Ruipérez-Valiente J.A., Rosenheck L., Kim Y.J. (2019) What Does Exploration Look Like? Painting a Picture of Learning Pathways Using Learning Analytics. In: Ifenthaler D., Kim Y. (eds) Game-Based Assessment Revisited. Advances in Game-Based Learning. Springer, Cham

DOI: https://doi.org/10.1007/978-3-030-15569-8_14

Chapter # - will be assigend by editors

WHAT DOES EXPLORATION LOOK LIKE?: PAINTING A PICTURE OF LEARNING PATHWAYS USING LEARNING ANALYTICS

José A. Ruipérez-Valiente, Louisa Rosenheck and Yoon Jeon Kim. Massachusetts Institute of Technology

Abstract: Game-based learning is becoming one of the major trends in education as it brings together numerous benefits. However, due to the open-ended and less linear nature of these environments, it is often complicated for instructors to really understand the learning process of students within a game. Learning analytics can play a meaningful role in transforming learning pathways in games into interpretable information for teachers. In this study, we propose three novel metrics that focus more on the learning process of students than on the outcomes. We apply these metrics to data from *The Radix Endeavor*, an inquiry-based learning game on STEM topics that has been tested in multiple schools across the US. We also report correlations between these metrics and in-game learning outcomes and discuss the importance and potential use of metrics to understand students' learning processes.

Keywords: Game-based learning, learning analytics, behavioural modelling, learning pathways

1. INTRODUCTION

To prepare students for success in our ever-changing knowledge economy, learning and teaching is moving toward valuing future-ready skills, also called 21st century skills or soft skills, including skills like problem-solving, interpreting information, and communication. Educational games and simulations are one important tool for building these kinds of skills. They can provide open-ended but scaffolded experiences in which students can test out ideas in a low-stakes environment, retrying levels or challenges until they succeed. However, one challenge in teaching interpersonal skills in general, and teaching them through games in particular, is that these skills are much harder to assess than the content knowledge and procedural skills that have been valued by schools in the past. While digital games for learning do have the affordance of being able to collect large amounts of very nuanced activity data, the most common types of analysis and reporting have most often focused on things like content-specific failures and successes, percent of correct attempts, and progress through the game. Dashboards or tools that count these achievements for each student and output certain metrics for teachers to track progress are a helpful starting place, and certainly convenient for teachers. However, richer metrics of game-based learning have the potential to not only show what students understand and can do, but also provide much deeper insights into patterns in how they are approaching problems and illuminate the different learning pathways they take. Learning analytics techniques that use generalizable methods but that are tailored to the specific game mechanics and assessment mechanics of a given game can play a key role in mapping out the learning pathways that students take and characterize the ways they are engaging with the game. As such, this chapter will use examples of measurement and data analysis methods from *The Radix Endeavor* (shortened as *Radix* from hereon), a multiplayer online STEM game, to demonstrate ways that we can understand not only *what* students are learning through games, but also *how* they are going about learning it. The main research questions for this study are:

1. How can we use activity data metrics to characterize exploration in digital learning games?
2. What specific methods of analysis can we implement to understand students' learning processes?

In this chapter we will start with a literature review of learning games and learning analytics to provide background for this study. We will describe the Radix game, pilot implementation, and data collection. We will then present three metrics that were designed and implemented—quest progression linearity, quest event focus, and time per quest—and describe the patterns that were found by applying these metrics to Radix pilot data. Finally, we will explain potential ways these metrics could be interpreted and used in a classroom setting, and discuss the broader impact that richer data on learning pathways could have on game-based learning.

2. LITERATURE REVIEW

Learning does not only involve acquiring content knowledge but also involves the development of supporting motivation and the establishment of skills and ways of thinking. In the past, this has not been easy to realize in traditional classroom settings due to physical and time constraints. However, research over the past decade has revealed that digital educational games can support meaningful and authentic learning, in deeper ways than more conventional forms of teaching (Connolly, Boyle, Macarthur, Hainey, & Boyle, 2012; Papastergiou, 2009; Vogel et al., 2006; Wouters, Van Nimwegen, Van Oostendorp, & Van Der Spek, 2013). When they play video games, people practice a set of 21st century skills that can be applied to studying and working in the real world (Prensky, 2006). Gee (2003) also argues that gaming has the potential to increase the impact and effectiveness of the work of individuals by bringing about synchronized intelligence, where humans and digital tools complement each other's abilities in order to achieve new goals. In addition, the National Research Council has reported that in the field of science learning in particular, simulations and games have the potential to advance multiple learning goals including conceptual understanding, science process skills, and discourse and argumentation (Honey & Hilton, 2011). Research into learning games has also revealed some of the game elements that best enable content also inspire interest, creativity, and social interaction (Squire, 2011). These value-added features that are specifically designed to support learning have in fact been found to magnify learning (Clark, Tanner-Smith, & Killingsworth, 2016). All of this evidence explains why the Joan Ganz Cooney Center's Level Up Learning survey reported that 74% of teachers are currently using digital games for instructional purposes with their students (Takeuchi & Vaala, 2014).

One tool that can help make sense of players' actions in open-ended digital games for learning is learning analytics (Berland, Baker, & Blikstein, 2014). Over the last decade the production of data has expanded at a stunning fast pace. In education, multiple virtual learning environments have been emerging, such as MOOC platforms, games for learning, intelligent tutoring systems and more. To analyze all these data we need a combination of theory, design and data mining techniques, and in order to fulfill these requisites the field of Learning Analytics (LA), an intersection between data science and learning sciences (Gašević, Kovanović, & Joksimović, 2017), has been gaining a lot of attention over the last years. The data analysis of these huge data samples has immense potential for the field. Nonetheless, LA should focus on the learning process, and therefore it also should be situated within the existing framework of educational research (Gašević, Dawson, & Siemens, 2015). In the field of education, learning analytics has explored numerous questions, such as course attrition (Kloft, Stiehler, Zheng, & Pinkwart, 2014), predicting the success of a student in a degree for admission purposes (Nghe, Janecek, & Haddawy, 2007), to predict if students are going to surpass a course or not (Calvo-Flores, Galindo, Jiménez, & Pineiro, 2006), to generate recommendations about learning resources in educational systems (Salehi & Nakhai Kamalabadi, 2013) or to predict the final grade of a student in a test (Pardos, Gowda, Ryan, & Heffernan, 2010).

