dashboards.

Keywords: Game-Based Assessment, Learning Analytics, Computational Social Science, Elo Rating System, STEM Education, K12 Education, Data Analysis, Educational Technology

1. BACKGROUND

Technology is gradually being applied into our everyday life affecting a lot of fundamental processes, including education. Accordingly, in recent years, we see a considerable change in the way of teaching and learning. Nowadays, it is common to involve game-based learning (GBL) in order to motivate students, improve their knowledge and increase the enjoyment of such an essential process as education. However, while there are obvious benefits, current educational systems are still suffering from numerous issues that require transformative changes. GBL holds the potential to improve many of the problems that are currently present within the educational process. One issue is that even though many teachers report a positive attitude towards games being used in K12 classrooms and believe that they can improve learning and curriculum, the actual number of teachers who are implementing digital games in their curriculum is contrarily low. Several reasons affect the possibility of implementing games into the formative assessment, including the doubts on how to actually effectively implement games in the classroom, how to support evidence-based decisions based on game data, or assess students using the games.
Within the context of these challenges, we argue that learning analytics (LA) could be a vital novelty being able to address the problems as mentioned earlier. It is proved that LA can play a meaningful role in transforming learning pathways in games into interpretable information for teachers.

While the work presented in this article focuses on the process for implementing LA, there are numerous other cross-cutting factors to consider during this process that are highly important for success. In the first place, it should be noted that the rise of research in LA has been given by the introduction of technology in education and that as it continues to be introduced more in the educational fabric (US Department of Education, 2016), we could expect a greater demand for implementing it. In turn, it could lead to greater ease in certain parts of introducing this process into education. Secondly, there is a need to highlight the need to anchor LA projects in real educational applications that can improve learning. However, there is a risk of implementing technology and analytics that are totally disconnected from the best pedagogical practices and educational theories developed in recent decades. Third, nowadays, people are more concerned about their privacy, and it is especially crucial to guarantee the privacy of students and teachers for the ethical development of this technology. This is a problem that has already been shaped by numerous policies (Drachsler & Kalz, 2016). On the other hand, it is a critical question whether educational systems want to move in the same direction as the large Internet companies, which continuously monitor their users (Slade & Prinsloo, 2013).

Accordingly, we must take into account the fact that students or their representatives must consent to their data being used for these purposes and have the freedom to choose which data can be used or if they wish to withdraw that consent at any point in the process; all of this requires major orchestration among the actors in this process, not only students but also teachers, designers and system administrators (Pardo & Siemens, 2014). This can lead to several difficulties in the implementation process, including one of the most crucial concerns - time which can be highly affected due to bureaucratic issues. Finally, we will highlight the need for the proper implementation and systematization of LA in education. It is essential to achieve the involvement of educational institutions and the generation of national educational policies, which could generate problems that may not initially be obvious (Macarini et al., 2019).

Despite all the challenges that could arise while introducing games into the educational process and using LA techniques accordingly, there are very important benefits, including, first of all, a better experience for students and teachers. This includes the academic achievement, enjoyment of the process from both sides, concentration and classroom dynamics in general. Moreover, we would be able to address the individual needs of every student, which is performed in the context of adaptive learning (Liu et al., 2017). Additionally, it will be possible to understand and improve the effectiveness of teaching practices and respectively inform institutional decisions and strategies. Respectively, the overarching aim of this work is to verify the potential of GBA and LA applied to the educational process through four case studies, each of which presents an important metric. In this study, we will explore four case studies of LA applied in a geometry game Shadowspect. The game was developed in order to train geometry and spatial reasoning skills by solving a series of geometry puzzles. The case studies will be focused on data-driven game design, learner modelling & adaptive learning, game-based assessment (GBA) of 21st-century skills, and teacher-oriented visualization dashboards.

The organization of the chapter is as follows: in Section 2, we will describe the LA process in GBL. In Section 3, we will introduce a geometry puzzle game Shadowspect. Section 4 will focus on a series of metrics that can provide comprehensive information about the learning process and the students. We will perform the puzzle difficulty, the competency of students by the use of the Elo rating system, and the persistence metric. Next, we will describe the implementation of a real-time dashboard of the metrics as mentioned earlier. Finally, in Section 5, we will discuss the obtained results and draw our conclusions.

2. THE LEARNING ANALYTICS PROCESS IN GAME-BASED

In this section, we will review the LA process in GBL and studies that fall within each one of these stages. We will delve into the following steps in the process:

1. Game-based learning environments. What is the context, and who are the students?
2. Raw data collection. Which data needs to be generated, and how to store it?

3. Data manipulation and feature engineering. What features are needed and how to obtain them?
4. Analytics and modelling. Which analysis and models should be implemented?
5. Educational application. What is the objective application and the final user?

In addition, there are other issues that must be taken into account in a transversal way, such as the technologies to be applied, the theories and learning sciences to be used, the privacy of users, as well as educational institutions and policies. Finally, it is important to mention that depending on the project, it is not necessary to go through all the steps that we will describe. Most researchers and experts in the field of LA focus only on steps 3 and 4. We now delve into each of these stages.

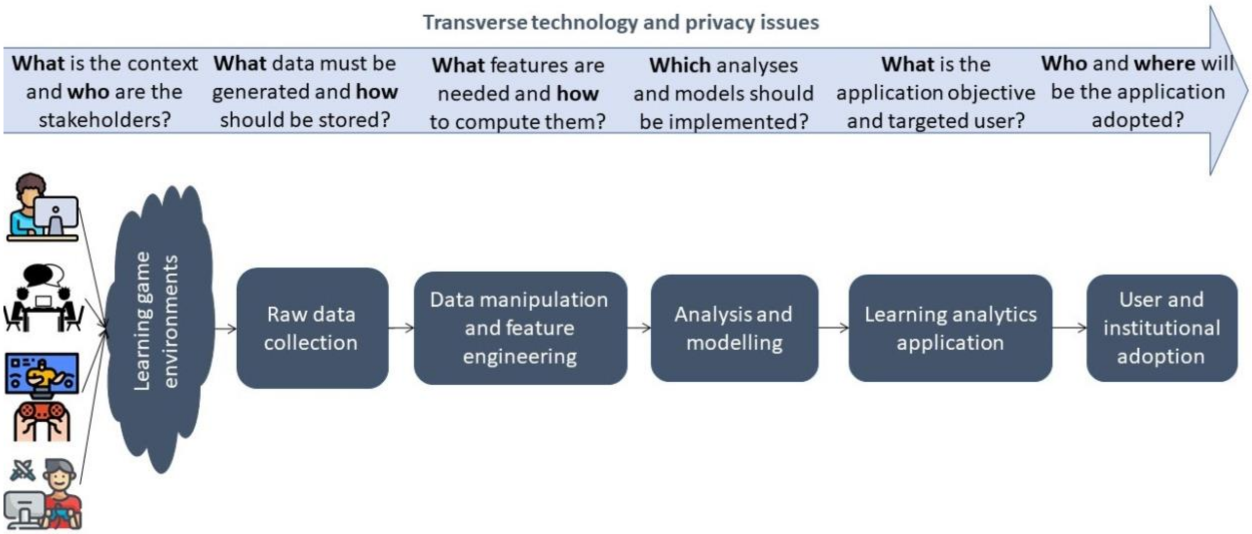

Figure 1. The Learning Analytics implementation process

In Figure 1, we illustrate the LA implementation process. The first step of the process takes place in the learning game environment and the users who interact with it. Traditionally, in distance or digital education, we can specify the following environments: Learning Management Systems (LMSs), such as Moodle, Sakai, dotLRN, among many others (Romero et al., 2008). Then, with the advent of Massive Open Online Courses (MOOCs), there was an explosion in terms of being able to collect large amounts of data from students around the world (Breslow, et al., 2013) which favoured the emergence of numerous studies in the area. LA projects have also been carried out in less common settings, such as smart tutors (Jaques et al., 2014) or educational games (Freire et al., 2020). Each of these environments has its specificities that make the implementation of their analytics hard to implement.

