# Patterns of Engagement in an Educational Massively Multiplayer Online Game: A Multidimensional View

José A. Ruipérez-Valiente, *Member, IEEE,* Matthew Gaydos, Louisa Rosenheck, Yoon Jeon Kim, and Eric Klopfer

*Abstract*—Learning games have great potential to become an integral part of new classrooms of the future. One of the key reported benefits is the capacity to keep students deeply engaged during their learning process. Therefore, it is necessary to develop models that can measure quantitatively how learners are engaging with learning games to inform game designers and educators, and to find ways to maximize learner engagement. In this work, we present our proposal to multidimensionally measure engagement in a learning game over four dimensions: general activity, social, exploration, and quests. We apply metrics from these dimensions to data from *The Radix Endeavor*, an inquiry-based online game for STEM learning that has been tested in K-12 classrooms as part of a pilot study across numerous schools. Based on these dimensions, we apply clustering and report four different engagement profiles that we define as: "integrally engaged," "lone achiever," "social explorer," and "non-engaged." We also use three variables (account type, class grade, and gender) to perform a cross-sectional analysis finding interesting, statistically significant differences in engagement. For example, in-school students and accounts registered to males engaged socially much more than out-of-school learners or accounts registered to females, and that older students have better performance metrics than younger ones.

*Index Terms*—Engagement, learning games, learning analytics, K-12 education, game-based assessment.

## I. INTRODUCTION

**D**IGITAL games have great potential to provide alternative ways to assess learning. As digital games have become an integral part of daily life for younger generations, activity in digital games generates rich data across multiple time points and contexts. With systems designed to collect their digital footprints, these data then can be used to make inferences about what they learn and how they learn without interrupting engagement in digital games. Because digital games often employ challenging, interesting, and complex problems, they can be used to generate evidence for 21$^{st}$ century competencies, which are traditionally difficult to measure using conventional forms of assessment. By combining the ability to extract rich and continuous data in games with the complex and authentic problems that games present, game-based assessments can provide a more comprehensive learner profile than traditional assessment methods. Although game-based learning (GBL) continues to gain importance as a method to keep students (especially younger students) engaged in learning [1] and improve learning outcomes [2], measuring engagement remains a significant challenge [3].

In this paper we present our vision of a multidimensional approach to engagement in the context of GBL, showing that some learners can be engaged with certain dimensions of tasks but not with others. The concept of engagement that we explore in this study is related to the degree of activity or attention someone gives to certain tasks over some period of time [4], which can be linked to their intrinsic interest, motivation, and learning [5]. The goal of our study is to provide a more comprehensive look at how player engagement can be measured through in-game behavior.

We analyze data from a pilot study using *The Radix Endeavor* (shortened to *Radix*), an educational Massively Multiplayer Online Game (MMOG) based on inquiry learning for STEM topics targeting middle and high school curricula. During this pilot, more than 10,000 students participated and it was tested as part of the curriculum in numerous states. Previous work has discussed the multifaceted appeal of MMOGs [6] and also their potential as educational tools [7]. Therefore, we are motivated to develop a multidimensional model to see how students are engaging with *Radix* in different ways, since the game was designed to engage teachers in students in multiple ways. More specifically, we establish the following objectives:

1) Define a series of metrics that provide measures of how students engaged in different dimensions with *Radix*.
2) Apply these metrics to the data that have been collected in the case studies with *Radix*, exploring the results and differences across cross-sectional variables.
3) Apply clustering analysis to find the most common engagement profiles of students and see how these change based on different cross-sectional variables.

The remainder of the paper is organized as follows. Section II reviews the background literature on student engagement, player engagement, and learning. Section III describes our methods, including a description of the game, *Radix*, as well as the data collection process and the context of the case study. Section IV presents our proposal of metrics in multiple

J. A. Ruipérez-Valiente is with the MIT Playful Journey Lab and the MIT Teaching Systems Lab. M. Gaydos and Y. J. Kim are with the MIT Playful Journey Lab. L. Rosenheck and E. Klopfer are with the MIT Education Arcade. Massachusetts Institute of Technology, Cambridge (MA), 02139, USA (e-mail: {jruipere, gaydos, louisa, yjk7, klopfer}@mit.edu).

dimensions and Section V describes the exploratory analysis after applying those metrics on *Radix* data. Then, we apply clustering analysis to find the most common engagement profiles in Section VI, eventually finalizing the paper with discussion and conclusions in Section VIII.

## II. BACKGROUND

In this section, we outline the importance of student engagement broadly, previous work on player engagement, and prior methods used to measure player engagement.

### A. Student Engagement and Learning

Student engagement in academic contexts is a broad construct that has been linked to a wide variety of desirable educational outcomes. Engaged students, for example, show improved learning, critical thinking, and grades [8] and are less likely to drop out of school [9]. Despite the importance of engagement in student learning, its definition is not universally agreed upon. Instead, there are multiple definitions stemming from different disciplines, broadly linking engagement to student motivation [10], and in the context of formal schooling, to developing relevant competence [11]. Engagement may be considered, for example, "a collection of mindful goal-directed states in which motivation arising from positive emotions serves to grab and sustain the learner's cognitive and motor competencies" [12]. It is comprised of three sub-components (behavioral, cognitive, and affective) [13], [14] or five sub-components (behavioral, cognitive, motivational, cognitive-behavioral, and social-behavioral) [10], and individuals may exhibit five different engagement profiles: highly engaged, moderately engaged, and minimally engaged; emotionally disengaged, and cognitively disengaged [15]. Common across these definitions of engagement is the notion that engagement may be thought of as enacted motivation, and that clearly defining engagement as well as reliably and accurately measuring its operationalization remains a challenge, particularly due to differences in how it manifests across different contexts [16].

### B. Player Engagement

Although there has been plenty of research on the importance of student engagement in improving learning outcomes (e.g., [17]), engagement in the context of games inherits the same challenges of engagement research more broadly. In particular, Fredricks *et al.* [5] highlight the need to improve the measurement of student engagement, especially using rigorous approaches that account for individual differences, different contexts, and multiple levels, all of which may change over time.

Consider, for example, the research on player engagement tied to Csíkszentmihályi's [18] concept of flow, a state of immersion that one experiences when engaged in experiences that one sees as enjoyable and valuable. Flow's immersive nature makes it difficult to study in real time, as interrupting players with questionnaires can drop them out of the experience that provided flow in the first place. Perhaps unsurprisingly,

prior work on in-game engagement has tended to develop survey instruments to gauge students' self-reported video game engagement, administered outside the context of game play (e.g., [2], [19]–[22]). Though in-game data is regularly used in video game entertainment as a means to retain players and increase profits, relatively few educational studies have been published on measures of in-game player activity as a data source for measuring engagement, either to improve learning or to improve our understanding of the construct (see [23], [24] for existing work in this field).

One study that provides insight into how complex the challenge of linking in-game data to player engagement is by Martey *et al.* [4]. Their work reflected the multi-faceted nature of engagement, combining multiple instruments to capture different aspects of player engagement, including self-report, content analysis, electrodermal activity, mouse movement, and game-click logs. They found that the different means used to measure engagement captured different aspects of the construct. For example, they found that attention-based engagement measures (i.e., self-report) were unrelated to behavioral (i.e., mouse-clicks) as well as physiological measures (i.e., electro-dermal activity), necessitating the use of multiple data sources to capture the multiple, distinct dimensions of game engagement. More recently, Cano, Fernández-Manjón, and García-Tejedor [25] examined student engagement through a serious game designed to help students with intellectual disabilities learn to navigate a local subway system. Their measure of player engagement focused on student in-game inactivity, confirmed by direct observation. In the context of game-based learning, measures of engagement have tended to examine the motivations of play [26], the subjective experiences of players [22], or the physiological responses accompanying engaged play [27]. Work has begun to provide theoretically sophisticated measures of engaged player behavior [23] accompanied by straightforward measures of engagement such as time on task [24].