Learning analytics can be especially useful in open-ended environments (like games) where the freedom of interaction is much higher (Blikstein, 2011). Therefore, the self-regulated strategies of learners and methods to measure those start to become more important in such environments (Segedy, Kinnebrew, & Biswas, 2015) and the use of optional activities in self-regulated environments have been found to have a relationship with learning outcomes (Ruiperez-Valiente et al., 2016). In the context of learning analytics applied to games, there have been new approaches in the past years, showing the challenge of developing new psychometric models for those environments (Gibson & Clarke-Midura, 2015) and one of the major goals is being able to use learning analytics for a trustworthy assessment of students' knowledge (Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 2012). Some previous work has shared ideas similar to ours in other contexts such as to measure the focus on actions to earn badges (Ruipérez-Valiente, Muñoz-Merino, & Delgado Kloos, 2017) or the linearity of students' following the recommendations of a system in online learning (Ruipérez-Valiente, Muñoz-Merino, Leony, & Delgado Kloos, 2015).

3. BACKGROUND

The Radix Endeavor is an inquiry-based online game for STEM learning developed at the MIT Education Arcade. It is an MMO-style game set in a virtual multiplayer world that is fairly open-ended and exploratory but that has set sequences of tasks for players to work through. The Radix world contains embedded biological and mathematical systems that involve the world's realistic but fictional flora, fauna, and civilizations. Players take on game tasks, or quests, that guide them to probe the game's systems and develop a firsthand understanding of math and biology concepts in a variety of topic areas. The game is exploratory, leaving a lot of experimenting and problem-solving up to the players. It incorporates a wide variety of content as well as STEM practices and soft skills. It is a long-form game, meant to be played over the course of a semester and revisited during each relevant curricular unit. In addition, it presents opportunities for players to collaborate both in and outside of the game, leading to a unique deep learning experience.

Figure 1. Screen shots of tools used in Radix quests.

When players enter the game for the first time, they begin a sequence of tutorial quests designed to get players used to moving around the world, using tools, and collecting data about their environment. Upon completion of the tutorial quest line, an array of topical quest lines is unlocked, including four in biology:

genetics, ecology, evolution and human body systems; and three in math: geometry, algebra, and statistics. While the quests are sequenced within a topic area, players are free to switch between quest lines according to their interests throughout their play sessions. Each quest line may have anywhere from four to ten quests within it, and each quest is made up of multiple smaller tasks which provide some scaffolding to players. The quest content is aligned with curriculum standards and the tasks are specific to the domain. For instance, in one of the genetics quests players must figure out how dominant and recessive traits work in order to breed non-toxic *glumbugs* for a chef to use in his cooking. In the algebra quest line, players explore a marketplace where they must barter with vendors who offer different rates of trade, using unit conversion concepts to maximize the *zorbits* they earn on the exchange (see Figure 1). In order to make sense of the in-game systems and complete their tasks, players have a number of tools at their disposal. Some tools are useful across quest lines, and some are domain-specific, but all tools are accessible at all times, regardless of the quest a player is currently working on. This design means that one of the skills players are practicing is selecting the tool that will be the most helpful or efficient to solve a problem. For example, the trait examiner and trait decoder let players identify the phenotypes and genotypes of a species, and the breeding station lets them breed plants and animals. These are most useful in genetics challenges, whereas the data library, which lets them do simple analyses of means and distributions, can be useful in a number of math and biology quests.

The specific interactions and problems presented in the quests are unique to a topic area in order to provide an environment where players are engaging in authentic inquiry. At the same time, there are elements of quest design that are consistent across quest lines and that are important to the game's pedagogical approach. Quests are introduced in context, to present an authentic problem in the fictional world. Players know generally what they need to do, but they are not told exactly what steps to take to solve the problem or which tools to use. They need to experiment with the systems to build some content understanding, usually iterating on their strategy based on the feedback they get from the game. When they turn in a quest, or present the solution, they are asked to not only hand in a game object or artifact, but also explain their reasoning or back up their claims. For example, along with the non-toxic bugs they must also create a Punnett square that shows which parents will breed the desired offspring. There is no penalty or disincentive for submitting incorrect solutions. Rather, players get some feedback and are invited to continue experimenting or try a new approach. This type of quest design is meant to create an inquiry experience where players explore and build their own knowledge in a low-stakes environment. This provides an opportunity for players to build and demonstrate skills such as creative problem-solving, experimentation, and supporting claims with evidence. It also provides an opportunity for designers and educators to recognize those skills and assess progress in their development.

The designing of specific game elements in Radix with the goal of generating evidence of learning was one of the project's goals and research questions from the start. We aimed to create a digital environment for inquiry learning that could provide feedback to both players and teachers about how players are approaching problems, using tools, and building conceptual knowledge in math and biology. These are skills that are difficult to measure with traditional tests, and we wanted to research how well a digital game could collect telemetry data for an embedded assessment approach. For example, quest tasks were designed to provide opportunities for players to build and demonstrate their inquiry skills. Game data was collected for actions relevant to quests and exploration, and that data was interpreted to provide teachers with feedback on what their students were struggling with. The feedback mechanisms were only built out for an initial level of two quest lines. In this study, we apply learning analytics techniques to form the basis of the next level of measurements that could be conducted around Radix gameplay to provide insights for how players approached the quests rather than simply describing what they were able to achieve or perform in the game.

4. METHOD

4.1 Pilot Description and Context

Radix launched as a free tool available across the US and internationally in late January 2014 and has been played in all 50 states and at least 7 different countries. The dataset used in this study was collected during the pilot period which ran through August 2015. While the game was designed with high school math and biology teachers in mind, Radix has been used by upper elementary, middle, and high school teachers as well as by a few instructors at community colleges and universities. Outside of the formal school environment, the game has also been picked up by various after school groups, enrichment programs, and the homeschool community who are using it with a wide variety of ages. During the pilot period, informal marketing and outreach was done to recruit teachers to participate in the pilot at various levels. This included reaching out to local and national teacher networks to publicize the game, as well as a number of press articles and blog posts showcasing the project and its opportunities for participation. Teachers created accounts for their students to play, but players who heard about the game via other channels were also able to create player accounts not associated with a school or teacher. Participating teachers were provided with some professional development opportunities and implementation resources but they were encouraged to tailor their implementations and use the game as they saw fit in their classroom. Most of them had their students play relevant quest lines at the time they were covering a given topic area in their class. Outside of school players naturally played as much or little as they chose to, working through quest lines according to their interests.

