IEEE Access
The journal for rapid open access publishing

IEEE Xplore Digital Library
http://ieeexplore.ieee.org/Xplore

Ideating and Developing a Visualization Dashboard to Support Teachers Using Educational Games in the Classroom

JOSÉ A RUIPÉREZ-VALIENTE[1,2], *Senior Member, IEEE*, MANUEL J. GOMEZ[1], PEDRO A. MARTÍNEZ[1], AND YOON JEON KIM[2]

***Abstract*—Technology has become an integral part of our everyday life, and its use in educational environments keeps growing. Additionally, video games are one of the most popular mediums across cultures and ages. There is ample evidence that supports the benefits of using games for learning and assessment, and educators are mainly supportive of using games in classrooms. However, we do not usually find educational games within the classroom activities. One of the main problems is that teachers report difficulties to actually know how their students are using the game so that they can analyze properly the effect of the activity and the interaction of students. To support teachers, educational games should incorporate learning analytics to transform data generated by students when playing useful information in a friendly and understandable way. For this work, we build upon *Shadowspect*, a 3D geometry puzzle game that has been used by teachers in a group of schools in the US. We use learning analytics techniques to generate a set of metrics implemented in a live dashboard that aims to facilitate that teachers can understand students' interaction with *Shadowspect*. We depict the multidisciplinary design process that we have followed to generate the metrics and the dashboard with great detail. Finally, we also provide uses cases that exemplify how teachers can use the dashboard to understand the global progress of their class and each of their students at an individual level, in order to intervene, adapt their classes and provide personalize feedback when appropriate.**

***Index Terms*—Educational Games, Learning Analytics, Game-based Assessment, Visualization Dashboard, Technology-enhanced Learning.**

I. INTRODUCTION

During the last decade, technology has started to make a significant impact on educational environments. In the era of the Internet, mobile technologies, and open education, the need for changes to improve the efficiency and quality of education has become crucial. Big data and analytics can contribute to these changes and reshape the future of education [1]. Also, as a part of the social distancing regulations stated after the global pandemic of the COVID-19, remote teaching and learning has become the norm across many educational institutions [2]. The role of technology in the field of education can have many applications. These are four of the predominant directions: 1) included as a part of the curriculum, 2) as an instructional delivery system, 3) as a means to aid instruction, and 4) as a tool to enhance the entire learning process [3].

One of the most prominent examples of technology in education is the use of digital games for learning [4]. Playing video games is one of the most popular activities in the world. According to [5], the role of video games in the American family is changing: nearly three-quarters (74%) of parents believe video games can be educational for their children, and more than half (57%) enjoy playing games with their child at least weekly. This has prompted a rapidly increasing interest in using games in educational settings, not merely because "it is what kids are paying attention to," but because well-designed games are very closely aligned with the design of good educational experiences [6, 7].

Despite the affordances of games to have a positive effect on the learning process, they are not so frequently used as part of classroom activities. Several factors contribute to this potential loss in learning opportunities [8]. First, many teachers do not have a good sense of what students are doing within a game, leading to uncertainty on how to best implement and utilize the game in their classrooms [9]. Second, logistical concerns, i.e. limited time available in the classroom, also preclude game activities in a classroom context. For example, students cannot devote much time to learning complex controls or watching useless tutorials to get over the at-times steep learning curves of a game. Other obstacles include the skepticism that a particular game is suitable for learning purposes, the support material and time needed by teachers to successfully implement a learning game, and the schools' limited budget.

Researchers have tried to address these challenges in an effort to bring educational games to the classroom. On the issue of games as a black box that prevents teachers from understanding students' interaction within the game, appropriate tools have been developed to help monitor the interaction. Specifically, a common approach has been to provide learning analytic dashboards that present low-level interactions in easy to understand visualizations. These visualizations provide opportunities for awareness, reflection, and sense-making, and bring to the front the potential to improve learning, enabling teachers to help students get better at getting better [10].

[1]University of Murcia, Faculty of Computer Science, Murcia (Spain)
[2]Massachusetts Institute of Technology, Playful Journey Lab, Cambridge (MA, USA)

This research has been partially supported by the National Science Foundation of the United States (award #1935450), the MIT-SPAIN "la Caixa" Foundation SEED FUND, and the Spanish Ministry of Economy and Competitiveness through the Juan de la Cierva Formacin program (FJCI-2017-34926), and the research and development contract LAGA (Learning Analytics and Game-based Assessment). We would like to recognize *Shadowspect* design and development team at the MIT Playful Journey Lab, along with the folks at Firehose, for their collaborative work bringing the game to life.

Corresponding author: José A Ruipérez-Valiente (e-mail: jruiperez@um.es).

Digital Object Identifier: 10.1109/ACCESS.2021.3086703

Despite these opportunities, teachers may be reluctant to use a game if the dashboard is not intuitive and easy to use. The design of a dashboard for games to be implemented in a classroom should therefore be user-centered and take into account the usability and the needs of the teachers.

In this work, we present our design process and the final product of a live dashboard that can support the use of games in the classroom by providing a number of metrics related to students' activities and gameplay. For this purpose, we use *Shadowspect*, a 3D geometry computer game where students can create geometrical objects such as cones or spheres to solve 3D puzzles, developing their geometric, dimensional and spatial reasoning skills. We collected low-level data from *Shadowspect* and transformed them into useful metrics to facilitate teachers' understanding of students' interaction with the game. This information can be presented graphically as interactive visualizations that can be easily manipulated by the teachers who are using the game in their classes. More specifically, for this research, we have the following objectives:

1) To present our metric design process and definitions that can help teachers understand students' interactions with *Shadowspect*.
2) To present our dashboard design guidelines and final live interactive dashboard that allows teachers to monitor students' interaction with the game, providing powerful and interactive visualizations that graphically represent the metrics.
3) To present a case study with two use cases from data collected in K12 schools across the US using *Shadowspect*:
   a) A first use case using these metrics to understand the global progress in an entire classroom.
   b) A second use case using these metrics to understand students' progress in a classroom at an individual level.

The rest of the paper is organized as follows. Section II reviews background literature on student engagement and learning analytics. Section III describes the methods, including *Shadowspect* as well as the data collection. Section IV presents the metric design process and the definition of each metric. Next, Section V introduces the system's architecture and the dashboard design guidelines, and then in Section VI we describe both use cases using the dashboard implemented. Then we finalize the paper with discussions, conclusions and future work in Section VII and VIII.

II. Related Work

In this section we present a review of the literature in some areas which are related to our work: In Subsection II-A we present literature related to educational games, in Subsection II-B we review some learning analytics studies, and finally, in Subsection II-C we present some works that have developed visualization dashboards.

A. Educational Games

Today's society is making electronic devices and technology indispensable tools for any task in our daily lives. Education is no exception, and in recent years systems have been created that support education with promising results [11]. With the implementation of technology in the classroom, educational games have emerged as a powerful tool [12]: games that bring knowledge or skill to the student by removing traditional learning methods' rigidity. The combination of three factors has led to this explosion of games for learning [13]: firstly, there is extensive related research that has given the field a robust scientific base. Secondly, there is the ease of acceptance of games and technology in general by the new generations born in the midst of the technological era. Finally, we find the importance of video games today, which have captured the attention of the majority of the population across platforms and contexts.

It has been demonstrated that these games for learning purposes, when well designed, provide numerous benefits for the training of students [7]. Motivation is an essential factor since learners see games as separate from formal education and normal study obligations. This motivation has positive consequences, as it has been shown that an increased motivation, leads to an increase in performance and commitment [14], which in turn can lead to better learning outcomes. Another benefit that games can provide is the interactivity and quick responsiveness. A well-designed educational game can provide the learner with immediate feedback on how they interact with the game, which is beneficial for the learning experience [15].

Now the question arises: if they bring so many benefits, what prevents their frequent use in the classrooms? There is a myriad of reasons [16], first of all, there is the rigidity of the educational system, which is usually reluctant to incorporate new mediums. Although there have been multiple improvements in this aspect, some stakeholders still do not accept the introduction of innovation in the classrooms, rejecting the inclusion of technology in many cases. There are also logistical challenges related to school schedules' rigidity or the lack of economic resources to provide students with the necessary tools. As budget is often quite limited, and the game development (real users, actual environment) is so complex and expensive, developers usually have no assurance about the effectiveness of their game designs and the adequacy of the final product to the users' skills [17]. Finally, a strong reason is the refusal of teachers to adapt learning systems, because there are no clear guidelines on how to do so and there is an overhead of time to learn new skills and adapt current practices.

As a potential solution to the issue of how to implement educational games in the classroom, we believe that learning analytics with the development of metrics can play a crucial role transforming the raw data into meaningful information for teachers. We go beyond the state of the art by aiding teachers that implement *Shadowspect* dashboard in their classes, with a tailored learning analytics dashboard to support that implementation.