In this study, we focus on engagement in terms of individuals' attention and activity, measured through in-game actions (i.e., mouse clicks) associated with the designed game tasks (e.g., quests). Building on the case study of game-based engagement by Ke, Xie, and Xie [28], we assume that, in this context, engagement is an evolving relationship between the player and game environment and that engagement in the game content is not something that develops necessarily and uniformly through game play, but rather may evolve through reflection on the game activity and may be exhibited in different ways by different players. This reciprocally developing nature of engagement is supported by prior work looking students and teachers interacting in context [29], [30]; however, these theoretically sophisticated views of engagement have only minimally been studied in games.

Focusing on the measurement of in-game engagement development in the context of a complex massively multi-player online game environment, we explore different measures' usefulness in capturing the relational and individual nature of engagement. The approach taken here thus enables measurements of engagement as they vary by different groups of players. Directly measuring player activity in order to estimate

player engagement in large-scale digital educational games may not only be useful for gauging player engagement as it changes over time, but also for developing assessments that take into account individual differences.

## III. METHODS

### A. The Radix Endeavor

*The Radix Endeavor* [31] is an inquiry-based, multi-disciplinary, MMO-style online game for STEM learning that includes a balance of guided tasks and open-ended exploration. It is inquiry-based in the sense that in it, players solve problems by exploring a topic, figuring out what questions need to be asked, and determining a pathway to answer those questions. It is an MMOG-style in that it involves players controlling an avatar in a third person perspective, is set in a virtual multiplayer world that is open-ended, and includes set sequences of tasks for players to work through as they explore the world and build conceptual understanding. It was developed at the MIT Education Arcade (see videos in the YouTube channel [32]), was launched in January 2014, and was free to play.

*Radix* is aligned with the Next Generation Science Standards [33] for biology and the Common Core State Standards [34] for math, incorporates STEM practices, and encourages students to develop $21^{st}$ century skills (e.g., critical thinking, collaboration) inside and outside of the game. It is meant to be played over the course of a semester and revisited during each relevant curricular unit.

### B. Educational Context

*Radix* was tested in numerous classrooms across the US and other countries in a pilot study, which ran from January 2014 through August 2015. During that time, teachers could create accounts for their students to play and informal marketing and outreach was done to recruit teachers to participate in the pilot, including reaching out to local and national teacher networks to publicize the game, as well as press articles and blog posts showcasing the project and its opportunities for participation. Participating teachers were provided with professional development opportunities and implementation resources, which included in-person sessions, monthly webinars, an online forum, information on alignment with standards, and suggestions for bridging curriculum. Teachers were encouraged to tailor their implementations and use the game as they saw fit in their classroom; most had their students play relevant quest lines at the time they were covering a given topic area in their class. However, we did not have a tight control of the way teachers implemented *Radix* in their classes.

Furthermore, 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. Additionally, players who heard about the game via other channels could create player accounts not associated with an educational institution. These players that were not affiliated with any specific school and teacher, played as much or little as they chose to, working through quest lines according to their interests. Because of the differences in control over and information about one cohort (in-school students) versus the other (out-of-school learners), we consider these to be distinct cohorts. *Radix* remained accessible and available to play on the website through late 2019.

### C. Data Collection

In this study, we include all the data collected until the $2^{nd}$ February 2018 from both in-school students and out-of-school learners, a total of 26,959 accounts. The inclusion and exclusion criteria are as follows:

- Teacher, staff, and research accounts were removed, leaving only in-school students and out-of school learner accounts.
- We eliminated accounts that had not reached a minimum interaction of two hours, which is the estimation from designers for learners to getting familiar with *Radix* game mechanics.
- We kept only the accounts that had completed the first three quests of the tutorial, since the two first quests do not require any specific interaction with the environment–just speaking with the non-player characters (NPCs)–and the third task is the first one where they need to use a tool to correctly finish the quest.

Conducting the analysis only on learners who met a minimum level of interaction allows us to be more confident about the results not being biased by spurious and random correlations. The final sample size and number of learners included in the analysis is 5,545. We define two different cohorts from these accounts:

- **In-school students** (5,114, 92.23%): These accounts were created by teachers for use in the classroom.
- **Out-of-school learners** (431, 7.77%): These accounts were created by online users and were not affiliated with a school. We do have access to the demographic variables for this cohort (see below).

For in-school student accounts, gender and class grade distributions are reported:

- **Gender**: Gender was reported for 5114 in-school students, with a distribution of 54.97% male and 45.03% female.
- **Class grade**: Class grade was reported for 5027 school learners. The range of class grades in our sample range from $1^{st}$ grade to $12^{th}$ grade. Because the STEM content focus goes from $6^{th}$ to $12^{th}$ grade, we provide further analysis of students in that grade interval, which represents a total of 4904 school learners. These learners are distributed by grade as follows: 20.27% for $6^{th}$ grade, 23.53% for $7^{th}$ grade, 17.99% for $8^{th}$ grade, 19.54% for $9^{th}$ grade, 12.30% for $10^{th}$ grade, 3.92% for $11^{th}$ grade, and 2.47% for $12^{th}$ grade. In order to map class grade to age, we assume that in the US system, students would normally turn 11 years old during the year of entry in $6^{th}$ grade.

Fig. 1. Two examples of different ways to engage with *Radix*: (a) shows a virtual social interaction with other peers in *Radix* and (b) shows an example of a quest to breed plants to learn the genetics basis.

## IV. PROPOSAL OF ENGAGEMENT METRICS AND DIMENSIONS

To help students build content knowledge while developing disciplinary practices and important 21$^{\text{st}}$ century skills, *Radix* relies on embedded content, inquiry-based pedagogy, and opportunities for collaboration and discussion. *Radix* was designed carefully to align the game mechanics learning objectives and data collected [35]. There are three broad categories of activity that players engage in while playing *Radix*. Each area has its own pedagogical purpose, while some actions are also designed to overlap in multiple categories:

- **Quests:** A quest in game environments refers to a task that a player needs to complete to achieve a reward or to progress in the game. There are around 120 different quests, and 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. Players have a variety of STEM-related tools available to use in collecting data and completing quests e.g., the breeding station Genetics, and the cartogram for mapping Geometry. Proper use of these tools embodies the skills and practices important to that discipline, and the tools are necessary to conduct investigations in the game world.
- **Social Interactions:** Players have a number of ways to interact with each other within the game, including public text chat, private chats, sharing items, and creating and joining a party, which is coordinated group play with shared data collection. These interactions are designed not only to facilitate collaboration on quest tasks, but to let players feel like part of a community and find support for problem-solving.

- **Exploration:** The *Radix* world and game environment are designed to foster exploration, both in and outside of quest tasks. Players can customize their avatars, walk around the world, and collect specimens from different biomes. Many of the tools are available regardless of whether certain quests are unlocked, meaning that players can, at any point, examine traits, measure objects, breed species, and much more, if they want to figure out how the world works.

Ideally, players engage with the *Radix* world in all three ways. Of course, that does not mean that every player needs to engage in all three activities on every quest. Rather, understanding how players use the game or are engaged by it was central to this study. In addition to these three qualitative dimensions, we see the overall general activity as a transverse dimension. We visually represent this multidimensional view of the ways students can engage with *Radix* in Figure 2. We can naturally interpret the intersections of the Venn diagram as students who engage with several dimensions of the model.