4.2 Data Collection

We used the data from the pilot study that we described in previous Subsection 4.1. The design of Radix emphasized a rich data infrastructure that could allow researchers to perform detailed analytics of students' interactions. Radix has a relational database with more than 20 tables that collects most of the interactions of students with the game, such as player metadata, tool usage, quest related events, or social interactions. As part of this study, we develop an algorithmic machinery that processes such data to create interpretable information such as the metrics that we present.

The data set includes over 14,000 Radix accounts; however, some of these accounts were not activated or barely interacted with the game. We therefore included only those accounts that were active within the game for at least one hour. With this filter, the number decreases to 5,493 accounts with 5,532 virtual characters. From the total, 4,841 (87.5%) of the characters were student accounts created as part of the pilot studies in schools and 691 (12.5%) of the characters were created by other online users. Some of these characters used Radix for over 22,000 hours, generated more than 1 million events, completed more than 68,000 quests and sent more than 60,000 social chat messages.

4.3 Data Analysis and Metrics

This study focuses on the processes students use to solve quests within the game, and then ultimately connects these metrics with an in-game learning outcome such as the percentage of correct responses. We have defined three brand-new metrics based on process mining techniques to investigate how students are interacting with a learning environment that is very open-ended and presents numerous possibilities and choices within the learning process of each student. The three metrics that we define are as follows:

- **Quest progression linearity**: This metric takes advantage of the multiple quest chains available in Radix as described in Section 3. Since students are free to jump from one quest chain to another, we investigated this issue by computing a percentage of quest chain changes by each student when they are still able to progress further within the current quest chain. For instance, if a student finishes quest GN1.1, from the Genetics topic which is part of the GN1 quest chain, and then completes ST1.1 which is part of Statistics ST1 quest chain, that would count as a quest chain change. However, if the student finishes GN1.8, which is the last quest of GN1 quest chain, and then completes ST.1.1, that would not count as a quest chain change since GN1.8 was the last quest of that quest chain and the student is forced to switch to a new one. Then, we computed a percentage as follows:

$$100 * (number\ of\ quest\ chain\ changes)\ /\ (number\ of\ quests\ completed)$$

- **Quest action focus**: Each of the quests of Radix is designed to be solved using experimental approaches by using specific tools to answer questions. Often, students will need to explore a bit before they are able to understand the requirements of the quest, what tools they need to use and how. In this metric, we explore the percentage of events that each student completes before solving a quest, which of these are strictly related to quest events, and which of them were not explicitly necessary to solve the quest (such as other tool events or social actions). Then, we computed a quest action focus percentage as follows:

$$100 * (action\ events\ related\ to\ quest)\ /\ (total\ events\ before\ quest)$$

- **Time per quest and average time difference per quest**: Since the path to solve each quest is not obvious once they receive the task, it might need exploration, experimentation and extra time depending on the strategy and knowledge of each student. Additionally, each quest might have a potentially different difficulty or length. Therefore, exploring the time required to solve each quest provides the potential to understand students' process and game dynamics. We computed the quest completion time between the acceptance of the quest and the quest being completed, omitting any times when the user was not interacting with the game. Since each student might resolve different quests, and each quest might potentially need different efforts, computing an average time per quest for each student would be biased by the quests that they completed. Therefore, to generate an informative per student metric, we computed the average time per quest, and then used the time spent by student in that quest to calculate the difference and compute an average time difference per quest for each student. This way, we can finally present an average number per student that indicates how many minutes faster or slower they are solving quests compared to the rest of students:

$$\sum_{i=1}^{Q} t_{j,i} - a_i$$

where Q is the number of quests in Radix, $t_{j,i}$ would be the time t to complete quest i by player j and a_i would be the average time for all players to complete quest i.

Since the Radix world is so open, we acknowledge that we cannot be completely accurate about measuring if student actions are devoted to finishing one quest or not; therefore some of these metrics represent an approximation of the ground truth. We explore these metrics at a student level, but also as global dynamics defining the Radix ecosystem, which can be useful for game design and understanding complex behaviors in open-ended game environments.

5. RESULTS

The three first subsections of the results describe the global dynamics of each metric whilst the fourth subsection connects together the three metrics at a student level with joint visualizations and correlations.

5.1 Quest Progression Linearity

To illustrate the dynamics of the quest system in Radix, we created a global graph of the most typical quest pathways followed by students, using Gephi (Figure 2). The graph was constructed by creating edges between quest completions. For example, if 10,000 students completed quest TUT1.1 and then went on to complete TUT1.2, that value would represent the weight between nodes TUT.1.1 and TUT.1.2. Then, we used the thickness of the edge (line) to encode this weight, showing the frequency at which students followed this path. Additionally, we used the size of the node/label to codify the centrality of the node within the network. Finally, the label represents the quest ID and the color represents the topic of the quest.

The global dynamics of the quest ecosystem are very clear in Figure 2. For example, we can see a high centrality for TUT1.1 quest, since is the first quest available in Radix, and then for EV1.1, GN1.1, EC.1.1, GM1.1, ST1.1, AL1.1 and HB1.1, as they are the first quest in each quest chain (or topic) and unlock after players solve the first tutorial quests. Moreover, we can see how the layout algorithm has grouped quests from the same topic close together based on the weight influence, which denotes that students usually solve quests from the same topic without jumping around. Additionally, we can see thicker edges between consecutive tasks of a quest chain, for example, TUT1.1, TUT1.2 until TUT1.7 which indicates that students generally solve consecutive quests from the same quest chain. These results are tightly coupled with the design and implementation of the game. The way the quests are presented in the game leads players sequentially through a quest line, although it doesn’t force them to complete tasks in the given order. In addition, many teachers who used Radix in class specified a particular quest line, encouraging students to focus on that topic area.