B. Learning Analytics and Metrics

Learning analytics is a field of research and practice that aims to collect and analyze data generated by a learner in a given environment [18], which can be applied to educational game data. The current data collection infrastructures allow systems to collect large datasets from students' interaction with educational games that need to be processed in order to be understood [19]. The design process of generating learning analytics metrics involves mapping the data to evidence. We can take some ideas from the evidence-centered design (ECD) framework, which views assessment as an evidentiary argument: an argument from which we observe students say, do, or make in a few particular circumstances, to inferences about what they know or do [20]. In this way, we can integrate these ideas into the design process of learning analytics metrics, to take into account the human nature of the field, incorporating expert-informed design decisions. Previous work [21] proposed a framework to use learning analytics in game-based assessment that was composed of three phases: 1) Design: The first step for design is to accomplish domain modeling, include design of a data infrastructure that can accommodate the game events; 2) Development: work on feature engineering to create variables related to the target competencies that we want to measure; 3) Evaluation: The last phase is the evaluation of the analytic model in terms of both construct validity as well as performance metrics. Within this process our work is situated in the development stage.

Learning analytics is a field of research and practice that aims to collect and analyze data generated by a learner in a given environment, which can be applied to educational game data.

These data can be analyzed, being not only useful for the evaluation of students [22], but it can also be used for future improvements in the design of educational games, to personalize the difficulty of the scenarios according to the student's abilities [23] or as in [24], to identify difficulties that students might be experiencing facing a task. Finally, one of the main advantages of learning analytics is to increase student engagement and improve learning, as engagement and learning are closely related [25, 14]. This personalized adaptation of scenarios and difficulty per student can go one step further with multimodal learning analytics, which aims to collect data external to the learning environment, such as the student's heart rate [26], and can be used for multiple purposes, such as to adjust the game difficulty based on the identified problems and levels of concentration [27].

One of the potential benefits of learning analytics is improving awareness. In the case of students, it can significantly improve the self-awareness, which can promote the students' motivation and their own planning within the activity. In the case of instructors, learning analytics can provide necessary information on the progress that students are doing to evaluate or adjust the learning method [28]. To provide this information effectively, metrics are often implemented to make sense of the raw data collected through a feature engineering process [29, 30]. Metrics can be understood as higher-level information measures extracted from the data and according to the specifications of each metric's purpose. For example, [31] developed metrics related to activity levels, difficulty and other patterns, so that the instructor could assess students based on those metrics. Each environment may have specific metrics; however, some are more common across environments, such as those related to the activity with numbers of events or active time.

Previous work has already developed metrics in the context of games for learning to measure engagement. For example, [32] differentiated four dimensions: the general activity, social, exploration, and game progress, finding four profiles of engagement. In this study, we implement some similar metrics, for example: for the general activity dimension, we implement a series of levels of activity and for the exploration dimension we analyze the funnel in the game puzzles. Another example of the importance of metrics for evaluative purposes is described by [33], which deals with the application of game-based learning of mathematics contents. It aims to study improvements with previous training through play and see if those metrics can be indicators of success. The study was carried out with students of about ten years old who completed mathematical tasks about rational numbers, using the game Semideus School. The control of the students' behavior and performance was recorded with different metrics, similar to our case but adapted to their type of game, and they proposed parameters such as time spent, the maximum level reached, number of games played, or general performance. As a conclusion, the use of metrics in game-based learning as part of the evaluation process shows great promise.

Once the metrics have been defined, the next step is to transmit them to teachers and/or students. One of the most effective ways to understand the metrics and data that have been applied in the literature is through the use of visualizations, since at first sight the most relevant aspects can be observed without the need for costly analysis of large amounts of data [34].

Finally, authors in [35] conducted a systematic literature review, analyzing 87 papers that reported evidence of the outcomes of the analysis of game analytics data and/or learning analytics data collected from serious games. Authors concluded that the application of data science can increase the still-limited application of serious games in education. Authors also noted that, when reporting measures, typical data such as completion times, interactions, or scores can and should be included; but research can benefit from moving on to more complex data extracted from in-game interactions. In this research, we use data science techniques to infer new knowledge and report measures such as completion times or percentage of completed levels. Moreover, we go beyond this state of the art by implementing more nuanced metrics such as those to detect common errors and sequences of actions, or by measuring students' competencies on geometry standards.

C. Visualization Dashboards

For the implementation of educational games in the classroom, we found inconveniences of several types such as the lack of human resources, the opposition of teachers to new teaching methods, or the fact that some teachers still believe that the implementation of educational games is a complex process that is beyond their reach. Our contribution with a dashboard for educational games helps diminish this last issue by providing the teacher with a simple yet potent interface for student analysis.

Dashboards as a tool to inform and transmit knowledge are used in many fields. Their importance and usefulness make them the subject of many studies. The correct design of dashboards is critical as we need to properly convey the information to the end-user. In [36] we can find an extensive guide that establishes the design bases for transmitting clear, fast, and convincing information. Although there are many design principles to follow, we can define four key points when designing a dashboard [37]: 1) they are visual presentations of information, 2) they show that information to achieve a goal, 3) they must be clear and summarized, and 4) they can be observed at a glance. The four premises described above for the design are consistent with the solution to one of the classic problems in the implementation of educational games in the classroom: the understanding of the student's activity. Furthermore, the learning analytics field has struggled with the adoption of such dashboards in practice, that is why certain authors have suggested involving the target users in the design process, making emphasis on a user-centered design approach [38, 39].

Previous studies have successfully made this dashboards in other types of learning environments, such as massive open online courses [40], or intelligent tutoring systems [41]. Other researchers have also made progress in providing approaches and frameworks for dashboard visualizations for classroom purposes. For example, the data storytelling approach [42] recommends that the visualizations should be driven by a particular purpose connected to teachers' intentions and goals. Similarly, other scholars are driven by the notion of translucence, where visualizations should make concepts teachers care about visible, raising teachers' awareness so they can make pedagogical decisions based on the visualized data [43, 44]. In addition, previous research has also identified goals and objectives for teacher-facing learning analytics dashboards. For example, they should provide feedback on students' learning activities and performance, pinpoint who may be at risk, and provide insights on the evolution of students' interactions on and with the learning platform [45, 46]. Targeted recommendations of visualization types (e.g., bar charts) with their corresponding category of feedback, for instance, have also been provided [46].

We propose a set of visualizations of the data collected in the educational game *Shadowspect*, as a tool for teachers to detect problems within a class [47, 48] or with a particular student [49], as proposed in the previous work [50]. With this visualization dashboard system, the raw data collected is transformed into metrics, and these are consumed via visualizations. This enables teachers to monitor what the students are doing with the game during the class period, intervene during the development of the activity when appropriate, or even use these metrics as part of the formative evaluation. This dashboard goes beyond state of the art in the implementation of educational games in the classroom, providing a more integral and robust solution to be used by teachers. Moreover, this solution has been co-designed with a cohort of fellows that are K12 math teachers [39, 43]. After following this human-centered learning analytics methodology [38], we have developed a dashboard with metrics that teachers found effective and with potential within their classes [51].

III. Methods

In this section we present in Subsection III-A *Shadowspect*, the interactive game that is used to perform this research, and in Subsection III-B the data collection retrieved using the game mentioned.

A. Shadowspect

Shadowspect [1] is a 3D geometry game designed as a formative assessment tool to measure math core standards (e.g. visualize relationships between 2D and 3D objects), so teachers can use it in their core math curriculum. In *Shadowspect*, students can create cubes, pyramids, ramps, cylinders, cones and spheres, which are considered primitive shapes. When students begin a puzzle, they receive a set of silhouettes from different views representing the composite figure they need to create combining these primitive shapes. After creating a primitive shape within the scenario, students can also scale, move, rotate or delete the shape in order to build a composite figure that matches with the silhouettes provided. Students can also paint the shapes in different colors. They can move the camera to see the figure they are building from different perspectives and then use the "Snapshot" functionality to generate a silhouette from a concrete perspective and see how close they are to the objective. Finally, students can submit the puzzle, and the game will evaluate if the composite figure matches all the silhouettes and provide them with feedback.

In the version of *Shadowspect* that we have used in this research, we have nine tutorial levels, nine intermediate and 12 advanced, bringing a total of 30 levels of increasing difficulty. The tutorial levels seek to teach the basic functionality of the game, so the students can learn how to build different primitives, scale and rotate them, how to change the perspective, take snapshots and so on. The intermediate levels allow students more freedom so they will not receive so much help to solve puzzles. Then the advanced levels represent a real challenge for experienced students.

B. Data Collection

The data collection 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 in order to use the game for two hours in their 7^{th} grade and 10^{th} 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.