We present now the process and steps taken to design and select the final metrics of the model that we present later in the Section:

1) Formation of a team between *Radix* designers, learning analytics researcher and learning scientist, that have deep knowledge of *Radix* to understand the dynamics of the game.
2) Team brainstorm to design metrics that can reflect the original pedagogical intentions of the *Radix* designers for each dimension.
3) Technical implementation of metrics constrained to the actual data available. Some metrics were not feasible in practice.
4) Intra-dimension feature selection process within each one of the categories, removing the most redundant features. Cross-dimension correlations between metrics might exist, but are not considered as part of this selection.

This last stage generated the final selection of dimensions and metrics that are included in the multidimensional model,
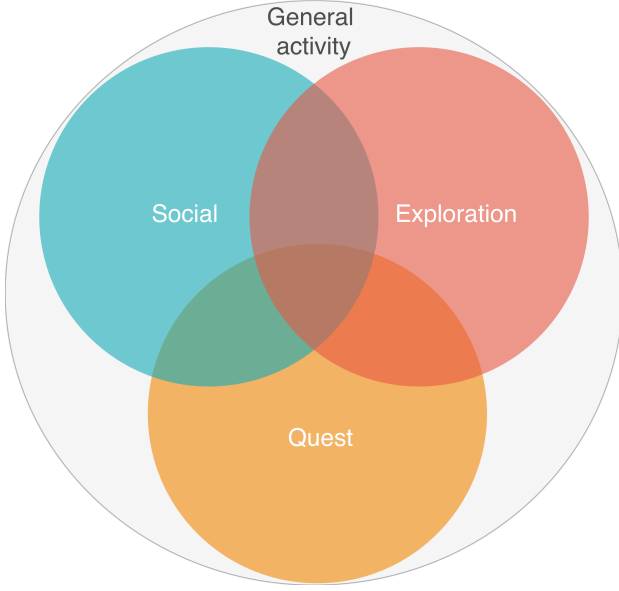
Fig. 2. Multidimensional view of how students can engage with *Radix*.

which are described in the following Sections. In addition to these metrics, we use percentage of correct quest attempts (p_correct) as a measure of performance or efficiency of the student solving quests.

### A. General Activity Metrics

The general activity of players in terms of time:
- **Active time** (active_time): Measure of the in-game active time in minutes with an inactivity cutoff.
- **Number of active days** (n_active_days): Number of different days that an account logged into *Radix*.

### B. Social Metrics

These metrics capture different actions that students can do to socialize and interact with peers within the game:
- **Number of zone chats** (n_zone_chats): Number of zone chat messages sent by the avatar. Zone chat messages can be read by all players within the same zone. Figure 1a shows an example of learners interacting with zone chat.
- **Number of private chats** (n_private_chats): Private messages sent to others through party chats (chats only open to users within the same party) and mail messages (private messages to a single user). This metric measures the total number of both of these chat message types.
- **Average number of characters per chat message** (chars_per_chat): This metric is computed by dividing the total number of messages for a player by the total number of alphanumeric characters of those messages providing the average length of the chat messages for a player.
- **Number of parties** (n_parties): Number of parties that a student has joined. In *Radix*, a party is a way of grouping together with a few other players to enable small group chat and sharing of in-game data.

### C. Exploration Metrics

These metrics indicate the degree of interaction with the available tools and the world of *Radix*:
- **Percentage of different zones** (p_zones): The world of *Radix* includes 28 different zones that students can explore and visit. Each zone contains different NPCs. There is also a world map available for players to know how to travel to a desired area. This metric measures the percentage of different zones visited by a given avatar.
- **Percentage of different tool events** (p_tools): There are 19 different tools or actions that a player can experiment within the *Radix* world in order to solve quests and learn STEM content. This metric measures the percentage of tools that the avatar has used.
- **Number of tool events** (n_tool_events): This metric n_tool_events provides a summation of all the tool events triggered by a student that indicates a degree of experimentation with all of the available tools.

### D. Quest Metrics

These metrics describe interactions with the quest system:
- **Percentage change quest chain** (p_quest_change): *Radix* quests do not have to be completed in a single, linear sequence, and students are free to jump from one quest chain to another. This item measures the percent of changed quests relative to how many quests a student completed.
- **Percentage of quest action focus** (p_quest_focus): This measures the percent of events completed that were quest-related relative to events completed that were not quest-related, for each solved quest. The interpretation is that a lower percentage of p_quest_focus might indicate students experimenting with the different tools either to find the solution or simply out of interest, or it could be due to confusion on how to actually reach the desired outcome.
- **Percentage of completed quests** (p_completed): This metric provides a percentage of the number of quests that have been completed by the student in *Radix*.

## V. EXPLORATORY ANALYSIS OF THE ENGAGEMENT METRICS

This section explores the results after applying the scripts to compute the metrics in the data collection reported in Section III-C. First, we look at the distribution of the metrics by account type in Section V-A, a cross-sectional analysis of the demographics in Section V-B, and last some insights about correlations in Section V-C.

### A. Distribution of Metrics by Account Type

Figure 3 shows a graph with a boxplot distribution separated by account type, which can be used to both depict a global overview for each metric but also to compare the two cohorts. We have limited the *y*-axis to a value of 300 to remove outliers and facilitate a better visualization, but outliers are included when generating all reported stats and analysis. The name
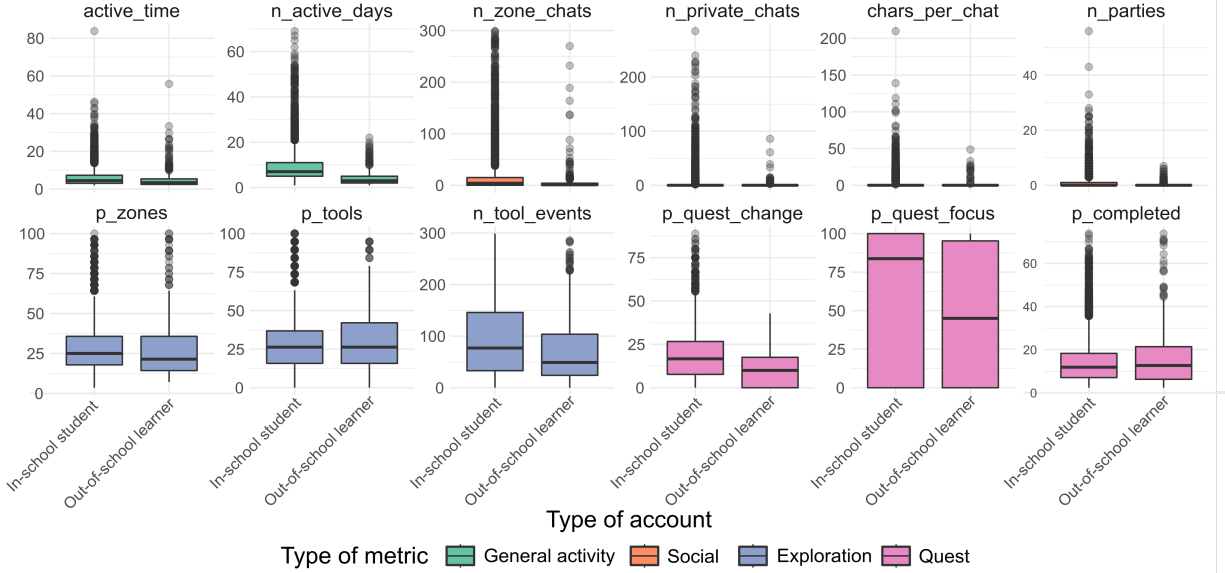
Fig. 3. Boxplot visualization of the distribution of all metrics separated by type of account.