On the other hand, in this context, the ability to easily collect data was strengthened with the advent of education through digital environments. The most common way that most virtual learning environments have followed is to save the traces of all the clicks that users make in the environment as events, in the format that is usually known in the literature, as clickstream data. This approach is not only followed in educational settings, but it is universal in a multitude of digital domains to model human behaviour (Bollen et al., 2009). Besides, in recent years, the use of sensors in education has also become increasingly common, both to capture audiovisual signals and biometric signals from students such as the heartbeat or electrodermal activity. It is especially promising for the evaluation of complex tasks. From the Shadowspect side, it generates a vast amount of clickstream data that we save and can generate the analytics out of it.

Once the data has been collected, the manipulation process begins. For the reason that these low-level events represent isolated actions of the students, they are not very informative in the raw format, and therefore it is necessary to go through a process known as feature engineering. During this process, data is transformed into actionable and valuable educational information. For example, these virtual environments will save the actions of each student, but this information will not be handy until we algorithmically calculate the total time that the user has spent actively interacting with the learning platform. This process requires a high technical knowledge level to manipulate the data, experience in the context, so that the characteristics are known to be useful (Karthik Ramasubramanian & Singh, 2019). In addition, it is usually one of the stages that requires the most effort in data analysis projects.

Once we have the required data, a key to understanding them and obtaining an educational benefit from them lies in their analysis. We now will describe a series of types of algorithms that are the most commonly used in studies with a straightforward application. Traditionally, many of the studies around analytics seek to understand students' interactions with the learning environment retrospectively. The methods to be applied are evident and direct when projects have clear objectives and future applications. For example, for generating predictions of future learning outcomes, there will be applied supervised learning algorithms capable of modeling the future based on historical data sets, such as predicting if they will get a certificate in MOOCs (Ruipérez-Valiente et al., 2017). Item Response Theory is capable of the adaptation of content or personalized recommendations to the user, for example, based on the user's current difficulty and skill level (Chen et al., 2005). These are a few examples of the most common modelling and analysis methods that are widespread in the LA area, but others can be applied as well. Finally, it should be noted that the data manipulation and feature engineering process, in conjunction with the analysis and modelling phase, is an iterative process that can be repeated until the desired results are achieved.

The analysis of the previous phase usually have an educational application associated with them. However, the reality is that on many occasions, this application is not usually put into practice and the research usually ends in the analysis and modelling phase. This means that most of the research is not transferred to practice, and therefore, it cannot be evaluated if it really has a positive educational impact or not.

In other cases, the most typical applications seen in LA projects include visualization interfaces, which are normally used by instructors to monitor how their students are progressing but can also be accessed by students to reflect on their own learning process and applications that introduce fully adaptive modules or systems. Moreover, another application is a recommendation system that sends personalized recommendations to the student about what might interest them. Finally, it is also expected that the final application of the analysis generates a high-level report on an educational context of an institution, course or subject, to better understand the context and adapt educational policies (Reich & Ruipérez-Valiente, 2019). These applications or final analyses should generate feedback in the educational contexts where the data was generated that allows improving the learning process, and with this, the cycle of the implementation of LA would be closed. Furthermore, the effect of these changes should be evaluated, which could be done using again a LA methodology. This evaluation is essential to be able to measure the impact of the changes introduced in the educational context.

In the coming years, the LA field and educational institutions, in general, face the challenge of the prospect of implementing the above-mentioned methods on a large scale and in a systematic way to finally be able to reach the high potential that all researchers converge on. To do this, there are numerous barriers that the entire community of actors involved in education must contribute to breaking down.

Policies should be developed to allow a correct implementation with educational impact, without losing sight of the importance of ethics and security in using these data, preserving the privacy to which students are entitled. These projects must be focused on users, empowering them and putting them at the centre of development in order to implement applications that can be used sustainably over time.

For this, it is recommended that institutions in the coming years focus on implementing LA applications such as visualization interfaces or content evaluation through analytics in a proper way. In a more distant future, other challenges await the area at the research level, such as transferring machine learning models when contexts and even platforms change or facilitating the interpretability of all the analytics to be more accessible to the users who have limited competencies in data analysis. The area also goes through a critical period to start developing more open educational science (Veeramachaneni et al., 2014), in which hypotheses are pre-registered, data and analysis are open, and articles published are free for any researcher and educator.

3. SHADOWSPECT: A GEOMETRY PUZZLE SOLVING GAME

As part of this research proposal, we will use a geometry puzzle-solving game Shadowspect[1] (https://shadowspect.org/). The objective of the game is to solve a series of puzzles by selecting, placing, and rotating 3D geometric primitives such as cubes, cylinders, spheres, cones or pyramids based on the three orthogonal views of a figure. Figure 2 shows a puzzle example in Shadowspect, with some zones delimited by red parallelograms and a letter to facilitate the following description. When users start a puzzle, they see its overall description (A), and they receive a set of silhouettes (B) from different views that represent the figure they need to build. Users can create (C) cubes, pyramids, ramps, cylinders, cones and spheres. Additionally, several puzzles set up constraints, such as using a maximum number of objects or a maximum number of shapes of each type. Learners can use various tools (D) to achieve in-game goals by moving, rotating, and scaling shapes around the stage to match the silhouettes provided. Additionally, users can delete and select multiple shapes at the same time. Students can change the camera view (E) to see the figure they are building from different angles and then use the Snapshot functionality (G) in order to generate the silhouette from the current view. Snapshots can help them know if their shapes match any of the solution silhouettes. Finally, users have to submit (G) their current shapes, and the system will evaluate if the solution is correct and provide them with feedback. Any interaction that users perform with Shadowspect is stored as clickstream data that will allow us to reconstruct the learning process that students undergo to solve each puzzle.

The data used for this work was collected as a part of assessment machinery development that later will be implemented in Shadowspect. The team recruited seven teachers to use the game for two hours in their 7th grade and 10th grade math and geometry classes. All students' interactions with the game were collected without any identifiable or personal data except for a nickname provided by each student. Even though we collected a large dataset with hundreds of students, we will use data from a single class of students to represent the typical situation that a teacher would face when implementing Shadowspect as part of the math curriculum. The data collection of the selected class involves 31 students that made around 54,829 events (an average of 1,768 events per user); students were active in the game environment for 33 hours (an average of 65 active minutes per student), and they solved a total of 448 puzzles (an average of 14 puzzles per student).