Although we have collected a large dataset with hundreds of students, we 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

[1] More information available at https://shadowspect.org/

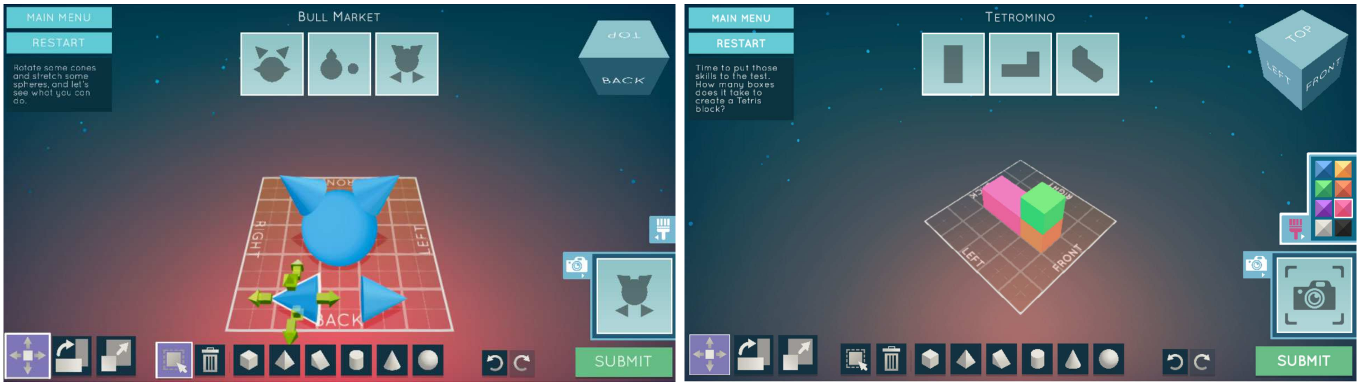


Fig. 1: Two puzzle examples in *Shadowspect*

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).

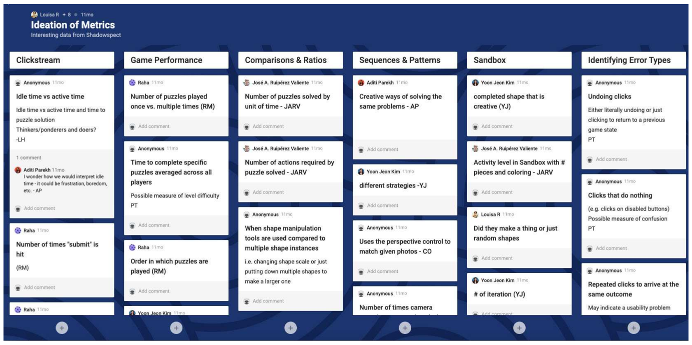
IV. Metric Design

In this section we explain the design of the metrics, which is divided into Subsection IV-A that describes the metric ideation process and Subsection IV-B that describes the definition of each one of the metrics.

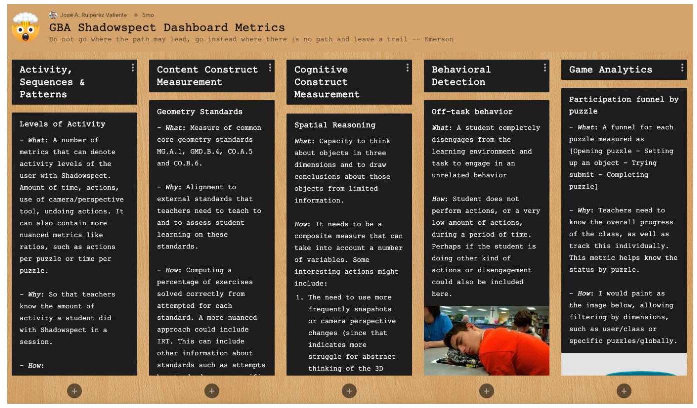
A. Metric Ideation Process

This subsection describes the process of ideating and defining a set of metrics that would be useful for teachers implementing *Shadowspect* in their classes. We describe next the steps that we followed:

1) **Define working group**: Since there have been more than ten people involved in this project as a whole, the first step was to define the working group that would be primarily involved in this process. The working group was composed of two learning designers, one educator, one assessment scientist and one learning analytics expert; the later one had the leadership of the working group. One key aspect is that since there was only one technical person experienced with data and analytics, this person had to greatly facilitate the process in terms of metrics that could be feasible and which suggestions were not viable.
2) **Initial ideation process**: The first task for the working group was to conduct brainstorming work. This initial session did not have very tight specifications, and every working group member was instructed to write down in a Padlet board straightforward ideas of things that would be interesting to measure using data from *Shadowspect*. The team then met to put in common these ideas and facilitate a conversation around the potential metrics. The Padlet board output for this session can be seen in Figure 2.1.
3) **Initial proposal of metrics**: This was the largest step of the process, and we had four sessions dedicated to it. These sessions had the following objectives:
   - *Session 1*: The first session was dedicated to ensuring that the whole team was aligned in terms of the objectives of the metrics and dashboard to an accomplished and user- and application-centered design. Therefore, we aligned the target user (teachers), the dashboard application (support for classroom implementation) and the high-level categories of the metrics.
   - *Sessions 2 and 3*: Each one of the team members was instructed to design just five metrics, specifying for each one of them 1) what do we want to measure?, 2) why is this significant [and to whom]?, and 3) how do we want to measure it and visualize it? Based on this list of metrics, sessions 2 and 3 were dedicated to discuss the proposals and merge those metrics that were similar enough into more generic metrics.
   - *Session 4*: Finally, the last session was employed in discussing the final selection of metrics, potential implementations and visualizations. This initial selection of metrics had a total of 11 metrics, and each one of them had responses to the what/why/how questions. A screenshot of the final Padlet board from this step is available in 2.2.
4) **Prioritization of metrics**: Since the scope of the project is limited, we performed a prioritization of the metrics. For this step in the process, we include the entire team of this project (over ten members). During this session, we provided an explanation of each metric, and we asked each one of the team members to review the Padlet board that was obtained as part of the previous step (i.e. Figure 2.2), and vote from one to five based on the following rubric:
   - *One Star*: I do not feel this is aligned with our principles and I would prefer not to have it.
   - *Two Stars*: I do not think we really need this, but it would be acceptable to have it.
   - *Three Stars*: It's interesting, but I do not think it is crucial.
   - *Four Stars*: It would be nice to have this metric, but not essential.



2.1 Padlet output with the initial brainstorming of ideas for metrics.



2.2 Padlet output with the final output of the initial proposal of metrics.

Fig. 2: Examples of product outputs from the working group that were the result of steps 2 and 3 of the metric ideation process.

- *Five Stars*: Yes, this is important and we need it!

5) **Narrow down the list of metrics**: We had a single session for this purpose in which we reviewed the ratings of the different metrics and the potential difficulty to implement them. Based on this, we agreed on a number of initial metrics to implement, which are the seven metrics that we present in this manuscript.
6) **Initial metric definition**: As the implementation moved forward, we had several additional meetings to co-define the initial implementation of each metric, in this process, again the learning analytics expert had a leading role in facilitating the technical side of this work.

B. Metric Definition

The output of the previous metric design process has been a set of seven different metrics, that we organize based on their purpose in three categories as follows:

- **Activity Metrics**: Includes Levels of Activity and Funnel by Puzzle.
- **Sequences and Patterns**: Includes Sequence Between/Within Puzzles and Common Errors.
- **Knowledge Inference Metrics**: Includes Levels of Difficulty and measurement of Geometry Standards using Multivariate ELO-based Learner Modelling.

Next, we explain the definition of each one of the metrics separately:

- **Funnel By Puzzle**: Before explaining the metric itself, we need to know 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. Thus, we use the funnel to illustrate the different possible stages that a student can reach while trying to solve a puzzle. We define the following four stages for the funnel: `started` (if the student has started the puzzle), `create_shape` (if the student has set up a primitive shape into this particular puzzle), `submitted` (if the student checked the puzzle solution) and `completed` (if the student has submitted the puzzle and the solution is correct). The metric outputs, for each student and puzzle, the number of times that this student reached each funnel stage in that concrete puzzle.
- **Levels of activity**: This metric implements a set of parameters that describe the levels of activity of the user with *Shadowspect*. These are straightforward parameters to compute based on a feature engineering process, such as the active time, number of events, different type of events, and number of different types of events like snapshots, rotations, movements, scaling, shape creations and deletions, among several others. For this case study we highlight only two of the parameters we mentioned, 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.
  - `active_time`: Amount of active time in minutes establishing an inactivity threshold of 60 seconds (i.e. if the time between two events is above 60 seconds, the user is considered to be inactive during that time and this slot is omitted from the computation).
  - `n_events`: Total number of events triggered within the game (every action performed by a student in *Shadowspect* is recorded as an event).
- **Levels of difficulty**: This metric provides a set of parameters that are related to the difficulty of the puzzles:
  - `completed_time`: This parameter is computed by dividing the amount of time invested in the game (`active_time`) by the number of completed puzzles.
  - `actions_completed`: This parameter is computed by dividing the number of actions (`n_events`) by the number of completed puzzles.
  - `p_incorrect`: This parameter is calculated by dividing the number of incorrect attempts by the total number of attempts (`n_attempts`) multiplied by 100.