of the metric is on the top of each boxplot and the box color indicates the type of metric as described in Section IV. A visual exploration can allow easy detection of differences in some metrics and a Multivariate Analysis of Variance (MANOVA) confirms that the metrics of both cohorts have significantly different means ($F = 54.72, p < e^{-16}$). A more in-depth look into the significance of each one of the metrics using simple one-way ANOVAs reveals the following findings: for the general activity metrics, the amount of `active_time` is significantly different ($F = 20.04, p < e^{-6}$) with a mean of 6.1 (in-school) vs 5 (out-of-school) hours invested, and `n_active_days` is also significant ($F = 140.78, p < e^{-16}$), with a mean of 9 (in-school) vs. 4 (out-of-school) active days. In-school students and out-of-school learners were active for a similar amount of time, around 5-6 hours, but in-school students were active on average twice as many days than out-of-school learners. Interestingly, this activity difference did not lead to more exploration as these three metrics do not show a statistically significant difference between the groups. Additionally, out-of-school learners had significantly higher `p_completed` ($F = 9.35, p = 0.002$). This difference in performance could be related to age or prior knowledge, though such conclusions cannot be drawn with confidence since we do not know the demographics of out-of-school learners. Another possibility is that the more frequent quest chain changes for in-school learners (`p_quest_change`, $F = 138, p < e^{-16}$), can have an impact in their `p_completed` progress, either diminishing it or suggesting an underlying problem (e.g., frustration, disinterest) that caused a pattern of regularly abandoning quest chains to start new ones. Finally, in-school students were more social, with higher `n_zone_chats` (average of 24 vs. 7.3 public messages, $F = 5.43, p = 0.019$), `n_private_chats` (average of 5.5 vs. 0.8 private messages, $F = 14.7, p = 0.0001$), and `n_parties` (average of 0.9 vs. 0.3 parties, $F = 27.64, p < e^{-7}$), but

there was no difference in terms of `chars_per_chat`. One surprising finding when facilitating the use of *Radix* in classes is that students would frequently use the chatting functionality of the game to talk with friends, even when they could also speak to them face-to-face. This translates to much higher levels of social activity within the cohort of school students.

*B. Cross-Sectional Analysis of Demographics*

In this section we perform a similar cross-sectional analysis as before, but using demographic variables to see how gender and class grade influence significant differences in how students engaged with the game.

First, in terms of gender, a MANOVA test also confirms differences in the engagement when comparing female and male students ($F = 18.15, p < e^{-16}$). If we unpack these differences (see Figure 4), the most interesting results indicate that both cohorts invested similar time and completed similar number of quests (`active_time` and `p_completed`), but female students have an average `p_correct` of 67.31% compared to 63.9% for male students ($F = 46.89, p < e^{-12}$). Additionally, `n_tool_events` average is 154 vs. 183 for female and male students respectively ($F = 13.24, p = 0.0002$), even though `p_tools` used are about the same. So male students performed more tool exploration and had a lower ratio of correct responses than their female counterparts, despite the fact that both groups invested a similar amount of time and completed a similar amount of quests. Finally, the social activity of the average male student in terms of `n_zone_chats`, `n_private_chats`, `chars_per_chat`, and `party_joined` doubles the level of activity of the average female students, being statistically significant for all metrics except for `chars_per_chat`. Previous studies have also found female students to be less socially active in online learning environments [36].
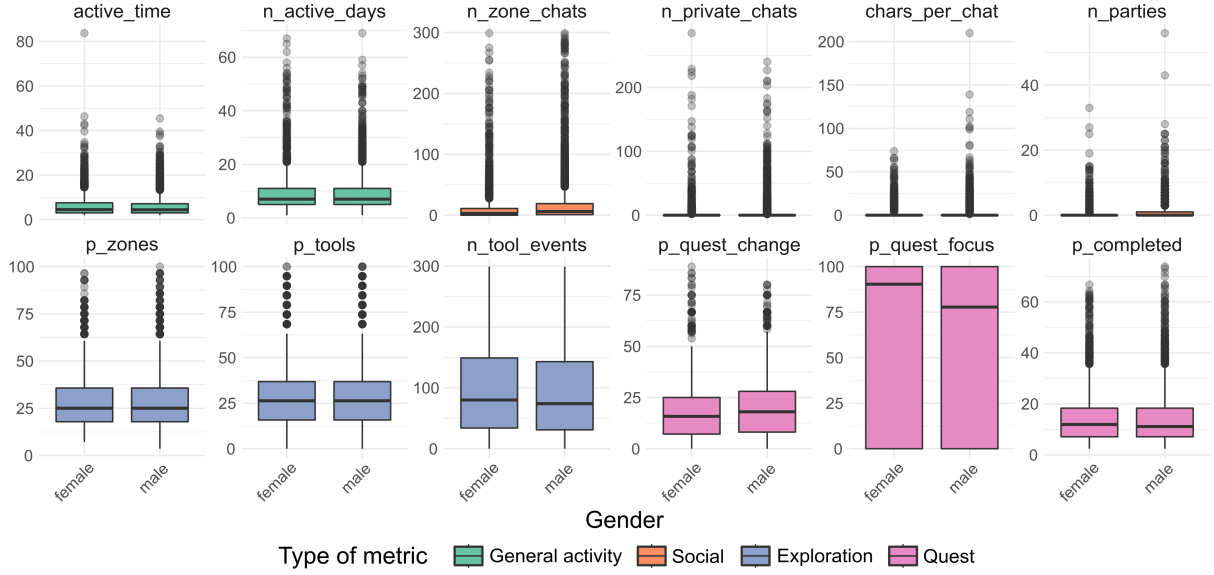
Fig. 4. Boxplot visualization of the distribution of all metrics separated by gender.
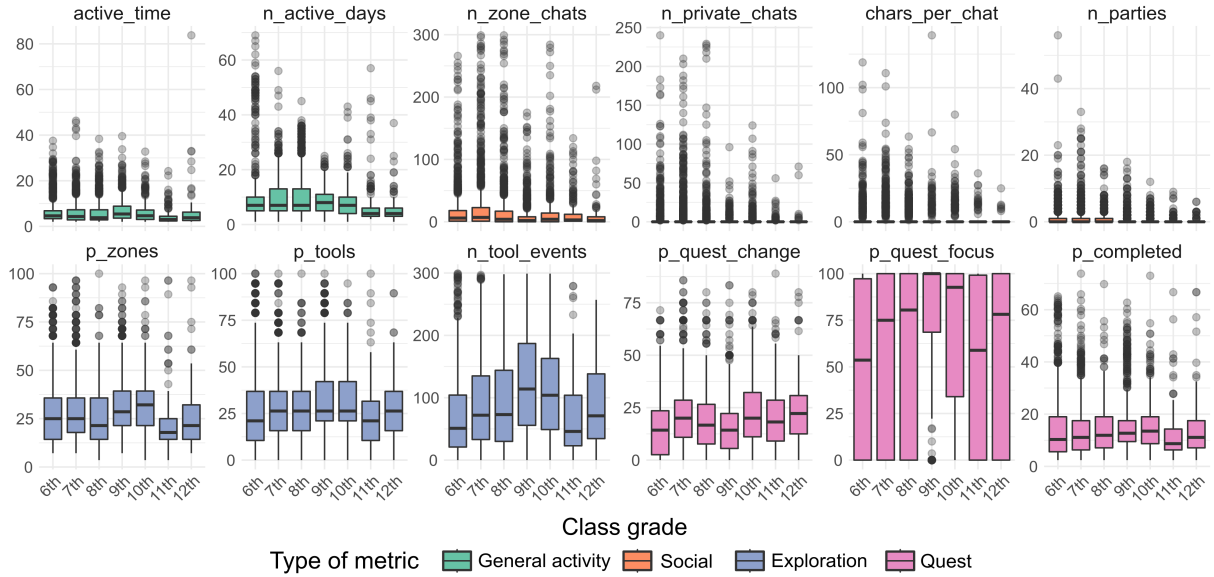


Fig. 5. Boxplot visualization of distribution of all metrics separated by class grade.