While Figure 2 explores the global dynamics, the individual learning path of each student can be completely different. Therefore, to illustrate this idea we present in Figure 3 two student examples, one that follows a highly linear quest progression and a second one that has performed frequent quest chain jumps during his/her learning process. Student A represents a very linear quest progression: the student completes consecutive quests from each quest chain, and only changes to a new quest chain after finishing the current one, changing quest chains only 4% of time upon completing an individual quest. On the opposite end of the spectrum, Student B advances by frequently jumping between quest chains; more exactly, upon completing a quest, 60% of time they changed quest chains. We will delve into the significance of how these different behaviors and strategies might influence learning outcomes and other metrics later in this paper.

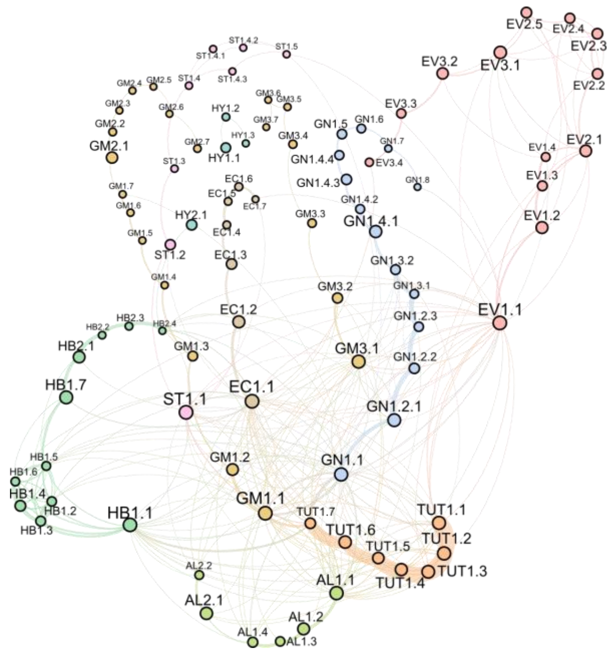

Figure 2. Graph network represent the global dynamics of the quest system.

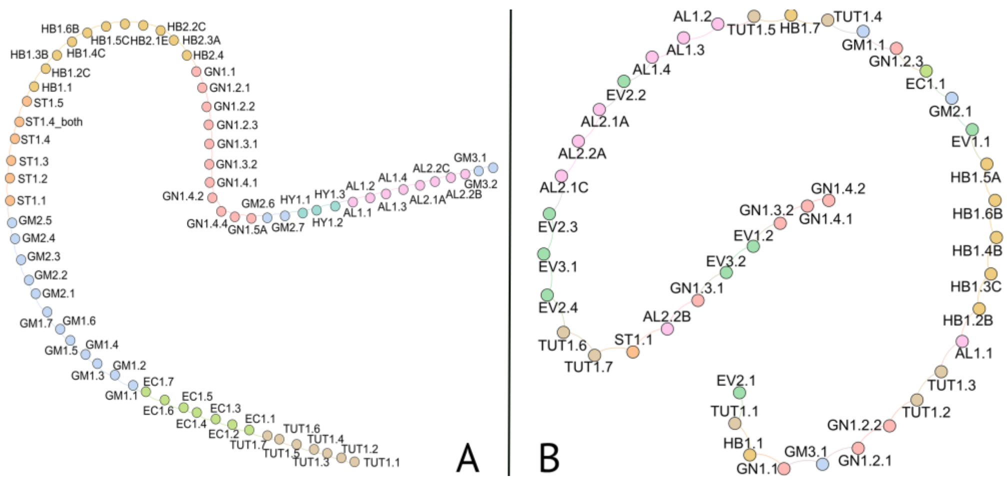

Figure 3. Graphs representing one student with a high level of progression linearity (A) and one with a low (B).

5.2 Quest Action Focus

Each quest of Radix is designed to require some degree of inquiry and exploration with the environment to understand the task and complete the requirements of the quest. Each quest chain focuses on specific STEM content that students will learn through experimentation with the environment by using specific scientific tools. Therefore, each quest chain is associated with a set of tools that allow students to gather the evidence to complete the task. However, students might need to experiment with the different tools before they have a clear sense of which tools are appropriate to the task and how to solve it. For example, in the GN1 (Genetics) quest chain students interact with the flora and fauna of Radix to learn how to breed for certain phenotypes, and then apply those skills to help villagers to cure diseases and more. Students have to identify the relevant genotypes and use Punnett squares to teach villages how to breed the traits of the flora and fauna themselves. To accomplish these tasks, they need to use four tools: the trait decoder and examiner, the Punnett square and the breeding station. The rest of the available tools or actions are not necessary for the GN1 quests, but students might experiment with them while they work to solve the quest.

That is why, as we can imagine, the dynamics in terms of events and actions of each quest should be different, and we use that fact to compute this quest action focus measure for five of the main quest chains AL1, GN1, GM2, EV1 and ST1, and Figure 4 shows a graph with the dynamics of four of them. Analogously to the previous section, we generated graphs to explore the dynamics of each quest chain separately, by creating an edge between each event generated by students before completing any of the quests from its chain. Again, the thickness of the edge encodes its weight, and the size of the node/label the centrality of the node within the network. The color encodes the type of node, and we use green for the use of tools related to quest requirements, orange for other tools, yellow for quest events, blue for party events and white for chatting events. Note that which tools are coded as a green or orange node should change from quest chain to quest chain according to the requirements of each quest chain.

There are a number of interesting things to note from the dynamics of each quest chain in Figure 4. First, we can see how the most central part of the network are always those actions related to the quest (green) and quest events (yellow), whereas other action events not related to the quest (orange) still show in the network but form subgroups as part of periphery. Interestingly, the periphery groups are similar in each graph; for example, we see the subgroup formed by the events "triangle use," "glass cutting reset," "window viewer use" and "window viewer reset," which represents the set of actions required to complete GM1. This might represent the behavior of jumping from one quest chain to another, and that is why the global dynamics of each quest chain capture these subgroups as well. The outer glow that some of the nodes have, for example traded items in AL1, represents the self-loop degree of an event, hence the thick self-loop of traded items in AL1, would mean that trading items consecutively was a very common two-gram sequence. This allows us to identify the main tools of each one of the quest chains. Finally, the high degree of centrality of chat events (white) that show that the social component is highly interspersed between quest actions.