  - `p_abandoned`: This parameter is computed by dividing the number of started puzzles by the number of completed puzzles.
  - `norm_all_measures`: A standardized and normalized measure of the four previous parameters together in a single value.
- **Sequence Between Puzzles**: *Shadowspect* presents a linear sequence of increasing difficulty puzzles. However, students do not have to follow this linear order. They can jump from any puzzle to another, regardless of its difficulty, pursuing their own interests and exploring the game. This metric also uses the four funnel stages defined on the funnel metric to analyze students' temporal interaction with the puzzles combined with the stage reached in each puzzle. That way, we can reconstruct the sequence of puzzles in chronological order for every student.
- **Sequence Within Puzzles**: In this metric, the objective is to obtain a sequence of actions of every student in each puzzle. By doing that, the teacher can know every single action a student has performed while solving a puzzle. We keep only the main events that are related to the puzzle solving process, which are starting a puzzle, manipulation events on a shape, a puzzle submission, 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.
- **Common Errors**: With this metric, we can automatically detect common errors in the sequences of actions that represent an attempt to a puzzle, by detecting incorrect patterns in the resolution of each puzzle. To do that, we compare in each puzzle the list of shapes provided by a student's solution with the correct list of shapes of that concrete puzzle. If the solution is incorrect, we register every manipulation event the student makes after the wrong submit (e.g. deleting a pyramid and creating a cone). Then, we group all these events by class, and we select those errors that are more common in each puzzle.
- **Multivariate ELO-based Learner Modelling**: ELO is a rating system for player ranking in games, that we have adapted to confront each student with a puzzle instead of two players. With the confrontation between puzzle and student, we can obtain each student's competence and the difficulty of each puzzle based on the history of the puzzle attempts. To compute the probability of a student solving a puzzle correctly, we consider their competency for each specific knowledge component and each puzzle's difficulty. For additional *formulae* you can consult previous work [52, 53].

V. Dashboard Design

The objective is to generate a real-time dashboard that teachers using *Shadowspect* in their classes can use to dynamically visualize the different metrics to support the sessions and provide personalized feedback. The section is divided in Subsection V-A where we describe the dashboard ideation process, Subsection V-B where we provide a technical overview of the whole system and Subsection V-C where we review the interface design principles that we have followed.

A. Dashboard Ideation Process

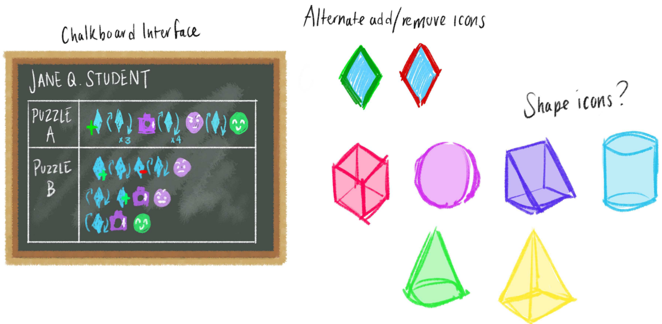
The ideation process to prototype and develop the dashboard was the result of the communication of three teams in the project that had the following responsibilities:

1) **Learning analytics team**: They were in charge of the technical implementation of the metrics using Python as well as the implementation of the dashboard with its final visualizations.
2) **Vizards team**: They were in charge of developing digital paper prototypes of potential visualizations for the metrics. Additionally, they also generated overall dashboard themes and examples of use cases to combine the visualizations.
3) **Co-design team**: They have been working with a cohort of school mathematics teachers to co-design the metrics and visualizations. Based on a series of activities with the teachers, we receive feedback that is incorporated into the final metric definition and visualization design.

The process to develop the final visualizations is iterative based on the feedback and on the interaction between the three teams. The main products from this dashboard design process were:

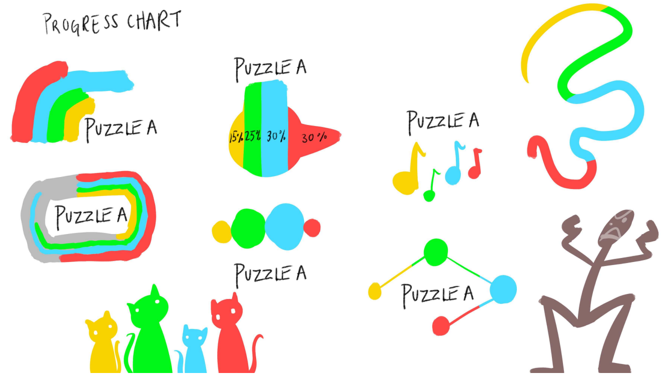
- **Final metric definition**: We made small adjustments to the metric definition that we already shared in previous Subsection IV-B.
- **Visualization prototypes for each metric**: We developed a set of visualization digital paper prototypes.
  For example, Figure 3a represents the digital paper prototype for Sequences Within Puzzles metric originally designed by the team. It is very similar to the final visualization that we will present in this paper as part of the dashboard, with icons representing each shape and action.
  Another example is available in Figure 3b, where we see different prototypes for the Funnel by Puzzle metric. We see the same colors used to represent the stages and different ways to represent the funnel concept.
- **Use case prototypes**: We prepared some potential use cases as digital paper prototypes regarding how teachers could combine the different metrics. This would help to generate useful connections between the metrics for the design of the dashboard.
  Figure 4a depicts the first example, where we have the Funnel by Puzzle metric for a specific level, representing the percentage of students that reached each funnel stage. We detect a puzzle with a bottleneck as depicted in the use case. Then, we look for the specific Level of Difficulty of that puzzle, which reveals that the puzzle has a medium-high complexity, explaining the bottleneck that we observe.
  In Figure 4b we see a transition between different metrics. We could zoom into every student, look into its Funnel

of activity icons 2.jpg



(a) Sequence within puzzles.

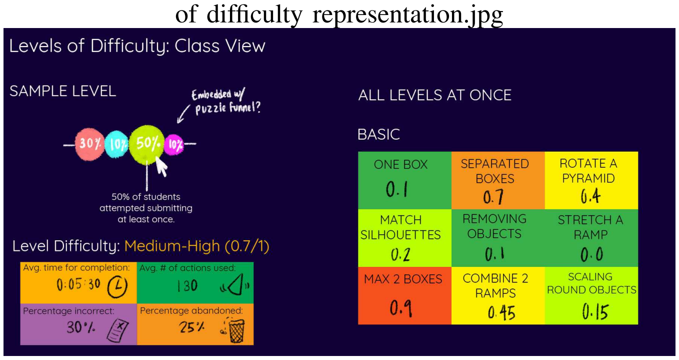
funnel 2.jpg



(b) Funnel by puzzle.

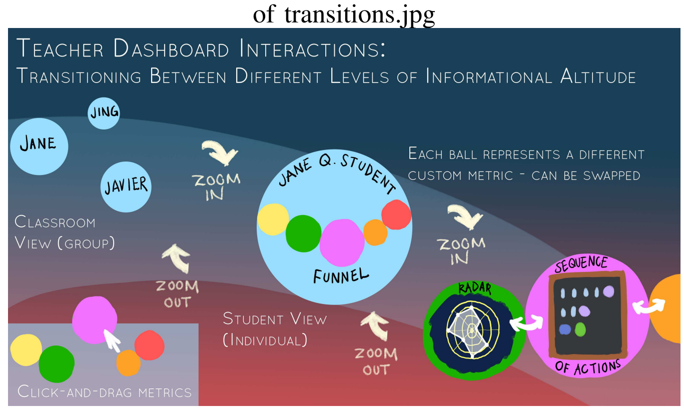
Fig. 3: Two examples of digital paper prototype visualizations.

of difficulty representation.jpg



(a) Funnel of Puzzles and Levels of Difficulty.

of transitions.jpg



(b) Funnel of Puzzles and Sequences Within Puzzles.

Fig. 4: Two examples of use case prototypes.

of Puzzles, and then continue to see more in-depth information with the Sequences within Puzzles and the Levels of Activity with a radar chart. This kind of transitions are very similar to the ones that we present in the case study in Section VI, and they highlight the potential to easily move between the different visualizations at a class or student level.

Based on the outputs of this design process, we then implemented the final dashboard product that we share in the following sections.