Second, we explore the changes in the engagement metrics using class grade as a cross-sectional variable (see Figure 5), which is also statistically significant according to the MANOVA test ($F = 32.54, p < e^{-16}$). We analyze these differences in terms of the grade progression, from younger to older ages, so that we can provide a qualitative meaning to these differences. We find that all class grades had a similar `active_time` but that younger students were active for more days (`n_active_days`, $F = 89.07, p < e^{-16}$). All class grades have similar `p_completed`, but older students have a significantly higher `p_correct` ($F = 9.39, p = 0.002$). Accordingly, all class grades reached a similar level of progress within *Radix*, but older students did so with a higher

correct ratio and were able to solve quests faster. Finally, all social metrics are also statistically different, showing a trend were young students were much more socially active within *Radix* than older students.

## C. Correlation Analysis

Figure 6 shows a correlation matrix visualization where the color codifies the value of the correlation. Uncolored cells mean that the correlation is not significant. Most of the correlations follow the typical positive pattern where doing more, means more, i.e., a higher `active_time` or `n_active_days` implies more `p_completed` or

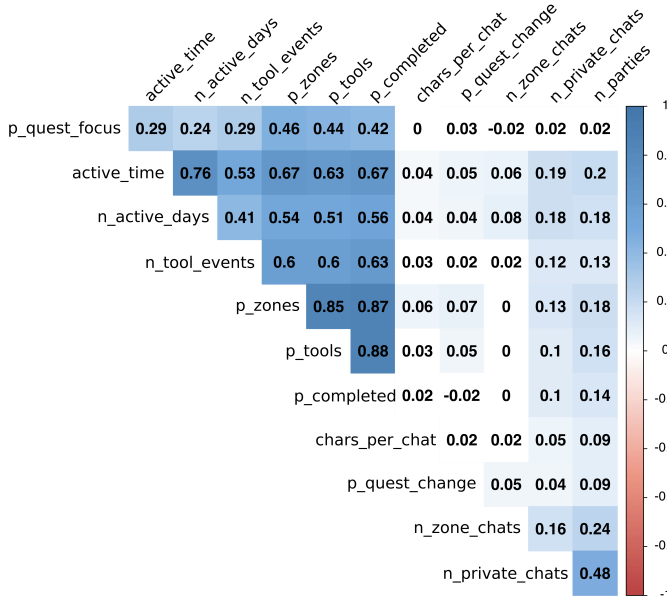| | active_time | n_active_days | n_tool_events | p_zones | p_tools | p_completed | chars_per_chat | p_quest_change | n_zone_chats | n_private_chats | n_parties |
|---|---|---|---|---|---|---|---|---|---|---|---|
| p_quest_focus | 0.29 | 0.24 | 0.29 | 0.46 | 0.44 | 0.42 | 0 | 0.03 | -0.02 | 0.02 | 0.02 |
| active_time | | 0.76 | 0.53 | 0.67 | 0.63 | 0.67 | 0.04 | 0.05 | 0.06 | 0.19 | 0.2 |
| n_active_days | | | 0.41 | 0.54 | 0.51 | 0.56 | 0.04 | 0.04 | 0.08 | 0.18 | 0.18 |
| n_tool_events | | | | 0.6 | 0.6 | 0.63 | 0.03 | 0.02 | 0.02 | 0.12 | 0.13 |
| p_zones | | | | | 0.85 | 0.87 | 0.06 | 0.07 | 0 | 0.13 | 0.18 |
| p_tools | | | | | | 0.88 | 0.03 | 0.05 | 0 | 0.1 | 0.16 |
| p_completed | | | | | | | 0.02 | -0.02 | 0 | 0.1 | 0.14 |
| chars_per_chat | | | | | | | | 0.02 | 0.02 | 0.05 | 0.09 |
| p_quest_change | | | | | | | | | 0.05 | 0.04 | 0.09 |
| n_zone_chats | | | | | | | | | | 0.16 | 0.24 |
| n_private_chats | | | | | | | | | | | 0.48 |

Fig. 6. Correlation matrix visualization of engagement metrics.

n_zones. However, there are a few details that are interesting. First, the correlation of general activity metrics with social metrics is low, which means that the social dimension does not strongly depend on the time spent within the game environment, but instead is more related to learner interests. Another interesting but low correlation is between p_quest_change and p_correct (-0.18), which would imply that changing quest chains frequently might have a negative impact on the performance of students.

## VI. Clustering Analysis

The results of Section V showed that there are a number of variables that have a heavy influence on the type of engagement shown by students using *Radix*. In this section, we want to obtain higher level profiles of engagement by applying clustering so that we can analyze each cluster separately. We utilized the following steps:

1) We use *k*-means as the clustering algorithm, a common approach when using a set of continuous variables as input. We use as input all the variables from the four dimensions: general activity, exploration, social and quest.

2) One of the issues of *k*-means is the selection of the appropriate number of clusters. We run *k*-means with *k* values from 1 to 15 and use the common elbow technique [37] to make the selection based on the within-clusters sum of squares change for each *k* value.

3) Based on the elbow technique, we finally select four clusters and apply *k*-means. We qualitatively describe each of the clusters and perform Pearson's Chi-squared test to see if the cross-sectional variables (type of account, gender, and class grade) have an effect in the distribution of clusters.

Figure 7 shows the distribution of variables by each cluster. Next, we perform a qualitative analysis of each one of these clusters:

- **Cluster 1** (9.53%): The smallest cluster in size represents the cohort of learners with the highest levels of engagement with all variables related to the general activity (average active_time of 14.72 hours and n_active_days of 18.45 days), exploration, and quest progress (average p_completed of 38%). They strongly engaged socially as well, with the highest number of n_private_chats and n_parties. However, the number of n_zone_chats is smaller than Cluster 4, so they communicated in a more private way with their party members. Perhaps it was this strong social interaction with other peers and numerous parties that kept them engaged long enough to achieve the highest p_completed of all clusters. All of the metrics are also higher than students in Cluster 2, except for p_quest_focus, which is slightly lower. Therefore, we can define this cluster as the "integrally engaged learners," that engaged with all the dimensions of *Radix*.

- **Cluster 2** (46.9%): The second cluster of learners is the largest in size, and thus the most representative. They have invested a similar amount of effort as Cluster 4 (around an average active_time of 6 hours and n_active_days of 8 days), higher than Cluster 3 but lower than Cluster 1. However, they had slightly higher exploration metrics and they completed more quests (an average p_completed of 15.5% vs 7%) than Cluster 4. Moreover, the key difference with Cluster 1 and 4 is that they barely engaged in social activities of any kind. Therefore, this cluster of learners engaged with the quest and exploration dimensions, but not socially, thus we can define them as "lone achievers."