[1] https://shadowspect.org/

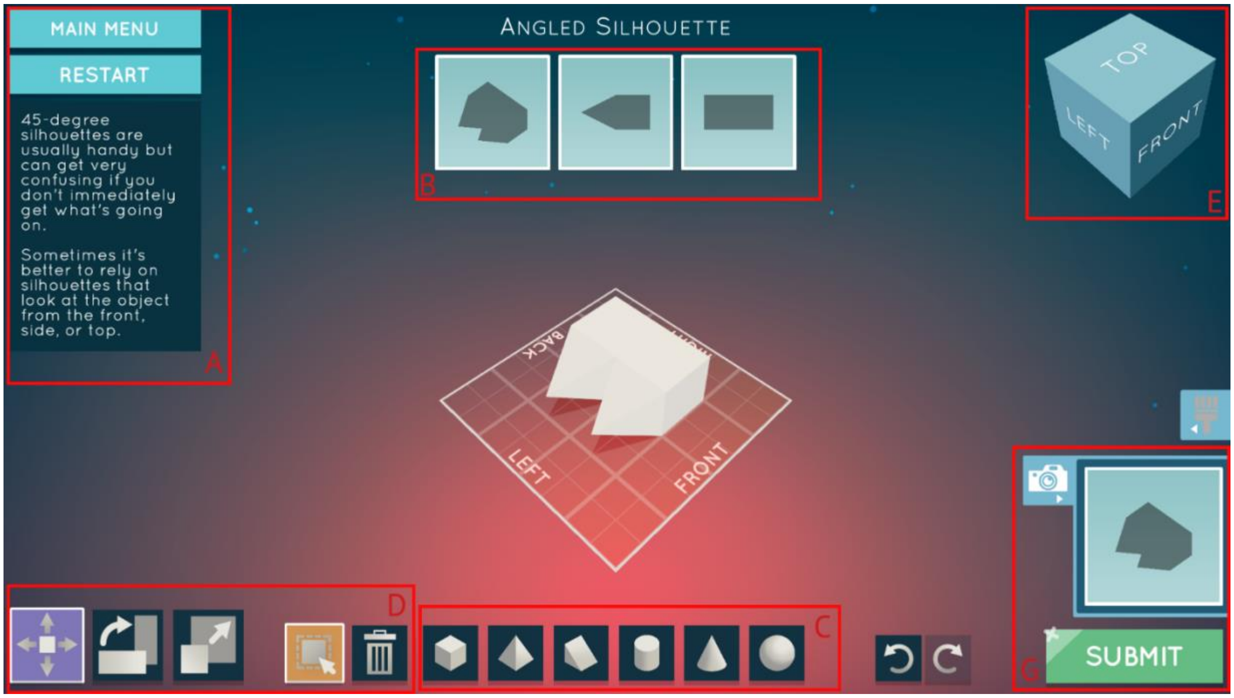

Figure 2. A puzzle example in a geometry game Shadowspect

4. LEARNING ANALYTICS CASE STUDIES IN SHADOWSPECT

In this section we present the four case studies that will exemplify the potential of LA in GBL.

4.1. Data-Driven Game Design

In game development, data from users has excellent potential to be used for accomplishing the learning and assessment goals. Analysis of these data results in deriving insightful metrics which can explain both players' overall enjoyment and efficiency of learning and qualities of assessment (Kim & Ruipérez-Valiente, 2020). The difficulty is one of the most common core game elements that game developers compute and apply data-driven methods on. The game designers state that the difficulty should balance boredom and frustration so that the players can experience a "well-shaped difficulty curve" (Aponte et al., 2009). Several authors (Gee, 2005), (Hamari et al., 2016) proved that the challenging and therefore motivating game atmosphere positively affects learning. In this case study, the authors explore the potential of LA within the explanation of how the metrics of difficulty were created and applied to the Shadowspect geometry game.

The primary objective is to estimate the relative difficulty of each puzzle in Shadowspect. Additionally, we aimed to determine the sequence of puzzles where the game has a well-shaped difficulty curve. Moreover, we questioned the amount of time that players of varying abilities would take to complete a number of puzzles. For fulfilling the goals as mentioned earlier, we performed a data-driven approach. Accordingly, we identified that difficulty consists of two measures: firstly, the level of effort required by learners to solve the puzzle and secondly, the relative complexity of the puzzle (i.e., fewer players can solve it). For example, difficulty value is affected if a user needs to make more actions or needs more time to complete the puzzle or if a vast majority of users are not capable of finishing the task. Based on these estimations, we computed the following four metrics per each puzzle level which directly influence the difficulty:

1. Average time per puzzle completed - total time spent in puzzles divided by the number of puzzles completed correctly.
2. An average number of actions per puzzle completed - a total number of actions performed divided by the number of puzzles completed correctly.
3. Percentage incorrect - the number of wrong submissions divided by the total number of submissions.
4. Percentage abandoned - number of different puzzles that were started and not completed correctly divided by the number of puzzles started.

For the reason that the metrics stated above are represented in different units such as percentage and time, we generated a composite difficulty measure. On these grounds, we first compute the z-scores of each metric and then normalize the sum of the four z-scores over the maximum. Accordingly, the final range of the composite puzzle difficulty of the geometry game Shadowspect varies from 0 (the easiest puzzle) to 1 (the hardest puzzle). In Figure 3, we illustrate the results of all four metrics per puzzle and the normalized composition of them called general difficulty measure. The puzzles are ordered in the way they appear in the game, so the game developers considered that the puzzle named “1. One Box” is the easiest, and the puzzle “Zzz” is the hardest. However, we can see that there are several puzzles that have a maximum difficulty, and at the same time, they appear in the middle of the game. This allowed us to reevaluate the distribution and allocation of puzzles into basic, intermediate and advanced groups of puzzles. Based on this difficulty representation, we had to review the overall difficulty curve of the entire game and fine-tuned the sequence of the puzzles.