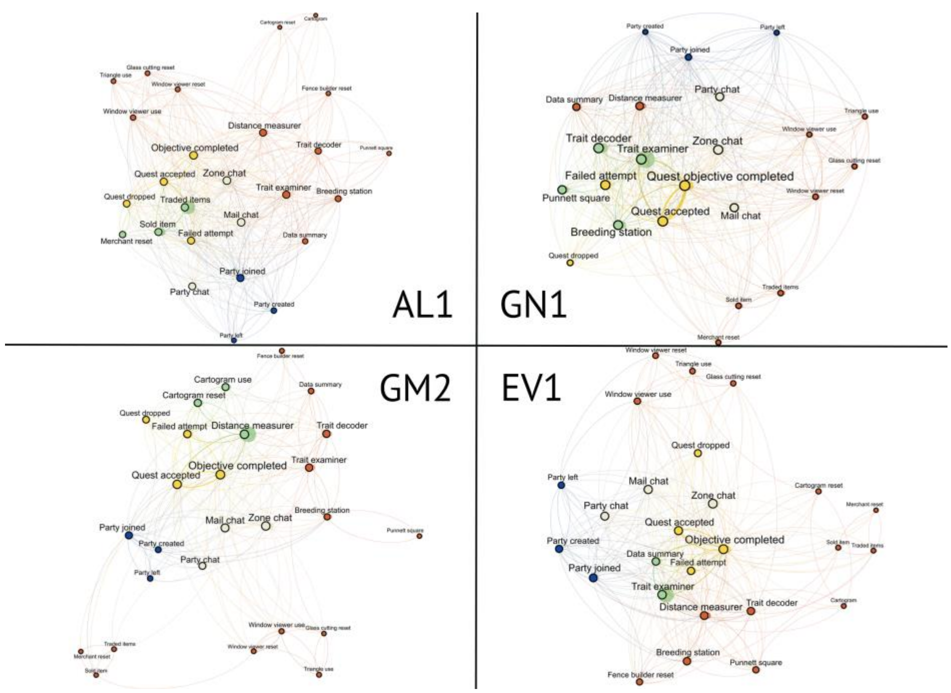

Figure 4. Graphs of action dynamics per quest chain.

5.3 Time per Quest

The last metric that we have proposed targets the time to resolve each quest of Radix. The algorithm gathered all events with timestamps that before the quest completion event, back to the last quest completion event and computes the effective active time to complete each quest. We can analyze this metric at both the quest and student level. At the student level, we can see a student's "efficiency" in quest completion by comparing the amount of time or events needed to that needed by other students. At the quest level, we can see approximate the amount of effort required to solve a quest by looking at whether it requires on average more or less time or events than other quests. Additionally, the percentage of correct solutions suggests the degree of difficulty of a quest.

In Figure 5, a boxplot visualization with the distribution of the time and percentage of correct responses per quest appears on the left. As a summary, we can see that the median of percentage correct is around 62% with a high variance, which denotes that there are some quests with very low correct ratios. The median time required per quest is 8.8 minutes, so generally quests do not take much time to complete, but we can see numerous outliers in the upper (more time) part of the distribution. For example, the quest EV1.1. (Evolution)—in which students need to explore to find the typical characteristics of *menjis* (a type of animal in the world of Radix)—required, on average, 110 minutes, but had a high correct response rate of 80%. Another example of the same quest chain would be EV1.4, where students have to make sense of *menjis* characteristics by responding to some questions, took on average 4.8 minutes but had a correct response rate of only 25%.

These dynamics can change a lot from one quest topic to another, so in Figure 5, each topic is broken out in an analogous visualization on the right. Geometry and Algebra have the highest, and Human Body has the lowest correct percentage ratio and lowest time per quest, which is likely due to the fact that the response method of this quest chain consisted of multiple choice questions, so students might have been using trial-and-error to guess the correct response. The quest topic that required the highest average time is Evolution, which is likely due to that domain requiring travel to different zones and data collection from a number of animals.

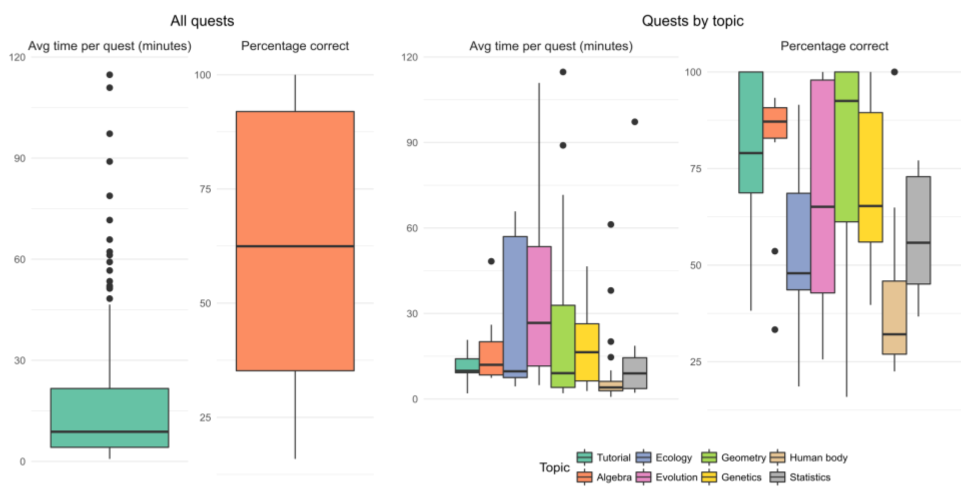

Figure 5. Boxplot distribution of the metrics by quest and by topic.