- **Cluster 3** (28.23%): This cluster of learners put in the lowest effort (around an average active_time of 3.6 hours and n_active_days of 5.6 days). Moreover, they have the lowest exploration metrics and they did not engage socially with other peers. Their interaction with the quest system had very low quest_focus and low p_completed. Therefore, this cohort of learners put in the least effort and had the lowest engagement with any of the dimensions, hence we define them as the "non-engaged learners."

- **Cluster 4** (15.33%): As we already reported, this Cluster had similar activity metrics to Cluster 2 but they had a lower value of p_completed, and they were very active socially, whereas learners from Cluster 2 did not engage socially at all. In fact, they have the highest value of n_zone_chats, so they have been talking publicly very often. However, the type of social activity they performed differs from the learners of Cluster 1, in that learners of Cluster 1 engaged more through parties and private messaging, while learners from Cluster 4 engaged more through the public chat. This might explain another key difference in chars_per_chat; the kind of messaging used by Cluster 4 learners was much
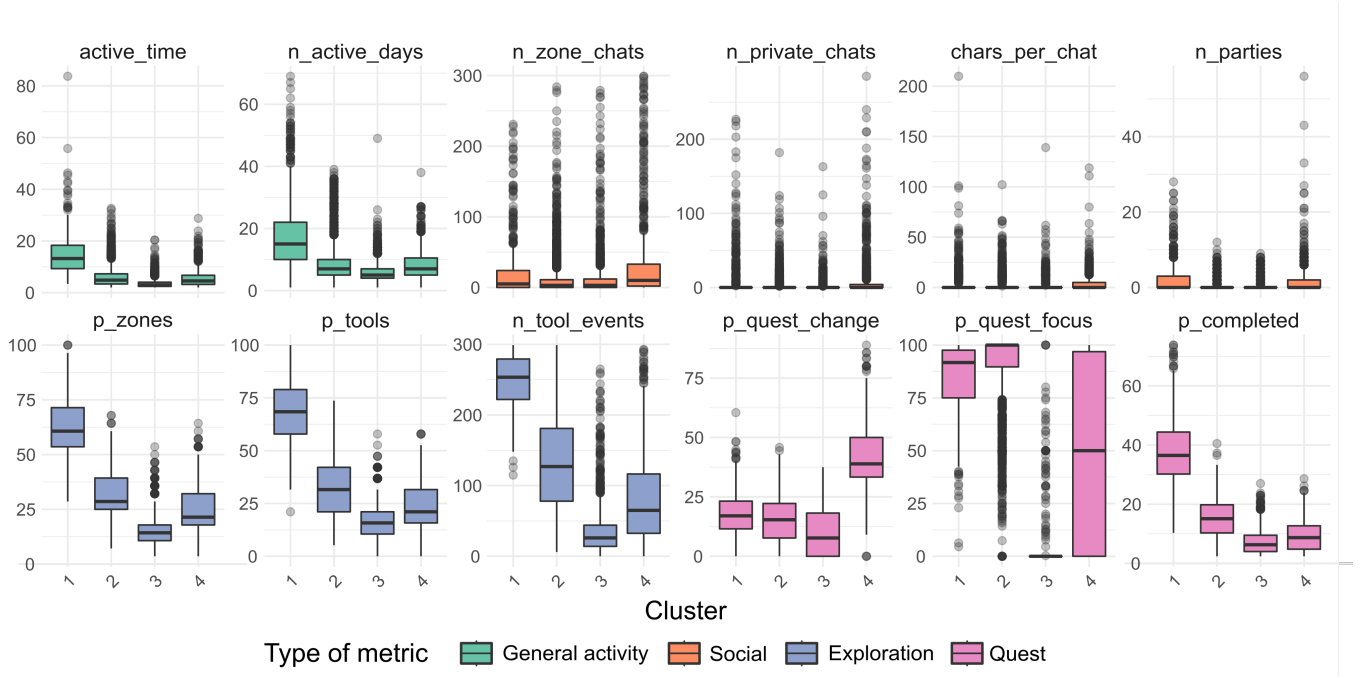
Fig. 7. Boxplot visualization of distribution of all metrics separated by cluster. The two performance metrics were not used as part of the clustering intput variables.

shorter in terms of character length, and potentially more trivial. Moreover, learners from Cluster 4 have the highest `p_quest_change` and much lower `p_quest_focus` than learners from Clusters 1 and 2, therefore their interaction with the quest system likely lacked focus and they were mostly exploring. The last tipping point is that they have the lowest `p_correct` of all clusters. Consequently, learners from Cluster 4 engaged socially and explored the world of *Radix*, however, they did not seriously advance or engage with the quest system, and we therefore define them as "social explorers."

Next, we look into how these clusters are distributed among the three cross-sectional variables that we have explored so far. We report these results in Table I, where we can see the positive significance of all of the Chi-squared tests, which indicates that the cross-sectional variables have an influence in how clusters are distributed. Each one of the columns contains the percentage of learners in each cluster for that value; this allows us to compare the percentage frequency of learners in each cluster for each cross-sectional variable value. First, in terms of account type, we can see that the percentage of out-of-school learners within the non-engaged cluster is higher by 20% and that the social explorer profile is a more frequent cluster for in-school students. For gender, as a cross-sectional variable, the social explorer cluster is more common for male students while lone achiever is more common for the female. Finally, for class grade, it is a bit harder to detect these kinds of qualitative changes, but we can see how the lone achiever profile is more common from 8th to 10th grade and that non-engaged learners are most common in 6th, 11th, and 12th grade. *Radix* was designed to present players with multiple entry points and multiple layers of interaction, with

a balance between guided quests and open-ended exploration. The emergence of these four clusters which are distinct but also have overlapping characteristics shows that players are in fact bringing their unique play styles and learning preferences into the game experience, and that the environment enables them to demonstrate that.

## VII. DISCUSSION

### A. Interpretation and Application of Results

This work is motivated by the fact that GBL has been gaining importance over the last decade and one of its main virtues is the capacity to keep students deeply engaged with their learning process. However, relatively few studies have focused on how to measure the engagement of students based on in-game metrics in the context of a learning game. Therefore, our motivation in this work has been to propose a multidimensional view of engagement, where some learners might engage with some dimensions and other learners with others in the context of GBL. We proposed a transverse dimension in terms of global activity levels, and then qualitative dimensions for exploration, social, and quest activity. We argue that although we have designed the metrics for each dimension targeting *Radix*, these dimensions are generic enough that the model could be applied with some adaptations to other learning games and that our framework could be adapted in a different context as follows:

- The general activity would measure the time or effort invested in the game.
- The exploration dimension measures how many and often the game activities/features have been used.
- The quest dimension can be mapped to the interaction with learning activities, content or achievement.

TABLE I
CLUSTER PERCENTAGE FREQUENCIES BY EACH ONE OF THE CROSS-SECTIONAL VARIABLE VALUES

| Cluster | Account type $(X^2 = 114, p < e^{-16})$ | | Gender $(X^2 = 46, p < e^{-16})$ | | Class grade $(X^2 = 359, p < e^{-16})$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | In-school student | Out-of-school learner | Male | Female | $6^{th}$ | $7^{th}$ | $8^{th}$ | $9^{th}$ | $10^{th}$ | $11^{th}$ | $12^{th}$ |
| 1: Integrally engaged | 9.25% | 12.92% | 9.70% | 8.70% | 14.11% | 8.75% | 4.36% | 9.75% | 7.88% | 5.58% | 10.57% |
| 2: Lone achiever | 47.61% | 38.53% | 43.99% | 52.03% | 35.05% | 43.31% | 48.91% | 65.81% | 54.18% | 37.56% | 43.09% |
| 3: Non-engaged | 26.76% | 45.66% | 27.37% | 26.01% | 37.66% | 26.41% | 28.00% | 15.71% | 18.65% | 37.06% | 29.27% |
| 4: Social explorer | 16.38% | 2.90% | 18.94% | 13.26% | 13.18% | 21.53% | 18.74% | 8.73% | 19.29% | 19.80% | 17.07% |

- The social dimension measures if students socialized with other peers (and if the game does not contain a social component, a multimodal approach using audio and video in classrooms could be attempted).