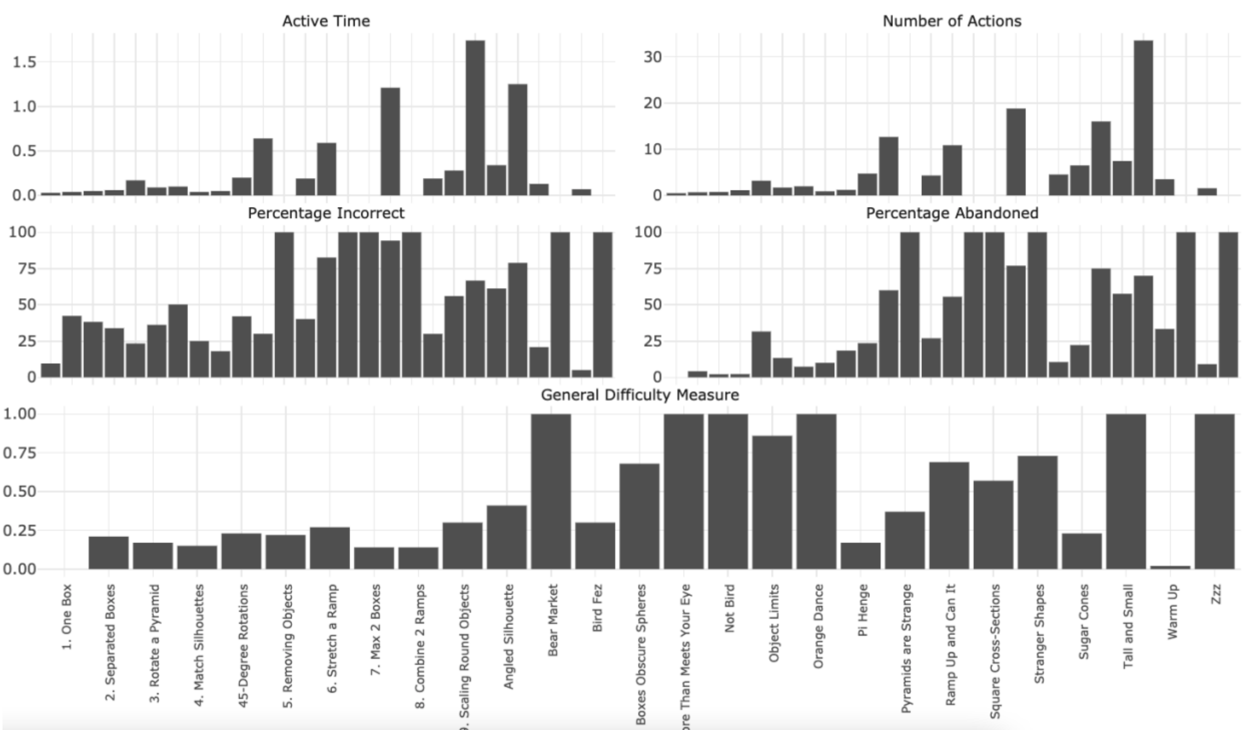

Figure 3. General difficulty measure and four metrics that compose it

For estimating the amount of time that players of varying skills would take to complete a number of puzzles, we calculated the amount of puzzles completed per unit of time. We concluded that during a one-hour session, a student with standard abilities would be able to finish on average 12 ± 1 puzzles. In contrast, in a two-hour session, the amount of puzzles would increase to 15 ± 1 puzzles.

By computing these empirical and realistic estimations, we will be able to provide them to teachers who want to use the Shadowspect geometry game in their classrooms. We proved that difficulty could be operationalized based on the varying notions in the literature (Kim & Ruipérez-Valiente, 2020).

4.2. Learner Modelling & Adaptive Learning

There is an extensive variety of methods that aim to measure the existing knowledge and forecast the future performance of users. They often fall within the context of a sequence of knowledge inference problems whose main goal consists of predicting or modelling the students' knowledge (or its absence) over questions as they interact with a learning platform at a specific time. In this case study, we will exemplify how we can use multivariate Elo learner modelling to predict if students are going to solve the next puzzle correctly and consequently adapt the sequence of puzzles.
Elo rating is a system for relative skill calculation of players (Pelánek, 2016) that has been predominantly used to rank players, for example, in competitive games such as chess tournaments. It was primarily developed for measuring how strong the players are, but it was also adapted into the context of educational research and was used for measuring both learner ability and task difficulty (Pankiewicz, et al., 2019). The basic idea of the Elo rating system is as follows: a score is assigned to each player, and after each game, this score is updated in proportion to how surprising the result of the game was. For example, if a weak player beats a strong one, the results are unexpected, and therefore the update is big.

In Elo rating system, first, we must obtain the probability that a student will answer correctly a question by using a logistic function with both the competence of the student θ_s and the difficulty of the question d_i while the correctness of an answer of a student on an item is $correct_{si} \in \{0,1\}$:

$$P_{correct_{si} \epsilon \{0,1\}} = 1/(1 + e^{-(\theta_s - d_i)})$$

Next, we calculate the probability of each student-question confrontation. Initial values of θ_s and d_i parameters are set to 0. The value of the constant K determines the behaviour of the system (i.e., if K is small, the estimation converges too slowly). The following equations represent updates for both the competence of the student and the difficulty of the puzzle:

$$\theta_s = \theta_s + K * (correct_{si} - P(correct_{si} = 1))$$

$$d_i = d_i + K * \big((P(correct_{si} = 1)\big) - correct_{si})$$

In the current case study, we were mainly interested in the competency of the student since we already calculated the puzzles' difficulty during the first case study. Therefore, in Figure 4, we represent the result obtained by using the equations stated above.

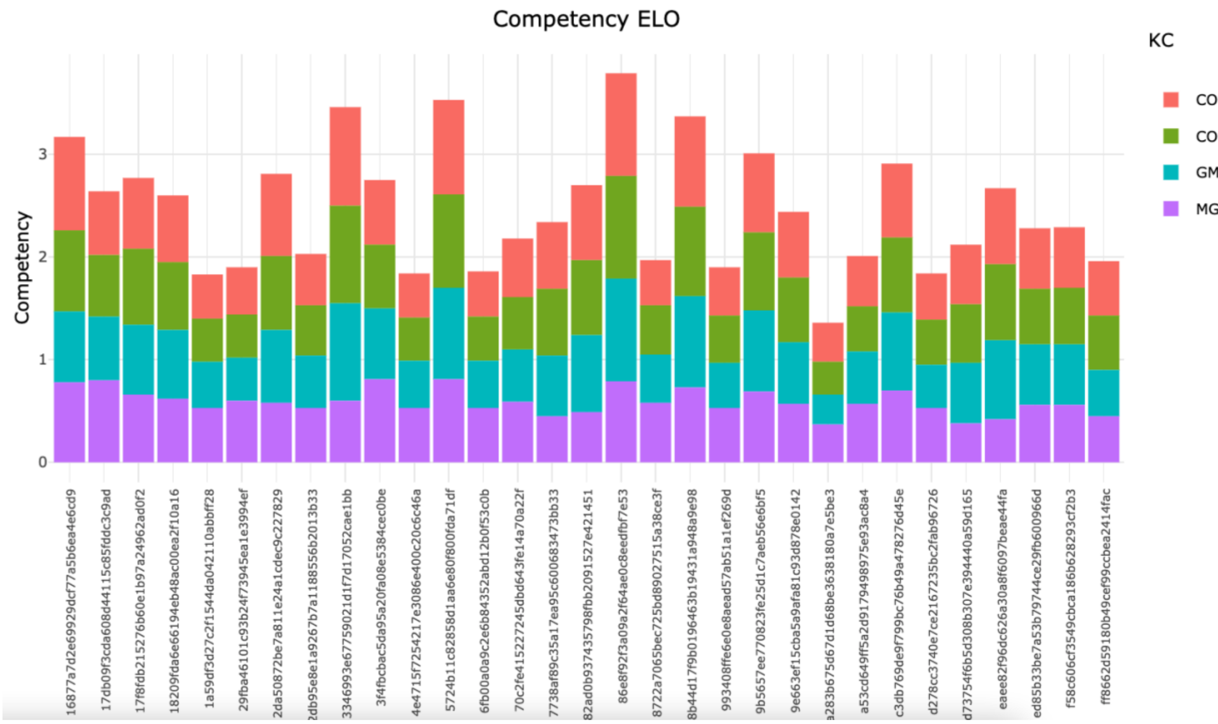

Figure 4. Students' competency by KC according to the Elo rating system

We illustrated the competency for each student for each Knowledge Component (KC) which are the skills needed to complete a puzzle successfully. In the visualization, each KC is differentiated by colours, normalized to 1; therefore, the maximum score of a student can be four since we work with four competencies. Across Shadowspect, experts defined four main KCs:

1. MG.1: Use geometric shapes, their measurements and their properties to describe objects.
2. GMD.4: Identify the shapes of the two-dimensional cross sections of the three-dimensional objects, and identify the three-dimensional objects generated by the rotations of the two-dimensional objects.
3. CO.5: Given a geometrical figure and a rotation, reflection or translation, draw the transformed figure using, for example, graph paper, tracing paper or geometry software. Specify a sequence of transformations that will take one given figure to another.
4. CO.6: Use geometric descriptions of rigid movements to transform figures and predict the effect of a given rigid movement on a given figure; in the case of two figures, use the definition of congruence in terms of rigid movements to decide if they are congruent.

As can be observed in Figure 4, the teacher, at a glance, could obtain the students' score for evaluation and can also identify where each student fails or if they have any mastered skills. Several interesting cases appear to be analyzed in the class visualization. First, we see that the user "86e8f92f3a09a2f64ae0c8eedfbf7e53" stands out above the rest with the best competence; in addition he draws attention that it has a high value for all the KCs, so we can consider that he has solved the puzzles and acquired the expected knowledge. In contrast, we can see the case of "a283b675d67d1d68be3638180a7e5be3" where the scores of the components are all weak. In this situation, the teacher can quickly see that the student has not shown interest in the task or has failed to acquire the necessary knowledge. To diagnose the problem, we can analyze the activity of such students to see if they have invested time, generated events, solved any puzzle, etc.

Finally, after seeing how beneficial this metric could be, it is also important to mention that it is not complicated to implement and adapt the Elo algorithm to the data. Moreover, the algorithm has few adjustment parameters, and it is also computationally very simple and fast (Veldkamp & Sluijter, 2019). By introducing the students' competency metric, we are confident that this can serve for the context of adaptive learning. By knowing how well each student is performing and what difficulties they are facing, the teacher would be able to adapt the learning process to address the special needs of every student. Moreover, the teacher can build the future classes based on the previous experience, for example, by explaining in more detail some topics that the earlier students found challenging. Additionally, the game itself automatically can be adapted to each user by providing hints or choosing the difficulty level according to the previous performance of the user. This would not involve the actions of the teacher; however, in a traditional classroom, it is still preferred that the teacher controls the class by using the metrics such as the puzzle difficulty and the competency of the user as described earlier.

4.3. Game-Based Assessment of 21st century skills

LA is widely applied to estimate the continually updated learners' knowledge, skills, and other attributes based on multiple observations in diverse contexts. Through GBA, we can behave in a diagnostic and formative way – to analyse how the student's way of thinking becomes "visible", thereby informing both learners and teachers about their train of thought, beyond right or wrong answers. Additionally, most children and teenagers play games on an almost everyday basis and consider them as an activity that makes life way more enjoyable. At the same time, game developers benefited from this situation and created a lot of games that are capable not only to entertain but also to educate even those users who are typically not very interested in learning. In this way, stealth assessment methods in GBA provide the opportunity of evaluating users without interrupting the gaming flow. Previous research has proved that learners are only able to show their actual skills and perform well if they are appropriately motivated and under no stress situations, and accordingly, GBA can help with this. As we mentioned earlier, gamification in learning is a perfect solution for maximising both enjoyment and engagement through capturing the interest of users and inspiring them to continue the learning process. In this case study, we will exemplify the potential of GBA in LA by measuring persistence in the Shadowspect geometry game. Persistence is a facet of conscientiousness that reflects a dispositional need to complete complex tasks and the desire to exhibit high-performance standards in the face of frustration (Ventura et al., 2013).

While exploring other works that have calculated the persistence metric, we found one example (Shute et al., 2015) where time spent on unresolved trials is a critical factor. Time in resolved practices is also important; however, it is more linked to the previous knowledge. The authors of (Shute et al., 2015) made an experiment consisted of exposing students to six problems, three easy and three impossible (the complete solution was not possible). In this way, we could see the time that each student dedicated to the tasks that are impossible, so it is a critical factor of the persistence that it presented. In (Israel-Fishelson & Hershkovitz, 2020), we saw how another measure of persistence is used. In this case, the indicator is the number of attempts to achieve solutions to different types of difficulty. The authors collected the number of attempts for each of the difficulties and the average.

Although there are few indicators on how to calculate persistence, it can be observed that time, both for completed and uncompleted activities, and the number of attempts are essential characteristics for persistence. Following some related works, it can be seen that the rest of the parameters are more linked to the specifications of each scenario where it has been implemented.

For the Shadowspect case study, in order to see if the student has been persistent or not in a puzzle, we face it as a problem of measuring a student's constancy in the face of adversity by considering several metrics, including a number of basic persistence parameters and others that help contextualise, give flavours and see the evolution over time. First of all, the more complex the puzzle is, the more persistent it should be since it is more challenging. Secondly, we will build the persistence metric on a time/attempts basis, with multipliers for the difficulty. Accordingly, we will consider percentiles of each of the parameters considered (time, attempts), supposing that the student was persistent if the respective value exceeds the value of 75%. Lastly, for each student, we will identify the puzzles in which he has been persistent. Then, we will calculate if the student globally has been persistent or not according to the number of puzzles in which he has been persistent. For example, to compute persistence at a puzzle attempt level, for each puzzle attempt, we point out the ratio with the features related to the persistence. To calculate persistence at a student level, we point out the ratio with all puzzle attempts for each student. Based on the assumptions and examples listed above, in Figure 5, we illustrate the persistence levels of a particular class where the beach figure means low persistence, desert stands for medium persistence, and mountain signifies high persistence. We can observe the fact that most of the students demonstrate a medium persistence level which is a common situation because initially, this metric was evaluated by the example of real students, and on average, they had medium persistence. Surprisingly, there are many more students who have low persistence in comparison with a few who have a high level of persistence. This could be explained by the fact that there was low interest or motivation in this particular class.