5.4 Distribution of Metrics by Student and Correlations

The previous subsection presented some global dynamics of the metrics that we have explored and this subsection reports the distribution of the individual metrics at the student level with correlations. Figure 6 shows a boxplot with the distribution of each one of the metrics per student. The first item shows the quest progression linearity with a median of 13.4% and mean of 16% of quest chain changes. Therefore, we see that generally speaking, students carry out a very linear learning path, following quest chains and infrequently switching between quest chains. We do see some outliers, with more than a 50% change of quest chains as discussed in Subsection 5.1., but this behavior does not represent the norm. The quest action focus has a median of 80% and mean of 65%, which represents that when working on a quest, the average student shows a high focus on quest-related events. Surprisingly, for the events that do not belong to quest actions, an average of 14% of events were "other action events" (i.e. tools not related to the quest) and 20% were social events (e.g. sending any kind of chat messages). However, as we saw in the global dynamics, many of the students are socializing within the game while resolving the quests. The third metric represents the time difference per quest, and is thus a measure of how fast students solve quests in comparison to their peers. The median is -6.5 and mean -4.6 minutes, but more importantly, it shows numerous outliers that indicate some students solving quests much faster or slower than the average. Finally, the percentage of correct responses has a mean value of 70%, and again we see a moderate variance that represents students with a higher or lower percentage of correct responses.

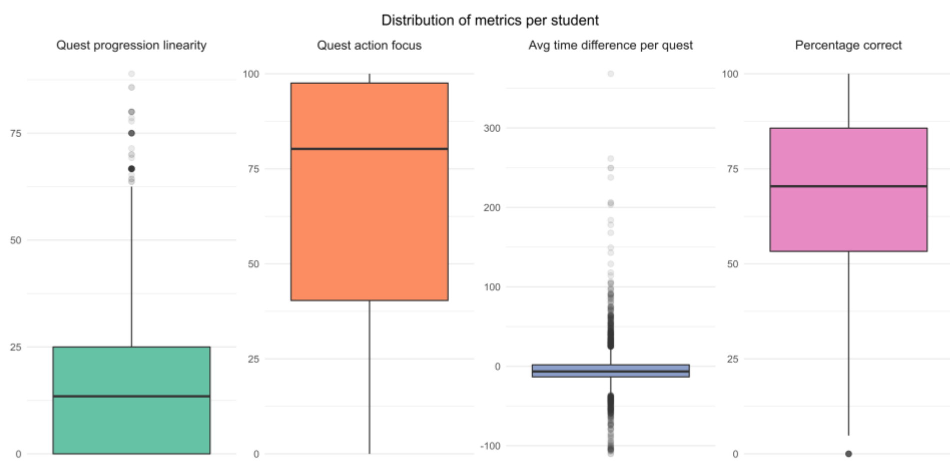

Figure 6. Boxplot distribution of the metrics by student.

Comparing differences in metrics related to students' learning paths between cohorts yields interesting results. In this case, we wanted to compare two cohorts, based on different types of accounts. First, those accounts that were part of the pilot study and were created by students in school, and second, those accounts that were independently created online by any interested individual. Table 1 shows the average for each metric and cohort, where all of the *t*-tests show a statistically significant difference between the means of the two cohorts. The quest progression linearity varies significantly, with school students following a less linear path (17% quest chain changes) than the online learners (5.98%). On the other hand, online learners engaged in fewer chatting events than school students, resulting in a higher quest focus for online learners; 64.2% of school students' events were quest focused versus 75% for online learners. We do not have a clear hypothesis that can explain these differences, but one possibility is that school students explored more in the game (hence with more quest chain jumps and less focus) and that online learners were more serious about advancing with the game. The difference in chatting events may have been facilitated by classroom use, with students playing Radix together in class. Finally, we can also see that school students were a few minutes slower in finishing quests and had a much lower percentage of correct responses.

Table 1. Average value of metrics split in two cohorts by type of account (all *t*-tests are significant with *p*-value below 0.01).

School student	Quest progression linearity	Quest action focus	Avg time difference per quest	Percentage correct
No	5.98 %	73.99 %	-1.77 minutes	80.95 %
Yes	17.47 %	64.23 %	- 5.08 minutes	67.52 %

Table 2 shows the correlation between the metrics. To remove the possibility of spurious correlations or diminished effects, we computed the set of correlations only for those students that interacted with Radix for at least 5 hours (N = 1397). We find a few interesting insights. First, if we look at the correlations of percentage correct to quest progression linearity, we find a low-moderate negative correlation of -0.26, which might indicate that students who are jumping between quest chains will have more failed quest attempts. Second, we also find a low positive correlation of 0.2 when comparing percentage correct with the quest action focus, indicating that students who create a higher proportion of events related to the quest tools are more likely to correctly solve the quest. While common-sense might suggest these results, more work is needed to understand the influence on learning. Finally, we find a low-moderate correlation of -0.32 between quest progression linearity and quest action focus, which indicates that if students are switching quest chains frequently, they are likely to have a lower quest action focus. This may be because they are jumping from one set of quest tools to another due to the frequent changes in quest lines.

Table 2. Correlations between the metrics (* indicates a *p*-value below 0.01).

	Quest progression linearity	Quest action focus	Avg time per quest	Percentage correct
Quest progression linearity	1	- 0.32*	0.09	- 0.26*
Quest action focus	- 0.32*	1	- 0.14	0.2*
Avg time per quest	0.09*	-0.14*	1	- 0.07
Percentage correct	- 0.26*	0.2*	- 0.07	1

6. DISCUSSION

The metrics presented here, describing quest progression linearity, quest action focus, and time per quest, help tell the story of how students are learning and exploring in a game like Radix. Specific learning outcomes are only one aspect of what a student gets out of playing an inquiry-based game, whereas the experience of exploration and discovery is an important part of a student's learning experience in the game. This pathway may vary from student to student depending on their personality, interests, and ways of thinking.

When we looked at quest progression linearity, we see that there is quite a bit of variation in the sequences in which students completed quests. Some students were more focused on one quest line at a time, while others jumped around, completing quests in different topic areas. Radix was designed for students to have this choice to allow students to follow their own interests, an important aspect of inquiry learning. It is, however, important to note that neither type of behavior is inherently better than any other, though these differences provide a source of rich information for teachers to understand what their students are doing in the game, what content they have explored, and what they are interested in. Once teachers understand this metric, they can use the information in their classroom context. For example, a teacher who has asked students to explore the world in an open-ended game-based learning lesson might be very interested to know which students dug deeply into a specific topic area and developed a deep interest in that domain, so that she could support their learning and offer resources for continued study. In addition, this teacher might like to know which students jumped around the most, because it might provide useful evidence of either a lack of focus or an independent motivation, depending on the student. While the quest progression linearity metric alone doesn't tell us what students are learning, it helps describe students' patterns of interaction, and allows some inferences into their interest level, which can inform how a teacher guides their learning.