Some other contexts might offer simpler or more complex multidimensional models.

We believe that in these new forms of learning and assessment, it is crucially important that we can not only measure to what extent students are engaging with these open-ended environments, but also characterize in what ways they are engaging. As education shifts its focus from content knowledge to skills and competencies, more attention needs to be paid to how those skills look when applied to authentic contexts, and how to support different learners in building those skills. Formative assessment tools that can paint a picture of how students engage and how their engagement changes over time can be highly informative for teachers to guide their students, and for students themselves to reflect on their own affinities and goals. For example, a student might be categorized as adopting a lone achiever approach in the first semester of a biology course, and if some of the learning goals for the year are collaboration and communication, the teacher might elicit examples of how other students are collaborating, highlighting how essential these competencies are in the workplace, and helping the student reflect on why they may not be as comfortable working with classmates. As the school year goes on, the teacher could get regular progress updates in the form of dynamic profile characterizations. If the teacher sees that, after a certain unit, the student has shifted to a stronger social explorer profile, the teacher can then recognize the student's efforts to leave their comfort zone, build new skills, and help the student set new goals for continuing to build these and other competencies. With this richer model of formative assessment, we are not judging different students as further ahead or falling behind; rather we can recognize their differences and meet each learner where they are. Tools for game-based learning that find patterns in player behavior could be built based on the methods described in this study, and would enable teachers to encourage mid-course play-style changes to broaden student learning, as well as enabling educators to value and support more abstract competencies and reflection. Combining this with more traditional data on achievement and content-specific ability could provide teachers with a much more well-rounded picture, enabling them to teach the whole learner.

The insights we get from the clustering methods described can also be valuable for a data-driven design approach. In the case of *Radix*, designers wanted the game to feel personalizable in a variety of ways. One is that there are different ways to solve a given quest depending on the tools and strategies players choose to use, and those pathways have in fact been observed in player data on genetics quests. Another area of player choice is in varying goals: players can choose to complete sequences of quests, or explore the world according to their own interests. The results of this study show that is in fact occurring to some degree, and depending on the pedagogical goals, the game mechanics and quest designs could then be tweaked to adjust the balance of these play styles. Beyond *Radix*, these methods could inform the design of future learning environments that might have specific intentions about creating an experience that was very social and collaborative, or very exploratory for example. Connecting the interaction design features that facilitate and motivate players to have certain experiences enables the learning design community to create more targeted interventions that build certain types of skills. By the same token, correlating types of interactions and features with experiences that elicit relevant behaviors and cognitive processes is of interest to learning science researchers who want to better understand how people learn skills and practices that have not been thoroughly studied. We can see from these examples that the results of clustering analysis of player actions can be useful for a wide variety of purposes and for multiple stakeholders, including game designers, learning scientists, educators, and learners themselves.

### B. Learning and Theoretical Implications

Beyond individual student experiences and reporting and feedback to stakeholders, this work has implications for understanding how students learn from and react to game environments more broadly. Multiple authors have brought up how badly designed virtual learning environments can trigger undesired behaviors from students. Cheating using the learning feedback from the platform in massive open online courses [38], gaming the system in intelligent tutoring systems [39], or how extrinsic motivators like badges can make students be more interested in an external reward than in actually learning [40] are all examples of off-task behaviors that have been shown to often lead to poorer performance on learning outcome measures.

The kind of analysis presented here is important for learning in at least three ways. First, it is important for the design of learning environments, as measures of engagement can allow us to better understand the degree to which the design of the game is aligned with actual game play and to better

understand the ways that different learners play the game. As engagement is important to learning, the design of educational games can be improved through better adjusting games to players in order to ensure that it is appropriately engaging for the target audience. Second, while some of the measures of player engagement provided here seem to be better than others, providing a more complex view of player engagement that accounts for different ways or times that players may be engaged is necessary for individualizing learning, either through the design of the game or the implementation model, which can have a big impact on students' engagement and their learning experience. Measures that provide increasing nuance to player activity can be useful for being more fair regarding what "counts" as good or engaging game play. Finally, engagement is only one of the constructs that can be assessed through analysis of in-game play, and other theoretical approaches using similar methods might yield interesting results. Attention, for example, has been shown to mediate or improve in the context of some game-based learning environments (e.g., [41]), and further work linking attention to in-game activity in this way may be helpful for connecting attention to the learning of content embedded within a game.

In this study, we have found four clear engagement profiles that we have defined as: "integrally engaged," "lone achiever," "social explorer," and "non-engaged learners." Some of these profiles have similarities with Bartle's taxonomy [42] for Multi-User Dungeons (MUDs), where he reported four main player profiles: achievers, killers, socializers, and explorers. Bartle's taxonomy and other studies of player types [26] based on psychological constructs measured through surveys taken outside of the context of play have recently been applied to in-game telemetry measures and learning outcomes. For example, [43] found that alignment between game players' preferences and game type was positively associated with gains in student interest, supporting the customization of game play to player preference in education contexts. While Bartle and others have assumed that these dimensions are independent, our view is that at a higher level or perhaps as enacted during game play, we can find profiles of players that engage with multiple dimensions at the same time. For example, the social explorer is a player who interacts with the world and other players simultaneously. This dynamic and multi-layered perspective on player behavior and player types provides a theoretically provocative avenue for further study, suggesting that at least in some cases, a static and singular view of player types fails to adequately capture how a game is played. It is unsurprising to think that players may vary between multiple different play styles throughout game play, especially as they exercise different preferences or strategies for success.

Further, while more data-intensive studies of differences in player activity have been carried out in the context of commercial games, this prior work is largely in pursuit of commercial aims, including diagnosing design problems [44], improving player retention [45], or improving the player experience [46]. Data-driven differences in player engagement profiles have not yet been well-connected to player learning, documented in the context of educational games or related to other measures and profiles of student engagement [15].

This study provides a first and necessary step for creating this bridge, as differences in game play profiles can be useful for assessing student progress in ways that accommodate such differences, for nudging students out of undesired engagement styles (e.g., disengagement) and modifying the game design to maximize learning opportunities based on differences in individual play styles.

Finally, it is worth noting that the engagement profile clusters found in this study can be influenced by the circumstances under which they played. For example, the time and days played should be considered relative to students' contexts, as parents and teachers likely had significant impact on when and how often students were able to play the game, and this might have influenced "integrally engaged" learners. If teachers encouraged the use of in-game features to communicate with other peers, that might boost the prevalence of "social explorer" learners, instead, if teachers recommended students to be quiet and focused, or to interact with their peers through talking, this could make the "lone achiever" profile more common. Finally, if teachers recommended to use the game only as a supplementary learning environment, the "non-engaged" profile could be frequent in that class. Therefore, teachers and parents need to consider how they present and structure time for game-based learning, and in this study it is unclear how external factors like time allowed and directives given influence engagement, or vice-versa. The measures developed here would enable future studies on *Radix* or other games to shed light on the relationship between a selected implementation model and engagement.