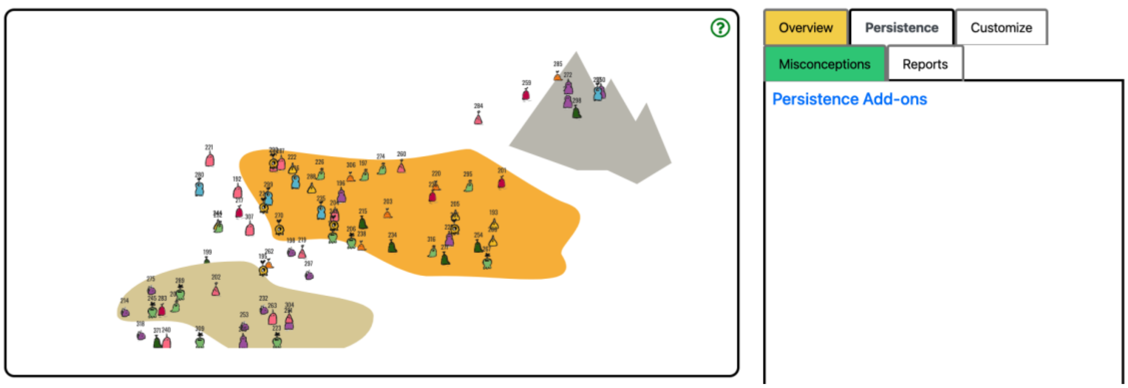

Figure 5. The persistence levels of a particular class

Moreover, teachers may build their own filters or create alerts for specific features they are interested in. In the particular case of persistence, an alert can be established when a student is experiencing difficulties in solving the puzzles and therefore appears as less persistent. This would allow teachers to take some actions in order to motivate the student and point them to the right solution. It brings flexibility for teachers and a more unique learning process for students.

4.4. Teacher-oriented Visualization Dashboards

While games proved to have a positive effect on the learning process, they are not so frequently used as part of classroom activities. One of the main reasons for that is related to the competencies of teachers. Moreover, most of them do not totally understand how the students interact with the game and how beneficial it is to implement and utilize it in their classrooms (Ruipérez-Valiente & Kim, 2020). On the other hand, there are more routine-related and logistic issues connected to the limited time available in the classroom and the impossibility of introducing games in the context of the regular classroom. Educational games were brought into regular classrooms to transform the skeptical mindset of all the people involved in the learning process. In order to help teachers better navigate and monitor the interactions of students, there were introduced various tools, including providing LA dashboards that present low-level interactions in visualizations easy to understand. While their power and strengths are apparent, it is not a trivial task to create an intuitive and easy to use dashboard. Therefore, the design of a dashboard for games for classroom use should be user-centred and take into account the usability and the needs of the teachers. In this case study, we will explore the potential of implementing a LA dashboard to support teachers that are using games in the classroom.

The working group for creating the dashboard was composed of two learning designers, one educator, one assessment scientist and one LA expert. Each member of the team proposed straightforward ideas of measures that would be interesting to consider for the dashboard using data from Shadowspect. After the initial ideas were formulated, we conducted the session where each proposed metric was discussed and voted for. All the metrics could not be included because the scope of the project was limited. Therefore, we decided that the dashboard should include only those metrics that can directly help teachers to improve the learning process at the same time, considering how difficult it would be to technically implement these metrics. The final dashboard consists of various metrics. Next, we will explain the most important ones that were not described before.

Funnel by puzzle metric requires first the explanation of what a funnel is. A conversion funnel is an e-commerce term that describes the different stages in a buyer's journey leading up to a purchase. Accordingly, in the Shadowspect case study, we use the funnel to illustrate the different possible stages that a student can reach while trying to solve a puzzle. We defined the following four stages for the funnel: *started stage* meaning that the student started the puzzle, *create_shape stage* when the student set up a primitive shape into this particular puzzle, *submitted stage* signifying that the student checked the puzzle solution and *completed stage* when the student submitted the puzzle and the solution is correct. Levels of Activity metric implements a set of parameters that describe the levels of activity of the student with Shadowspect. These are straightforward parameters to compute based on a feature engineering process, such as the active time, the number of events, different types of events, and the number of different types of events like snapshots, rotations, movements, scaling, shape creations and deletions, among several others. For our case study, we highlight only two of the parameters as mentioned earlier, namely, active time and the number of events; since these are the most important to look at when analyzing students' interaction with the game, however, we would like to denote that all of them are available for the teacher.

Finally, Sequence Within puzzles metric obtains a sequence of actions of every student in each puzzle. By doing that, we can know every single action a student performed while solving a puzzle. We only keep the main events that are related to the puzzle-solving process, which are starting a puzzle, manipulation events on a shape, puzzle submitting, snapshots and perspective change. To reduce the number of rows in the data, we collapse identical consecutive events, adding a field that indicates the number of times that an event has been performed in a row.

Next, in Figure 6 and Figure 7, we represented two examples of a real-time dashboard that we generated. Through them, teachers can see the metrics of the entire class or filter by a particular student. These dashboards can be used in a regular classroom for dynamically visualizing the different metrics to support the sessions and provide personalized feedback. For example, by analysing the Sequence within Puzzles metric represented in Figure 6, the teacher can see the sequence of actions, denoting that the student performed a large number of actions in order to solve the puzzle. As the student made a significant amount of actions between submits, we know that the student was mindfully trying to solve the puzzle, instead of making random actions.

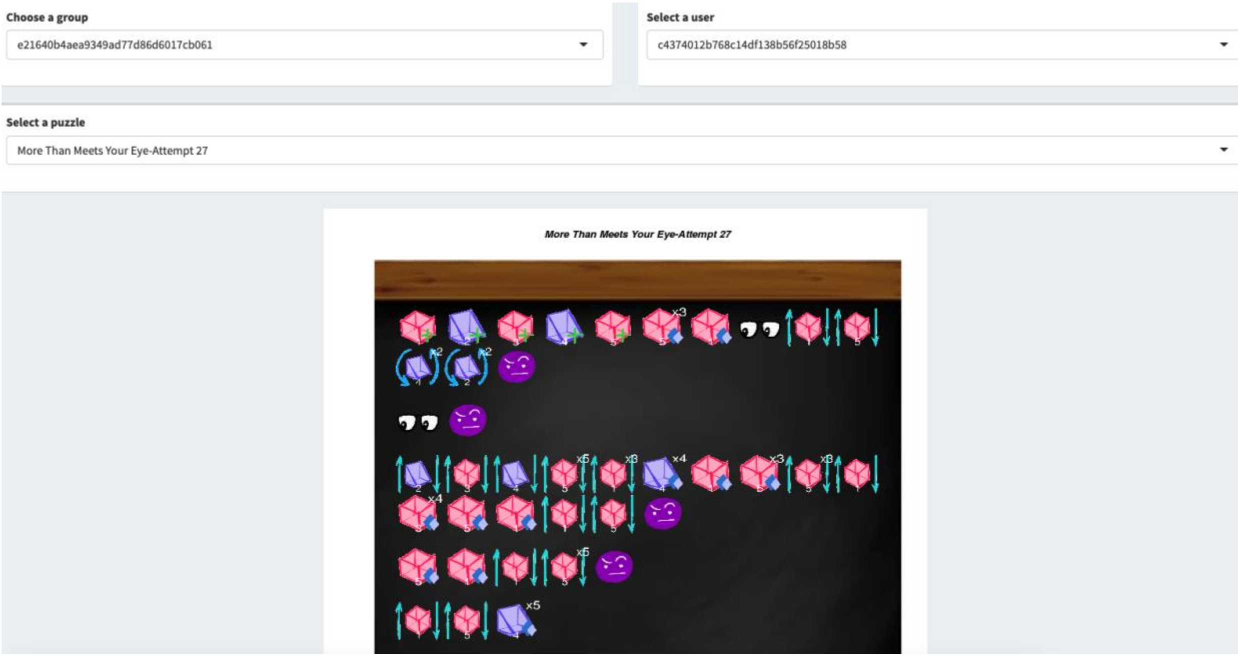

Figure 6. A real-time teacher-oriented visualisations dashboard (Sequence within Puzzles metric)