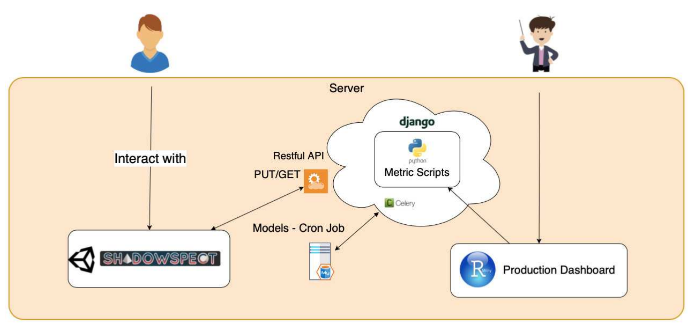


Fig. 5: System's architecture diagram

B. System's Architecture

In Figure 5 we have a diagram which contains the different elements that make up the architecture:

- **Shadowspect**: This module represents the geometry game that was developed for game-based assessment purposes as described in Subsection III-A. The game has been built using Unity Engine and deployed as a web application hosted in a web server. This facilitates accessing the game from multiple devices and without having to install any software to do so. Also, the game was developed as lightweight as possible, since we need students using Chromebooks or similar low capacity computers in schools to be able to use it.
- **Server backend with Django**: The main backend of the server has been built using the Django framework based on Python. *Shadowspect* communicates with this Django server using a RESTful API. Django also communicates with a MySQL database where all the necessary models have been defined. One of the challenges is keeping the metrics data up to date to make this a real-time dashboard.

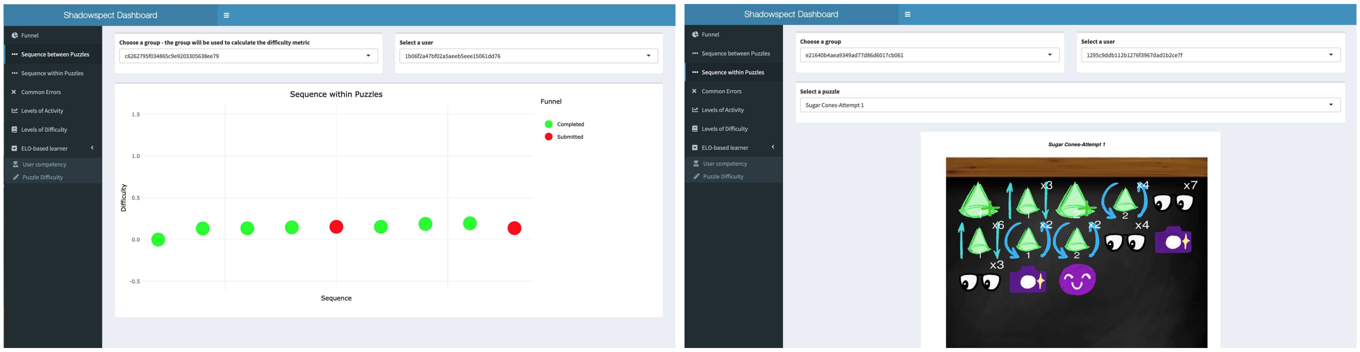


Fig. 6: Two examples of visualizations within the Shiny dashboard interface.

To do so, we use Celery, a task queue implementation for Python web applications used to asynchronously execute work outside the HTTP request-response cycle [54], so that we can schedule a cron job to execute the Python scripts every ten minutes and keep the metric's output updated.

- **Analytics processing**: Each one of the metrics that we have defined is a separate function that computes the required data output as defined in a Python script. These functions are called by the cron job that updates the data, the Python scripts directly import data from the MySQL database, computes the metric, and stores the metric output in the MySQL database, so that is directly accessible without delay.
- **Dashboard**: We have developed the dashboard using Shiny's R framework, and we have deployed it on ShinyApps web server. This brings a good number of benefits, such as that the entire deployment pipeline is very easy as it does not need any hardware or configuration of the system. ShinyApps is also secure-by-design with each application using its own protected environment, and access is always SSL encrypted. Finally, the resources allocated to the dashboard are scalable and we do not need to worry about balancing backend resources based on the system's current workload.
- **Users**: We have two kinds of users. On the one side, we have the students, that interact with *Shadowspect* generating the trace data with their interaction with the game. On the other side, we have the teachers, that are using *Shadowspect* in their classes and are the ones that can access the Shiny dashboard production environment to visualize what their students are doing.

C. Interface Design Principles

This section presents an overview of the interface and the visualization criteria applied for its design. As many authors note, we follow the principle that "Everything should be made as simple as possible, but not simpler" [55]. As teachers will use this dashboard, we want to prioritize making visualizations easily interpretable, so that they can use the information provided effectively. We have developed visualizations for each of the metrics defined as part of the previous steps.

Figure 6 shows two examples of the interface, the first one with the Sequence Between Puzzles and the second one with the Sequence Within Puzzles. As shown in the Figure 6, we can choose between the different visualizations using the sidebar with a tab for each visualization. We have carefully considered the following design decisions about the interface that we detail now:

- **Selection of group, users and/or puzzles**: The teacher can select the different groups, users and puzzles available with the selection boxes.
- **Visualization graphic**: The graphics are generated using *plotly* which already provides a certain level of interaction, for example, popping up some extra information when passing the mouse over certain elements of the visualization (e.g. showing exact values). This extra information provided facilitates the understanding of the visualization by teachers. In addition, visualizations also have a legend that helps to interpret the colors.
- **Interactivity between visualizations**: We have implemented links between related visualizations to make navigation more comfortable and intuitive following similar ideas from the use case prototypes. For example, we can click on any dot of the Sequence Between Puzzles visualization, and the dashboard will automatically show the Sequence Within Puzzle visualization for that concrete student and puzzle attempt.
- **Graph-type selection**: We have prioritized using the same type of graphs as often as possible (e.g. bar plots). If there is a time dimension, this is represented by the x-axis to see the evolution. Moreover, we have developed from scratch our own Common Errors and Sequence Within Puzzles visualizations, that use icons created explicitly for this purpose to facilitate the interpretability of these charts.
- **Color selection**: We have carefully decided the color palettes to use, since a right choice of colors in visualizations can help the users to detect some interrelations and patterns within data easily [56]. For example, for the visualizations based on metrics that use the funnel term, we have defined a qualitative color palette with red color for submitted incorrect levels, green color for completed levels, blue for started levels and finally yellow for levels

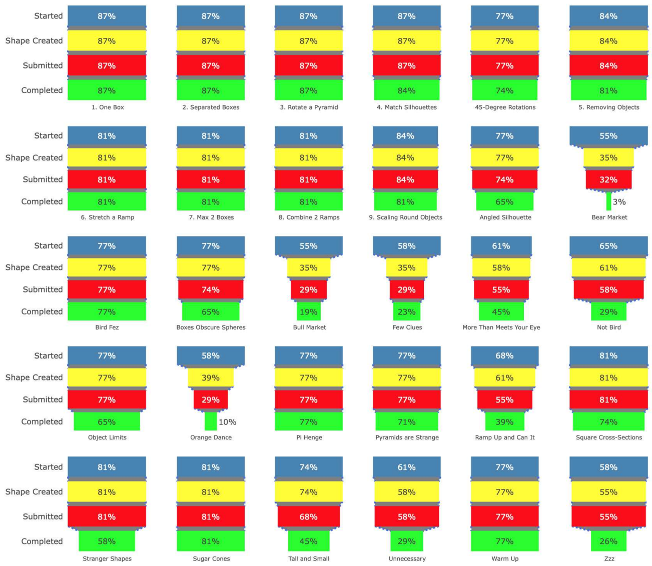


Fig. 7: Visualization of Funnel by Puzzle visualization for the selected class.

where the student at least created a shape.

VI. Case Study

In this section we present two use cases of how a teacher could use the live dashboard developed and its visualizations to monitor the overall classroom process (Subsection VI-A) but also individual students (Subsection VI-B).

A. Use Case 1: Classroom Analysis

This first use case exemplifies how the teacher can use the dashboard to analyze the global class status. This can be very useful to detect issues that are systematically affecting the whole group and be aware of the class's current progress. Many of the metrics can be connected as a part of a teacher's workflow, since the information that one provides can be helpful to contextualize another metric. To do this, we analyze some of the metrics that have been defined with specific examples using data from the real class that we are analyzing.

Figure 7 represents the Funnel by Puzzle for the selected class. In this visualization, the teacher can have a direct overview of the progress in all the puzzles. We observe puzzles where students have accomplished proper progress, e.g. "Bird Fez" which 77% of students have completed correctly. However, it can also be used to detect those puzzles where students are struggling, such as "Orange Dance" where we see that almost 60% of the students have started it, however only 10% have managed to finish it. This puzzle shows an obviously very low completion rate, and in order to delve into this issue, a teacher might want to further look into this issue, for example by looking into the Levels of Difficulty of such puzzle.

Figure 8 shows the Levels of Difficulty for each of the class puzzles. We can see the different parameters as specified in the metric definition. In the case of puzzle "Orange Dance," the metric indicates that it is the second one in terms of the difficulty level. Therefore, this helps contextualize the current situation of struggle with this puzzle.