The quest action focus metric provides another type of insight into how players explore the game world. In Radix, while a player solves quests they can also engage in more interest-based activities. Providing this information to a teacher could be helpful in order to understand how players are engaging with the tools of the game. If a number of students are using the trait examiner tool during a geometry quest, for example, teachers may want to bring that up in a class discussion, finding out what players were trying to do or what they were interested in, in order to support student interests and tie those into the curriculum more tightly. Perhaps those students were sitting near each other, noticed a rare species that happened to live in the area they were walking through, and all decided to find out more about its traits. The teacher might decide to have some discussion about how that experience connected to their science class, or pose the question of why that rare species lives in a particular biome. Similarly, if students seem to be using the chat feature more than usual during a particular quest line, teachers may recognize that there is something of note there—whether it be a challenge that students need to work through together, something they are excited about, or an indication that students were off-task (itself something worth probing further). While the metrics themselves don't tell teachers what exactly is going on, they give a sign that there is something going on that may be worth discussing, thereby tightening up the feedback loop between lesson plans, student experiences, and teacher feedback.

Lastly, being able to compare time spent on quests across quest lines and across students in relation to either a class, school, or larger community of players can also give teachers a richer picture of how students are spending time in the game. If many students spent more time on evolution than on ecosystems for example, and had more failed attempts or demonstrated misconceptions in other science lab activities, a teacher may realize that students need more review on evolution concepts. As another example, if a student spends less time completing the algebra quest line than they did to complete any of the geometry quests (in relation to class averages), but keeps coming back to algebra tools during other quest lines, a teacher might decide to give that student deeper challenges in the area of algebra, where they have an interest and ability.

In addition to providing information just for teachers, if built into the game, all of these metrics could facilitate valuable self-reflection for students not only on *what* they learned, but *how* they learned. Students can explore questions of how efficient they are in their learning, whether they are spending time on the areas they are most interested in, and where they might need to focus their efforts in and outside of the game. These tools can deepen learners' engagement in self-assessment processes, supporting independent learning habits. For all stakeholders, understanding an array of factors about how students are learning and the variety of learning pathways present within one activity can emphasize the richness of the learning experience and the importance of assessing and characterizing more than content knowledge and skills. Providing richer metrics that help us see differences in learning pathways and open reflective conversations can encourage the community to value constructs that are traditionally harder to measure. This approach fits well with formative assessment practices, in that it paints a richer picture that informs what activities to do more or less of, new strategies to try, and connections that can be made between concepts. At a higher level, these three metrics represent a new way to look at measurement and feedback in open-ended digital learning environments. These metrics and others using a similar approach could be applied to different learning games and digital interactives to expand our understanding of learning experiences and refine the way digital learning is embedded into teaching practices.

Metrics and analytics like the ones presented here enable us to measure patterns of exploration. This type of measurement is not unrelated to educational assessment, but it’s important to note their distinct goals and uses. The aim of educational assessment is primarily measuring learning outcomes and collecting empirical evidence of learning gains. Our work on measurement of patterns of exploration does not, however, let us make specific claims about student learning. Rather, it uses clickstream data to discover analytics that can tell us how people are engaging with a game or digital learning experience. We can identify and categorize patterns of interaction and engagement that may vary across individual learners and their learning contexts. These two approaches—educational assessment and analytics of engagement—are complementary because together they provide a complete picture of both what students are learning as well as how they are learning it. In order to guide students along a learning pathway that is productive, educators and learning designers need to know what knowledge and skills their students are building, and also have some understanding of the mechanisms being used to get those results. Either one without the other does not fully explain a student’s learning. By combining learning analytics and educational assessment together in games like Radix, games can provide more robust and actionable interpretations of how students are learning from playing games.

7. BROADER IMPACT AND FUTURE WORK

The way we have defined patterns of exploration and begun to measure them, as described in this chapter, can be applied to other learning games as well. We have presented some examples of how these metrics could be used by teachers to inform instruction, and by students to enable self-reflection in the context of Radix. Building these approaches into an educational game can expand the game itself into a more complete game-based learning system. Continually updated analytics mean that patterns of exploration identified for a given student or class can feed back into the game in the form of adaptive leveling or customized scaffolding. In addition, these informative analytics can be communicated to teachers who can make meaning out of it to provide personalized feedback and support on an ongoing basis. By recognizing these patterns in this way, either the teacher or the game itself can provide valuable data-driven scaffolding and feedback, thereby making the learning experience more relevant to students. All users and stakeholders can more easily recognize the multiple pathways that lead to learning, and celebrate the variety of ways students choose to explore concepts. Our long-term goals in game design and learning analytics is to inspire and assist other designers to incorporate measurement of patterns of exploration into their games, simulations, and digital interactives for learning. We believe that measuring these patterns will not only provide informative data to teachers and students, but that it can reveal the types of exploration actually happening in learning games. Moreover, making the patterns visible can push designers to shape their game environments to be more inquiry-based, student-centered, and constructivist, incorporating more progressive pedagogies that support deep learning and the building of future-ready skills.

The possibilities for future work are far-reaching. We have seen these three metrics exert a small influence on the percentage of correct solutions of a student, which can be seen as a learning outcome within the game environment. However, there is low fidelity of implementation within this study, since students who are working together might be sharing solutions, and because teachers varied widely in how they asked students to use the game. In the future, to more closely connect patterns of exploration with learning outcomes, we might use the pre- and post-tests performed as part of the pilot study in Radix to compute learning gains and find potential relationships with the process metrics we have described. One of the problematic areas of learning analytics in general, and in a game-based learning system specifically, is that the models and metrics are hardly generalizable due to important differences in context. Therefore, we would like to work on developing more general process mining methods that can be applied in multiple game-based learning systems with minimal adaptation. We would like to apply those methods to a variety of learning games designed in our research group and from other organizations. This would enable us to research how patterns of exploration interact with the learning of both content and skills. Through working with practitioners, these learning analytics methods would also provide a rich environment to understand how teachers and students can use the measurement data to become aware of the exploration they are engaging in and more effectively guide their learning journeys. We believe this approach and the ability to better use patterns of exploration represent one of the cornerstones in open-ended and complex environments for learning.