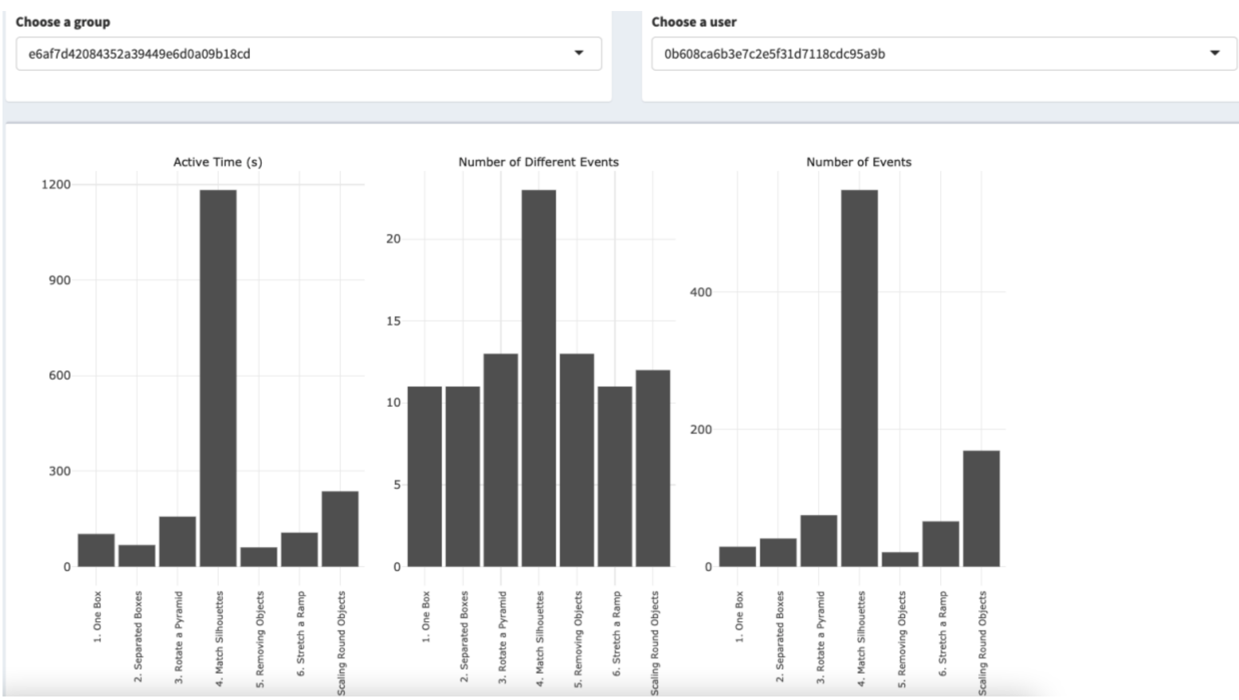

Figure 7. A real-time teacher-oriented visualisations dashboard (Levels of activity metric)

5. CONCLUSION

The objective of this research was diverse: first, to explain how beneficial it is to introduce GBA and LA into the educational process but also what challenges it brings. The following goal consisted of describing the LA process. We described the context of the environment, the raw data collection and its storage in the scope of the paper. Next, we delved into the steps of data manipulation, feature engineering and data analytics. In the last step, we characterised the educational application of this process.

After describing the LA process and motivating the research in this area, we introduced Shadowspect - a geometry puzzle game created for the development of geometry and spatial reasoning skills. In its context, we aimed to propose a series of metrics that can provide comprehensive information regarding the learning means of students while solving the puzzles in the game. We computed the puzzle difficulty, then we calculated the competency of students by the use of the Elo rating system, and lastly, we presented the persistence metric. Finally, we described the implementation of a real-time dashboard with simple but detailed visualisations of the metrics as mentioned above that can allow teachers to track the students within their class. Accordingly, the teachers will be able to evaluate or detect problems quickly and effectively, addressing the unique needs of each student and giving them personalised attention.

The metric design, their definition and calculation is a challenging process, especially when implementing them in games. This is due to the fact that, in general, modelling learning in games is stimulating because they are open environments where students should keep a friendly and motivating atmosphere all the time. To address this problem, we depicted a user-centred design process (Shum et al., 2019) targeting metrics that can support their implementation in our particular case. This process involved the work of a multidisciplinary team that was able to bring diverse perspectives to the LA design process (Suthers & Road, 2013). We selected the most essential metrics that were not a very challenging process to implement and that were easy to interpret from the teachers' side. The designed LA dashboard heavily relied on the metrics designed before. In the end, it represents simple but powerful visualisations of metrics for teachers who can easily use them in their educational process after a short introduction from the developers. On the other hand, our approach is aligned with current concepts of designing translucent LA with teachers (Martinez-Maldonado et al., 2020) in order to make learning visible, improve awareness and accountability.

We believe that our work presented a groundbreaking advancement for the implementation of GBA and LA into the educational process in the classroom. By performing the case studies as mentioned earlier, we have proved that GBA has a vast potential to be easily adapted in the class. The dashboard that we presented depicts an excellent opportunity for educators to perform a live monitoring of their students directly during the class. Data-driven game design, particularly for educational games, could help designers and researchers to ensure that a game is not only playable and enjoyable but satisfies other educational purposes such as learning and assessment. This can serve for supporting formative assessment and moreover developing more LA dashboards in GBA and other environments that we characterised during our work.

ACKNOWLEDGMENT

Blinded for review.