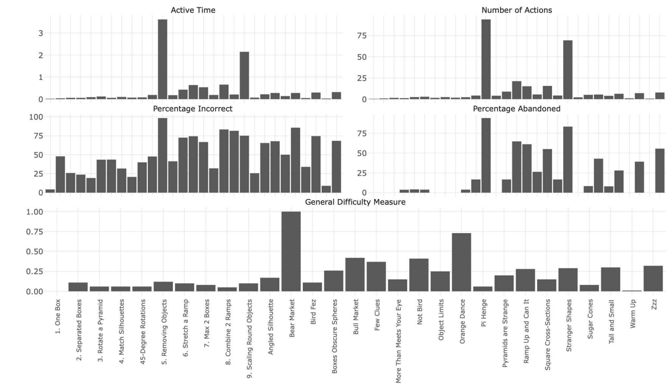


Fig. 8: Visualization of Levels of Difficulty for the selected class.

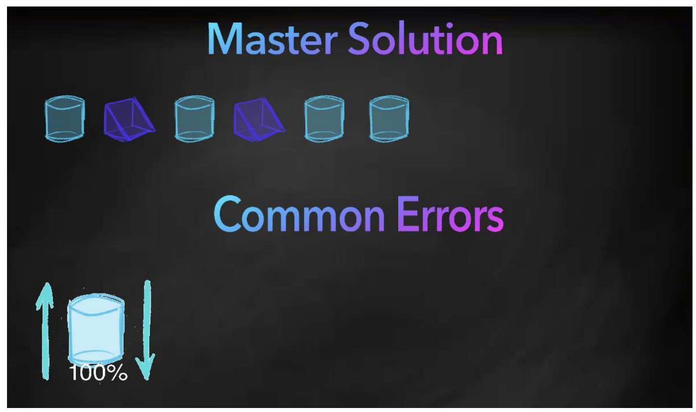


Fig. 9: Visualization of Common Errors for "Orange Dance" puzzle.

One final step that a teacher might want to take is to analyze the most Common Errors metric for the puzzle "Orange Dance," seeking to understand misconceptions solving the puzzle that can be addressed globally in a session for the entire class. Figure 9 shows the visualization for Common Errors, and it shows that the most common error detected is related to the position (movement) of the cylinder. This enables the teacher to adapt the session explaining in geometrical terms where the cylinder must be placed within the 3D environment to generate the views that solve the puzzle.

B. Use Case 2: Individual Student Analysis

In this use case, we present an individual student's analysis in a concrete classroom using individual-oriented visualizations to monitor that student and locate possible problems. First, we select a particular group of students in the classroom to see how they progress in the completion of the different game puzzles. We will then represent how a teacher can observe students' progress and difficulties at individual levels based on different metrics.

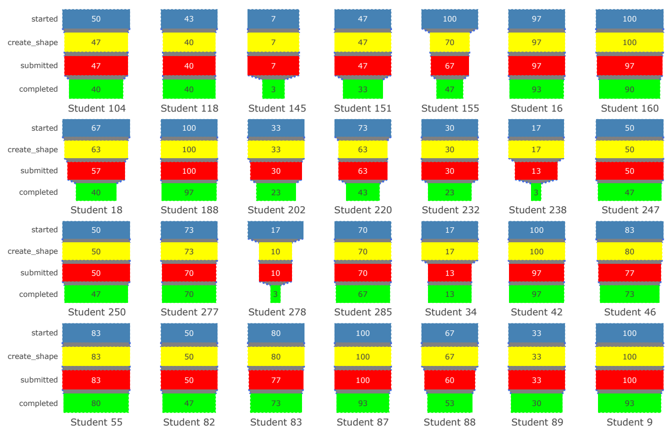
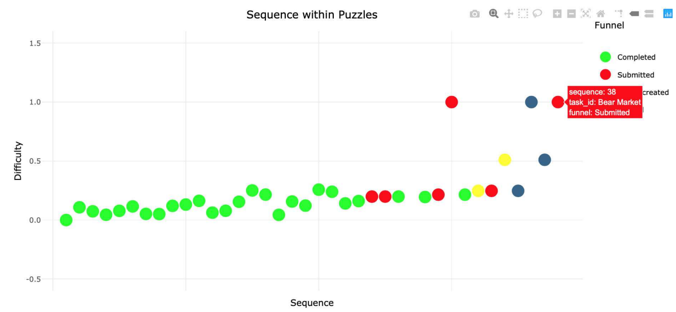


Fig. 10: Funnel by user visualization of the selected class.

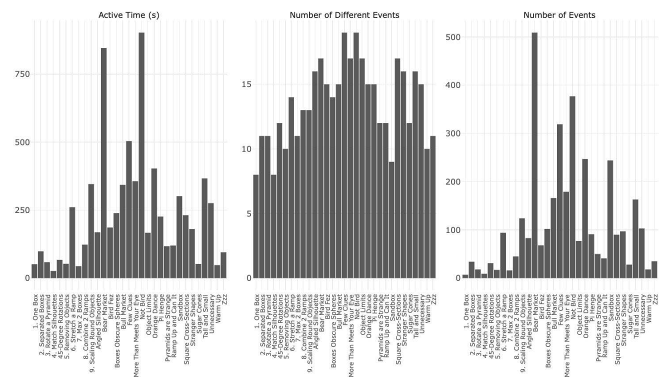
In Figure 10 we use a Funnel by User visualization, which is based on the Funnel by Puzzle metric we defined previously. In this concrete visualization, there are 28 students with a funnel corresponding to each student, which has the four different stages we defined in the funnel metric. The number inside the funnel represents the percentage of puzzles where that concrete student has reached that funnel stage. For example, we can focus on Student 160, which is a high performing student. This student has started all puzzles (100% of them) and completed correctly 90% of them. The teacher may want to know in which puzzles the student had difficulties. We can add a higher level of detail on students' progress with the Sequence Between Puzzles visualization, shown in Figure 11.1.

Figure 11.1 shows three different metrics at the same time: The x-axis with the dots represent the sequence of puzzles of the student, while the color of the dots represents the funnel stage of each puzzle, and then we have incorporated the difficulty metric of each puzzle by adjusting the position on the y-axis. The plot shows that almost every puzzle has been completed immediately, but the student has been having some problems solving one concrete puzzle, "Bear Market." In Figure 11.1 we also see an example of a tool-tip in the dashboard, showing the number of attempts in the sequence of puzzles, the name of the puzzle and the funnel stage reached. We can see that "Bear Market" has a 1.0 value in the difficulty metric. That indicates that this is the most challenging puzzle for this group. The student has tried to solve that puzzle three different times but failed.

Then, from Figure 11.2, we can draw some conclusions about student's interaction with the game. From the previous visualization, we know that puzzle level named "Bear Market" was submitted and then we see the `active_time` and `n_events` in this puzzle has been one of the highest of all puzzles. So we know the student has spent a significant amount of time trying to solve it, and we could now say that the student has experienced difficulties with this puzzle. The teacher could now want to know how the student has interacted with "Bear Market" and see if the actions make sense or the student has been acting randomly. In Figure 12.1



11.1 Sequence between puzzles.



11.2 Number of seconds spent and number of events performed.

Fig. 11: Sequence Between Puzzles and Levels of Activity for Student 160.

we can see the sequence of actions, denoting that the student has performed a large number of actions. As the student has made a significant amount of actions between submits, we know that the student has been mindfully trying to solve the puzzle, instead of making actions arbitrarily.

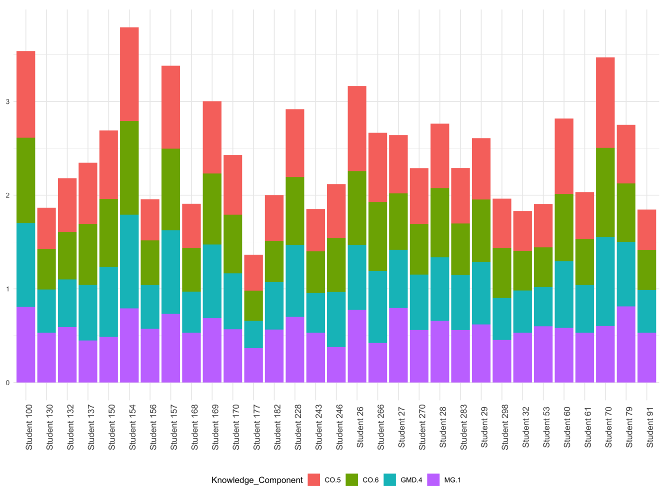
Finally, the teacher may want to know a summary of the competences acquired by the student. That summary of competences is shown in Figure 12.2. This visualization, corresponding to the ELO-based metric, shows a bar plot for each user, with the competency level of each knowledge component, showing that Student 160 has reached very high values for every competence. This overall analysis can show the teacher that this is a good performing student that has achieved a high competency in the game, and that has apparently only struggled with "Bear Market" puzzle, and thus a potential intervention for the teacher would be to clarify with the student how to solve this specific puzzle.

VII. Discussion

The overarching issue that this paper is tackling is the effective implementation of learning games in the classrooms [57]. Within this context, we believe that our contributions considerably expanded current knowledge on two directions: learning analytics metrics and dashboards for educational games.