REFERENCES

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, *19*(1–2), 205–220. https://doi.org/10.1007/s10758-014-9223-7

Blikstein, P. (2011). Using learning analytics to assess students ' behavior in open-ended programming tasks. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 110–116). https://doi.org/10.1145/2090116.2090132

Calvo-Flores, M. D., Galindo, E. G., Jiménez, M. C. P., & Pineiro, O. P. (2006). Predicting students' marks from Moodle logs using neural network models. *Current Developments in Technology-Assisted Education*, *1*(2), 586–590.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79–122.

Connolly, T. M., Boyle, E. A., Macarthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71.

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice*, *3*(1), 63–78.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, *1*(1), 20.

Gibson, D., & Clarke-Midura, J. (2015). Some psychometric and design implications of game-based learning analytics. In *E-learning systems, environments and approaches* (pp. 247–261). Springer.

Groff, J., Clarke-Midura, J., Owen, V. E., Rosenheck, L., & Beall, M. (2015). Better learning in games: A balanced design lens for a new generation of learning games.

Honey, M. A., & Hilton, M. (2011). Learning science through computer games and simulations. Committee on science learning: computer games, simulations, and education. Board on Science Education, National Research Council. Washington, DC: The National Academies Press.

Kim, Y. J., Ruipérez-Valiente, J. A., Tan, P., Rosenheck, L., & Klopfer, E. (2019). Towards a Process to

Integrate Learning Analytics and Evidence-Centered Design for Game-based Assessment. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*.

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In *Workshop on Analysis of Large Scale Social Interaction in MOOCs* (pp. 60–65). Retrieved from http://www.aclweb.org/anthology/W/W14/W14-41.pdf#page=67

Nghe, N. T., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. In *Frontiers in Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual* (p. T2G--7).

Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1–12. https://doi.org/10.1016/j.compedu.2008.06.004

Pardos, Z. A., Gowda, S. M., Ryan, S. J. D., & Heffernan, N. T. (2010). Ensembling Predictions of Student Post-Test Scores for an Intelligent Tutoring System. In *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 189–198).

Prensky, M. (2006). *Don't bother me, Mom, I'm learning!: How computer and video games are preparing your kids for 21st century success and how you can help!* Paragon house St. Paul, MN.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., & Delgado Kloos, C. (2017). Detecting and Clustering Students by their Gamification Behavior with Badges: A Case Study in Engineering Education. *International Journal of Engineering Education*, *33*(2–B), 816–830.

Ruiperez-Valiente, J. A., Munoz-Merino, P. J., Delgado Kloos, C., Niemann, K., Scheffel, M., & Wolpers, M. (2016). Analyzing the Impact of Using Optional Activities in Self-Regulated Learning. *IEEE Transactions on Learning Technologies*, *9*(3), 231–243. https://doi.org/10.1109/TLT.2016.2518172

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado Kloos, C. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, *47*(Learning Analytics, Educational Data Mining and data-driven Educational Decision Making), 139–148. https://doi.org/10.1016/j.chb.2014.07.002

Salehi, M., & Nakhai Kamalabadi, I. (2013). Hybrid recommendation approach for learning material based on sequential pattern of the accessed material and the learner's preference tree. *Knowledge-Based Systems*, *48*, 57–69. https://doi.org/10.1016/j.knosys.2013.04.012

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, *2*(1), 13–48.

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2012). Tracing a Little for Big Improvements: Application of Learning Analytics and Videogames for Student Assessment. *Procedia Computer Science*, *15*, 203–209. https://doi.org/10.1016/j.procs.2012.10.072

Squire, K. (2011). *Video Games and Learning: Teaching and Participatory Culture in the Digital Age. Technology, Education--Connections (the TEC Series)*. ERIC.

Takeuchi, L. M., & Vaala, S. (2014). Level up Learning: A National Survey on Teaching with Digital Games. *Joan Ganz Cooney Center*.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, *34*(3), 229–243.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249.

AUTHOR INFORMATION

José A. Ruipérez-Valiente
Massachusetts Institute of Technology
600 Technology Square, 2nd floor. Cambridge (MA), 02139, USA.
jruipere@mit.edu
http://joseruiperez.me/
He completed his B.Eng. and M.Eng. in Telecommunications at Universidad Católica de San Antonio de Murcia (UCAM) and Universidad Carlos III of Madrid (UC3M) respectively, graduating in both cases at the top of his class. Afterwards, he completed his M.Sc. and Ph.D. in Telematics at UC3M while conducting research in learning analytics and educational data mining at Institute IMDEA Networks. During this time, he completed two research stays of three months each, the first at MIT and the second at the University of Edinburgh. He has received more than a dozen academic and research awards and has published over 35 scientific publications in high impact venues. Currently he is a postdoctoral associate at the CMS/W department at MIT where he is part of the Teaching Systems Lab and Playful Journey Lab, and also collaborates with the Education Arcade in applying data science to large scale free online courses and game-based environments to enhance human knowledge on how we learn.

Louisa Rosenheck
Massachusetts Institute of Technology
700 Technology Square, 3rd floor. Cambridge (MA), 02139, USA.
louisa@mit.edu
https://education.mit.edu/team/
She is a designer and researcher of educational technologies in the MIT Education Arcade and Playful Journey Lab. She manages the design, content, and development of educational games and simulations, and oversees the research on how games can be effectively used in both formal and informal educational settings. She also brings the playfulness of games into her assessment design work, creating digital and non-digital tools to help both students and teachers assess creative work and soft skills. In addition to these areas, she supports other designers and organizations in creating innovative ed tech and curriculum. Through capacity building partnerships on local and international ed tech programs, she aims to increase the deep learning in games and tools being produced for education.

Yoon Jeon Kim
Massachusetts Institute of Technology
600 Technology Square, 2nd floor. Cambridge (MA), 02139, USA.
yjk7@mit.edu
http://yoonjeonkim.net/
She is the director of the Playful Journey Lab within 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 Assessment working group (https://playful.mit.edu/playfulassessment), where she and her colleagues explore playful, authentic, yet rigorous assessment approaches in both digital and non-digital learning environments. The core of her work is close collaboration with practitioners—empowering teachers to innovate around classroom assessment and use playful and authentic assessment tools that can truly impact student learning.