REFERENCES

Aponte, M., Levieux, G., & Natkin, S. (2009). Scaling the Level of Difficulty in Single Player Video Games Entertainment Computing – ICEC 2009. *International Conference on Entertainment Computing*, 25–35. internal-pdf://aponte_2009-2317115653/Aponte_2009.pdf%5Cnhttp://dx.doi.org/10.1007/978-3-642-04052-8_3

Bollen, J., van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M. A., & Balakireva, L. (2009). Clickstream data yields high-resolution Maps of science. *PLoS ONE*, *4*(3).

https://doi.org/10.1371/journal.pone.0004803

Breslow, Lori; Pritchard, David E.; DeBoer, Jennifer; Stump, Glenda S.; Ho, Andrew D.; Seaton, D. T. (2013). Studying Learning in the Worldwide Classroom Research into edX's First MOOC. *Environmental Psychology and Nonverbal Behavior*, *2*(4). https://doi.org/10.1007/BF01173772

Chen, C. M., Lee, H. M., & Chen, Y. H. (2005). Personalized e-learning system using Item Response Theory. *Computers and Education*, *44*(3), 237–255. https://doi.org/10.1016/j.compedu.2004.01.006

Drachsler, H., & Kalz, M. (2016). The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges. *Journal of Computer Assisted Learning*, *32*(3), 281–290. https://doi.org/10.1111/jcal.12135

Freire, M., Serrano-laguna, Á., Iglesias, B. M., Martínez-ortiz, I., Moreno-ger, P., & Fernández-manjón, B. (2020). Learning, Design, and Technology. *Learning, Design, and Technology*, 1–29. https://doi.org/10.1007/978-3-319-17727-4

Gee, J. P. (2005). Learning by Design: Good Video Games as Learning Machines. *E-Learning and Digital Media*, *2*(1), 5–16. https://doi.org/10.2304/elea.2005.2.1.5

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Israel-Fishelson, R., & Hershkovitz, A. (2020). Persistence in a Game-Based Learning Environment: The Case of Elementary School Students Learning Computational Thinking. *Journal of Educational Computing Research*, *58*(5), 891–918. https://doi.org/10.1177/0735633119887187

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8474 LNCS*, 29–38. https://doi.org/10.1007/978-3-319-07221-0_4

Karthik Ramasubramanian, & Singh, A. (2019). Machine Learning Using R. In *Frontiers in Artificial Intelligence and Applications* (Vol. 321). https://doi.org/10.3233/FAIA200024

Kim, Y. J., & Ruipérez-Valiente, J. A. (2020). Data-driven game design: The case of difficulty in educational games. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12315 LNCS*, 449–454. https://doi.org/10.1007/978-3-030-57717-9_43

Liu, M., McKelroy, E., Corliss, S. B., & Carrigan, J. (2017). Investigating the effect of an adaptive learning intervention on students' learning. *Educational Technology Research and Development*, *65*(6), 1605–1625. https://doi.org/10.1007/s11423-017-9542-1

Macarini, L. A., Ochoa, X., Cechinel, C., Rodés, V., Dos Santos, H. L., Alonso, G. E., Casas, A. P., &

Díaz, P. (2019). Challenges on implementing learning analytics over countrywide K-12 data. *ACM International Conference Proceeding Series*, 441–445. https://doi.org/10.1145/3303772.3303819

Martinez-Maldonado, R., Elliott, D., Axisa, C., Power, T., Echeverria, V., & Buckingham Shum, S. (2020). Designing translucent learning analytics with teachers: an elicitation process. *Interactive Learning Environments*, *0*(0), 1–15. https://doi.org/10.1080/10494820.2019.1710541

Pankiewicz, M., & Bator, M. (2019). Elo Rating Algorithm for the Purpose of Measuring Task Difficulty in Online Learning Environments. *E-Mentor*, *82*(5), 43–51. https://doi.org/10.15219/em82.1444

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438–450. https://doi.org/10.1111/bjet.12152

Pelánek, R. (2016). Applications of the Elo rating system in adaptive educational systems. *Computers and Education*, *98*, 169–179. https://doi.org/10.1016/j.compedu.2016.03.017

Reich, J., & Ruipérez-Valiente, J. A. (2019). The MOOC pivot. *Science*, *363*(6423), 130–131. https://doi.org/10.1126/science.aav7958

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, *51*(1), 368–384. https://doi.org/10.1016/j.compedu.2007.05.016

Ruipérez-Valiente, J. A., Cobos, R., Muñoz-Merino, P. J., Andujar, Á., & Kloos, C. D. (2017). Early prediction and variable importance of certificate accomplishment in a MOOC. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10254 LNCS*, 263–272. https://doi.org/10.1007/978-3-319-59044-8_31

Ruipérez-Valiente, J. A., & Kim, Y. J. (2020). Effects of solo vs. collaborative play in a digital learning game on geometry: Results from a K12 experiment. *Computers and Education*, *159*(September). https://doi.org/10.1016/j.compedu.2020.104008

Shum, S. B., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, *6*(2), 1–9. https://doi.org/10.18608/jla.2019.62.1

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M., & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers and Education*, *86*, 224–235. https://doi.org/10.1016/j.compedu.2015.08.001

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Suthers, D., & Road, E. W. (n.d.). *P1-Suthers.Pdf*. 2–5.

US Department of Education. (2016). Future ready learning. *2016 National Education Technology Plan*, 1–106.

Veeramachaneni, K., O'Reilly, U.-M., & Taylor, C. (2014). *Towards Feature Engineering at Scale for Data from Massive Open Online Courses*. http://arxiv.org/abs/1407.5238

Veldkamp, B. P., & Sluijter, C. (2019). *Theoretical and practical Advances in Computer-based Educational Measurement*.

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers and Education*, *60*(1), 52–58. https://doi.org/10.1016/j.compedu.2012.07.003

ADDITIONAL READING

Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2019). A multivariate Elo-based learner model for adaptive educational systems. *EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining*, *1*, 228–233.

Dicerbo, K. E. (2014). International Forum of Educational Technology & Society Game-Based Assessment of Persistence. *Source: Journal of Educational Technology & Society*, *17*(1), 17–28.

Gomez, M. J., Ruipérez-Valiente, J. A., Martínez, P. A., & Kim, Y. J. (2021). Applying learning analytics to detect sequences of actions and common errors in a geometry game. *Sensors (Switzerland)*, *21*(4), 1–16. https://doi.org/10.3390/s21041025

Holman, C., Aguilar, S., & Fishman, B. (2013). *GradeCraft*. 260. https://doi.org/10.1145/2460296.2460350

Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2017). Applying learning analytics for improving students engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments*, *25*(2), 220–234. https://doi.org/10.1080/10494820.2016.1278391

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, *23*(4), 579–595. https://doi.org/10.1287/mksc.1040.0073

Scruggs, R., Baker, R. S., & McLaren, B. M. (2020). Extending deep knowledge tracing: Inferring interpretable knowledge and predicting post-system performance. *ICCE 2020 - 28th International Conference on Computers in Education, Proceedings*, *1*, 195–204.

KEY TERMS AND DEFINITIONS

Clickstream Data: The traces of all clicks that users make, normally, in the game context.

Game-Based Assessment: A field which goal is to use games in order to educate and learn.

Item Response Theory: A group of mathematical models whose goal is to find a relationship between latent traits.

Knowledge Component: The skills needed to complete a task correctly.

Learning Analytics: A collection of students-related data and their analysis in the educational context.

Learning Management System: A software application for various tasks including administration, documentation, tracking, automation and reporting of educational courses or learning processes.

Massive Open Online Course: An online course whose main goal is to educate and normally provide an option of free and open registration and a publicly-shared curriculum.