First, in terms of learning analytics metrics, the objective was to build up metrics to support teachers by processing

12.1 Sequence Within Puzzles for "Bear Market" performed by Student 160.



12.2 ELO-based competency for the selected group.

Fig. 12

the raw data [35]. However, this metric design and definition process is especially challenging when dealing with games, as the open environments significantly broaden the spectrum of possibilities [58]. In order to overtake this issue, we have depicted a user-centered design process [38] targeting metrics that can support them implementing *Shadowspect*. This process has involved a multidisciplinary team that can bring diverse perspectives to the learning analytics design process [59]. Our methodology has resulted in seven metrics that go beyond the frequently found straightforward indicators [17, 33], by implementing nuanced learning analytics metrics with very clear application-oriented goals.

Second, in terms of the dashboard design, the objective was to build a learning analytics dashboard to represent the previously implemented metrics in simple yet powerful visualizations for teachers. We have depicted a process that was novel in two ways when compared with most previous literature. The first one is a vizards team that has been developing digital paper prototype visualizations before the implementation [60], and the second one is a co-design team working with K12 geometry teachers to co-design the final visualizations and refine metric definitions [39]. These contributions are aligned with the expected future of learning analytics dashboards, offering new design approaches [61]. Moreover, our approach is also aligned with current ideas to design translucent learning analytics with teachers [43], in order to make learning visible, improve awareness, and accountability; in that sense, one of the main novelties has been purposely designed icons aligned with *Shadowspect* game mechanics that facilitate this process.

While the implementation of our metrics and dashboard have been designed explicitly for *Shadowspect*, many of the depicted design guidelines, such as the ideation and design processes, architecture, prototyping ideas or interface design principles can be re-applied to other tools and across contexts. Therefore, future work can build on top of these ideas to develop their own learning analytics dashboards in educational games and other environments.

We believe that our work represents an important advancement for the implementation of learning games in the classroom, and to specifically support formative assessment using a game tech ecosystem [62]. In previous work, [63] noted out the potential of using learning analytics for assessment purposes, as it provides a multitude of information that the student can use to adapt the personal learning environment as much as possible to their own strengths and weaknesses, and this work goes into this direction. The implementation of such real-time tools provides valuable new information from students' interaction that educators can use to actively influence game activities to improve students' learning outcomes [64]. These contributions can improve the support of teachers using learning games, hence facilitating more effective implementations in the classroom [57]

VIII. Conclusions and Future Work

The objective of this research was threefold: first, to propose a series of metrics that can provide comprehensive information regarding the process of students with the puzzles in *Shadowspect*. Second, to implement a real-time dashboard with simple but detailed visualizations of these metrics that can allow teachers to track the students within their class, so that they can evaluate or detect problems quickly and effectively. Third and last, exemplify with uses case how this new approach represents an opportunity for educators to provide personalized attention to their students and help them in their learning process.

Teachers can do a live monitorization of their students during class, enabling just-in-time interventions that aim to provide support at the right time by adapting to each individual's needs. One of the main limitations of this work is that we do not yet conducted a proper validation. Thus, the use cases presented have not been made in collaboration with teachers. Part of our future work aims to deploy both *Shadowspect* and the dashboard with teachers in their classroom to evaluate the solution properly.

As a part of our future work, we will be developing new metrics to continue expanding the dashboard and its possibilities. More nuanced metrics and visualizations will allow students to visualize their mistakes and areas of improvement. Also, we will be working on obtaining evidences of the interpretability of these visualizations and to guarantee that they are explainable so that teachers can easily intervene. *Shadowspect* is designed as a formative assessment tool, and thus we can also use this dashboard for students so that they can receive feedback and improve their self-awareness. In this way, we can use *Shadowspect* as a robust learning tool that can be easily implemented by teachers in the classroom and that emphasizes the formative feedback to the student. This study has proposed a new dynamic approach that can be helpful to facilitate the systematic implementation of educational games in the classrooms of the future.

References

[1] George Siemens and Phil Long. Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5):30, 2011.

[2] Yara Maher, Sherin M Moussa, and M Essam Khalifa. Learners on focus: Visualizing analytics through an integrated model for learning analytics in adaptive gamified e-learning. *IEEE Access*, 8:197597–197616, 2020.

[3] R Raja and PC Nagasubramani. Impact of modern technology in education. *Journal of Applied and Advanced Research*, 3(S1):33–35, 2018.

[4] Sara De Freitas. *Learning in immersive worlds: A review of game-based learning*. Jisc, 2006.

[5] ESA. 2019 essential facts about the computer and video game industry. Technical report, Entertainment Software Association, 2019.

[6] Raph Koster. *Theory of fun for game design*. " O'Reilly Media, Inc.", 2013.

[7] James Paul Gee. Are video games good for learning? *Nordic Journal of Digital Literacy*, 1(03):172–183, 2006.

[8] John Kirriemuir and Angela McFarlane. Use of computer and video games in the classroom. In *DiGRA Conference*, 2003.

[9] Jos A. Ruiprez-Valiente and Yoon Jeon Kim. Effects of solo vs. collaborative play in a digital learning game on geometry: Results from a k12 experiment. *Computers & Education*, 159:104008, 2020.

[10] Katrien Verbert, Erik Duval, Joris Klerkx, Sten Govaerts, and José Luis Santos. Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10):1500–1509, 2013.

[11] Juha Sorva, Ville Karavirta, and Lauri Malmi. A review of generic program visualization systems for introductory programming education. *ACM Transactions on Computing Education (TOCE)*, 13(4):1–64, 2013.

[12] Alice Kaplan Gordon. *Games For Growth; Educational Games in the Classroom.* Science Research Associates, College Division, 1970.

[13] Richard Van Eck. Digital game-based learning: It's not just the digital natives who are restless. *EDUCAUSE review*, 41(2):16, 2006.

[14] Juho Hamari, David J Shernoff, Elizabeth Rowe, Brianno Coller, Jodi Asbell-Clarke, and Teon Edwards. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in human behavior*, 54:170–179, 2016.

[15] Renah Razzaq, Korinn S Ostrow, and Neil T Heffernan. Effect of immediate feedback on math achievement at the high school level. In *International Conference on Artificial Intelligence in Education*, pages 263–267. Springer, 2020.

[16] Muriel Algayres and Evangelia Triantafyllou. An educational model for integrating game-based and problem-based learning in data-driven flipped classrooms. In *International Symposium on Emerging Technologies for Education*, pages 145–154. Springer, 2019.

[17] A. R. Cano, . J. Garca-Tejedor, C. Alonso-Fernndez, and B. Fernndez-Manjn. Game analytics evidence-based evaluation of a learning game for intellectual disabled users. *IEEE Access*, 7:123820–123829, 2019.

[18] Dragan Gašević, Vitomir Kovanović, and Srećko Joksimović. Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. *Learning: Research and Practice*, 3(1):63–78, 2017.

[19] Cristina Alonso-Fernandez, Antonio Calvo, Manuel Freire, Ivan Martinez-Ortiz, and Baltasar Fernandez-Manjon. Systematizing game learning analytics for serious games. In *2017 IEEE Global Engineering Education Conference (EDUCON)*, pages 1111–1118. IEEE, 2017.

[20] Robert J Mislevy and Geneva D Haertel. Implications of evidence-centered design for educational testing. *Educational measurement: issues and practice*, 25(4):6–20, 2006.

[21] Yoon Jeon Kim, José A Ruipérez-Valiente, T Philip, R Louisa, and Eric Klopfer. Towards a process to integrate learning analytics and evidence-centered design for game-based assessment. In *Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference*, pages 204–205, 2019.

[22] Yoon Jeon Kim and Dirk Ifenthaler. Game-based assessment: The past ten years and moving forward. In *Game-Based Assessment Revisited*, pages 3–11. Springer, 2019.

[23] Cristina Alonso-Fernández, Ana R Cano, Antonio Calvo-Morata, Manuel Freire, Iván Martínez-Ortiz, and Baltasar Fernández-Manjón. Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior*, 99:301–309, 2019.

[24] Pedro A Martínez, Manuel J Gómez, José A Ruipérez-Valiente, Gregorio Martínez Pérez, and Yoon Jeon Kim. Visualizing educational game data: A case study of visualizations to support teachers. In *Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?* CEUR-WS, 2020.

[25] Boon-Seng Chew. An efficient framework for game-based learning activity. In *2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 147–150. IEEE, 2017.

[26] Manuel Ninaus, Katerina Tsarava, and Korbinian Moeller. A pilot study on the feasibility of dynamic difficulty adjustment in game-based learning using heart-rate. In *International Conference on Games and Learning Alliance*, pages 117–128. Springer, 2019.

[27] Meng-Jung Tsai, Li-Ju Huang, Huei-Tse Hou, Chung-Yuan Hsu, and Guo-Li Chiou. Visual behavior, flow and achievement in game-based learning. *Computers & Education*, 98:115–129, 2016.

[28] Manuel Freire, Ángel Serrano-Laguna, Borja Manero, Iván Martínez-Ortiz, Pablo Moreno-Ger, and Baltasar Fernández-Manjón. Game learning analytics: learning analytics for serious games. In *Learning, design, and technology*, pages 1–29. Springer Nature Switzerland AG, 2016.

[29] Sten Mäses, Bil Hallaq, and Olaf Maennel. Obtaining better metrics for complex serious games within virtualised simulation environments. In *European Conference on Games Based Learning*, pages 428–434. Academic Conferences International Limited, 2017.

[30] Jan L Plass, Bruce D Homer, Charles K Kinzer, Yoo Kyung Chang, Jonathan Frye, Walter Kaczetow, Katherine Isbister, and Ken Perlin. Metrics in simulations and games for learning. In *Game analytics*, pages 697–729. Springer, 2013.

[31] Jina Kang, Min Liu, and Wen Qu. Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72:757–770, 2017.

[32] Jose A Ruiperez-Valiente, Matt Gaydos, Louisa Rosenheck, Yoon Jeon Kim, and Eric Klopfer. Patterns of engagement in an educational massive multiplayer online game: A multidimensional view. *IEEE Transactions on Learning Technologies*, 2020.

[33] Kristian Kiili, Korbinian Moeller, and Manuel Ninaus. Evaluating the effectiveness of a game-based rational number training-in-game metrics as learning indicators. *Computers & Education*, 120:13–28, 2018.

[34] Bireswar Laha and Doug A Bowman. Identifying the benefits of immersion in virtual reality for volume data visualization. In *Immersive visualization revisited workshop of the IEEE VR conference*, pages 1–2, 2012.

[35] Cristina Alonso-Fernandez, Antonio Calvo-Morata, Manuel Freire, Ivan Martinez-Ortiz, and Baltasar Fernández-Manjón. Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, 141:103612, 2019.

[36] Stephen Few. Information dashboard design. *S. Few*, 2003.

[37] Yesom Yoo, Hyeyun Lee, Il-Hyun Jo, and Yeonjeong Park. Educational dashboards for smart learning: Review of case studies. *Emerging issues in smart learning*, pages 145–155, 2015.

[38] Simon Buckingham Shum, Rebecca Ferguson, and Roberto Martinez-Maldonado. Human-centred learning analytics. *Journal of Learning Analytics*, 6(2):1–9, 2019.

[39] Kenneth Holstein, Gena Hong, Mera Tegene, Bruce M McLaren, and Vincent Aleven. The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge*, pages 79–88, 2018.

[40] Jose A Ruiperez-Valiente, Pedro J Munoz-Merino, Jose A Gascon-Pinedo, and C Delgado Kloos. Scaling to massiveness with analyse: A learning analytics tool for open edx. *IEEE Transactions on Human-Machine Systems*, 47(6):909–914, 2016.

[41] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 257–266, 2017.

[42] Vanessa Echeverria, Roberto Martinez-Maldonado, Roger Granda, Katherine Chiluiza, Cristina Conati, and Simon Buckingham Shum. Driving data storytelling from learning design. In *Proceedings of the 8th international conference on learning analytics and knowledge*, pages 131–140, 2018.

[43] Roberto Martinez-Maldonado, Doug Elliott, Carmen Axisa, Tamara Power, Vanessa Echeverria, and Simon Buckingham Shum. Designing translucent learning analytics with teachers: an elicitation process. *Interactive Learning Environments*, pages 1–15, 2020.

[44] Rita Prestigiacomo, Roger Hadgraft, Jane Hunter, Lori Locker, Simon Knight, Elise van den Hoven, and Roberto Martinez-Maldonado. Learning-centred translucence: an approach to understand how teachers talk about classroom data. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 100–105, 2020.

[45] Gayane Sedrakyan, Jonna Malmberg, Katrien Verbert, Sanna Järvelä, and Paul A Kirschner. Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107:105512, 2020.

[46] Gayane Sedrakyan, Erik Mannens, and Katrien Verbert. Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50:19–38, 2019.

[47] Laure France, J-M Heraud, J-C Marty, Thibault Carron, and Joseph Heili. Monitoring virtual classroom: Visualization techniques to observe student activities in an e-learning system. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pages 716–720. IEEE, 2006.

[48] Vili Podgorelec and Saša Kuhar. Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, 114(8):111–116, 2011.

[49] Shamus P Smith, Daniel Hickmott, Erica Southgate, Ross Bille, and Liz Stephens. Exploring play-learners' analytics in a serious game for literacy improvement. In *Joint International Conference on Serious Games*, pages 13–24. Springer, 2016.

[50] Ángel Serrano-Laguna, Iván Martínez-Ortiz, Jason Haag, Damon Regan, Andy Johnson, and Baltasar Fernández-Manjón. Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, 50:116–123, 2017.

[51] Mohamed Amine Chatti, Arham Muslim, Mouadh

Guesmi, Florian Richtscheid, Dawood Nasimi, Amin Shahin, and Ritesh Damera. How to design effective learning analytics indicators? a human-centered design approach. In *European Conference on Technology Enhanced Learning*, pages 303–317. Springer, 2020.

[52] Solmaz Abdi, Hassan Khosravi, Shazia Sadiq, and Dragan Gasevic. A multivariate elo-based learner model for adaptive educational systems. *International Educational Data Mining Society*, 2019.

[53] Radek Pelánek. Applications of the elo rating system in adaptive educational systems. *Computers & Education*, 98:169–179, 2016.

[54] Matt Makai. Celery, 2019. [Web; last access on 12-07-2020].

[55] Christa Kelleher and Thorsten Wagener. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, 26(6):822–827, 2011.

[56] Cynthia A Brewer. Color use guidelines for mapping. *Visualization in modern cartography*, 1994:123–148, 1994.

[57] Meihua Qian and Karen R Clark. Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63:50–58, 2016.

[58] Laila Shoukry, Stefan Göbel, and Ralf Steinmetz. Learning analytics and serious games: Trends and considerations. In *Proceedings of the 2014 ACM International Workshop on Serious Games*, pages 21–26, 2014.

[59] Dan Suthers and Katrien Verbert. Learning analytics as a "middle space". In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pages 1–4, 2013.

[60] Yvonne Vezzoli, Manolis Mavrikis, and Asimina Vasalou. Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. In *Proceedings of the Tenth international conference on learning analytics & knowledge*, pages 73–82, 2020.

[61] Katrien Verbert, Xavier Ochoa, Robin De Croon, Raphael A Dourado, and Tinne De Laet. Learning analytics dashboards: the past, the present and the future. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 35–40, 2020.

[62] Malcolm Bauer, Caroline Wylie, Tanner Jackson, Bob Mislevy, Erin Hoffman-John, Michael John, and Seth Corrigan. Why video games can be a good fit for formative assessment. *Journal of Applied Testing Technology*, 18, 2017.

[63] Dirk T Tempelaar, André Heck, Hans Cuypers, Henk van der Kooij, and Evert van de Vrie. Formative assessment and learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 205–209, 2013.

[64] Miroslav Minović and Miloš Milovanović. Real-time learning analytics in educational games. In *Proceedings of the first international conference on technological ecosystem for enhancing multiculturality*, pages 245–251, 2013.

José A. Ruipérez-Valiente (Senior Member, IEEE) received the B.Eng. degree in telecommunications from Universidad Catlica de San Antonio de Murcia, and the M.Eng. degree in telecommunications and the M.Sc. and Ph.D. degrees in telematics from Universidad Carlos III of Madrid while conducting research with Institute IMDEA Networks in the area of learning analytics and educational data mining. He was a postdoctoral associate with MIT. He has received more than 20 academic/research awards and fellowships, has published more than 70 scientific publications in high impact venues, and participated in over 18 funded projects. He currently holds the prestigious Spanish Fellowship Juan de la Cierva with the University of Murcia.

Manuel J. Gomez received his B.Sc. in Computer Science from the University of Murcia, Spain, specialized in applied computing and data science. He completed his last academic year with an end-of-degree project based on Learning Analytics and Data Science applied to educational games. He is a M.Sc. Student in Big Data and also a member of the CyberDataLab at the University of Murcia working as a research intern.

Pedro A. Martínez received his B.Sc. in Computer Engineering at the University of Murcia, specialized in applied computing and data science. He is currently studying the MSc on New Technologies in Computer Science specialized in intelligent and knowledge-based technologies with applications in medicine. He is a member of the CyberDataLab of the University of Murcia where he is working as a research intern.

Yoon Jeon Kim is a research scientist at MIT Open Learning. Her work has centered on the topic of innovative assessment and how technological advancement influences what we are measuring about student learning and how we are measuring it. Before joining MIT, she was involved in many game- and simulation-based assessment development and evaluation projects using evidence-centered design in combination with educational data mining techniques. Currently, she leads the Playful Journey Lab, where she and her colleagues explore playful, authentic, yet rigorous assessment approaches in both digital and non-digital learning environments.