### C. Technical Contributions and Technological Implications

On a more granular level, research on learning through games and classroom implementation of game-based learning must be supported with specific data analysis methods and assessment metrics. One challenge of conducting game learning analytics is being able to find ways to extract meaningful information from large datasets that sometimes collect all interactions of players with the game [47]. In this work, we have defined a number of engagement metrics in different dimensions that we hope will inspire others grappling with how exactly to use the oceans of data being collected. Here we explain the design and selection process that we have gone through in order to develop the measure of engagement.

Our findings suggest that people engaging in games for learning might interact differently depending on the characteristics of the game feature. Player differences and preferences typically accounted for in entertainment game design should also be taken into account as part of the design and development of games for learning. One way to do so is by applying a top down approach using Evidence-Centered game Design (ECgD) [48], where we can define a number of dimensions that we want to measure with the game, design features that can generate evidence of those dimensions, and then mechanics in the game that will facilitate players to interact with such features [49]. Based on this research, you should expect that some players will tend to engage more naturally with some of the dimensions than others, and so,

designers could introduce game mechanics to try to balance that if desired, e.g., in the case of *Radix*, we could introduce quests that require players to socialize with other peers to be completed, or require them to explore the world of *Radix*. In this way, we can intertwine several dimensions together. One typical application to make these indicators available to instructors is to make dashboards so that they can visualize what students are doing within a virtual learning environment in order to adapt their practice or intervene to help an issue, which is certainly applicable in game-based learning analytics as well [47]. However, some dashboards show very isolated measures and teachers struggle to interpret them [50]. In this direction, a richer multidimensional model is important as it can help to put together a more accurate and nuanced measure of student activity and learning. These dimensions can be helpful for the organization of the interface in different key dimensions to facilitate interpretability, e.g., one previous study on dashboards in online courses organized the interface around three dimensions, "exercise indicators," "video indicators," and "course activity indicators" [51]. This multidimensional study based on game trace data can be perfectly complemented by additional data obtained from other tracking technologies and processed via multimodal learning analytics approaches. We believe that there is a lot of potential especially in the following two directions:

- To improve the quality of the engagement measurements using additional data sources besides game trace data; for example, including electrodermal activity [52] or emotion detection based on gestures [53]. These additional signals can help validate engagement metrics based on trace data so that we can develop more robust models. Additionally, this analysis could discover how biometric and emotion signals change when players interact with the different dimensions of the game.
- To extend the type of player interactions that we are able to capture with data sources from the physical world such as video and audio. This approach can help to augment in-game interactions with physical world interactions, so that we can know if a student is talking to another one, went to play together with a friend at their desk, and so on. This can help validate and improve the trace data measures: for example, it might be the case that some students do not engage socially within the game because they are socializing with their peers in the physical world while they are playing, and we can thus improve the social dimension engagement measures with these additional data sources.

The goal in pursuing this type of multimodal data collection and applying new methods of analysis to game-based learning is to be able to take advantage of trace data to understand how both learners in general and individual learners can be supported in their learning of key competencies and 21$^{st}$ century skills. This work necessitates bringing together specific technical approaches with overarching learning theories, as we have begun to do in this study.

## VIII. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

We believe that supporting learning engagement and evaluating learning experiences depends significantly on the ability to measure learner engagement, which might be especially challenging in open-ended worlds like *Radix*. This study focused on proposing a multidimensional view of engagement, based on social, quest, and exploration qualitative dimensions, and a transverse one regarding general activity. We then explored the results of the implementation based on the main engagement profiles that we defined as "integrally engaged," "lone achiever," "social explorer," and "non-engaged learners." We concluded by extensively discussing the educational applications, and the technological, learning, and theoretical implications. Our findings could be used in the future to design learning games that are appropriate for each cohort and combine multiple dimensions, new learning technologies that can support learner engagement, and new case studies or theoretical research directions encouraged by our results.

We would also like to acknowledge a number of limitations of this work, the foremost being that since this is retrospective exploratory analysis based only on data traces, we cannot establish any confirmatory reasons of the results that we report, where more experimental approaches could be more helpful. In general, we are blind to a number of contextual factors about how teachers implemented and students used *Radix* from class to class, details such as how much in-class time teachers gave to students, how many different sessions, if they worked individually or not, or how the game contents where embedded in the rest of curriculum, are unknown to us; future studies can aim to understand the impact of how teachers implement games in the classroom in how students engage with the game. Additionally, it is worth noting that we used the typical elbow technique to find the appropriate number of clusters and limited the number of metrics per cluster, therefore a higher number of clusters or additional metrics per dimension could provide more granularity in the engagement profiles that could be found. And so, for these reasons, our recommendation is not to focus so much on the exact numbers or statistically significant results, that could vary from one study to another due to different factors, but more on the qualitative findings, where we report on distinct multidimensional engagement profiles, and how these metrics and profiles change across a number of cross-sectional variables. We believe that the lessons learned that we report, can play a significant role in the implementation of case studies, design of new learning games and future research, as we extensively described in the discussion.

As a part of future work, we would like to better connect in-game metrics with player outcomes such as learning gains and self-efficacy. Previous work in the area [54] has used behavioral metrics to predict learning gains and investigate the individual effect of such metrics in how much a student learns; we would like to replicate those analyses in a *Radix* pilot study by using the pre- and post-tests that many school students completed. We would also like to investigate the impact of providing this information to teachers and students, particularly to improve students' awareness of their own learning

processes and teachers' classroom practices. We would also like to delve into investigating if these game environments, used in school classes, can help improve the performance of those students that suffer from anxiety in class, or facilitate the socialization of shy students with other peers through activity in the virtual world. Future work should plan on combining trace data from the game with other external data sources like tracking wearables, audio, or video. The multimodal combination of these different data sources can create much richer and complex models to help understand the interplay between all these signals, and how to use them to improve learning within this educational context.

## ACKNOWLEDGMENTS

## REFERENCES

[1] S. de Freitas, "Learning in immersive worlds: A review of game-based learning," Joint Information Systems Committee, Bristol, UK, Working Paper, 2006. [Online]. Available: http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf

[2] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Comput. Human Behav.*, vol. 54, pp. 170–179, Jan. 2016.

[3] A. I. A. Jabbar and P. Felicia, "Gameplay engagement and learning in game-based learning: A systematic review," *Rev. Educ. Res.*, vol. 85, no. 4, pp. 740–779, Dec. 2015.

[4] R. M. Martey, K. Kenski, J. Folkestad, L. Feldman, E. Gordis, A. Shaw, J. Stromer-Galley, B. Clegg, H. Zhang, N. Kaufman *et al.*, "Measuring game engagement: multiple methods and construct complexity," *Simulat. Gaming*, vol. 45, no. 4-5, pp. 528–547, Aug./Oct. 2014.

[5] J. A. Fredricks, M. Filsecker, and M. A. Lawson, "Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues," *Learn. Instruction*, vol. 43, pp. 1–4, June 2016.

[6] N. Yee, "The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments," *Presence: Teleot. Virt. Environ.*, vol. 15, no. 3, p. 309329, June 2006.

[7] F. Paraskeva, S. Mysirlaki, and A. Papagianni, "Multiplayer online games as educational tools: Facing new challenges in learning," *Comput. Educ.*, vol. 54, no. 2, pp. 498–505, Feb. 2010.

[8] R. M. Carini, G. D. Kuh, and S. P. Klein, "Student engagement and student learning: Testing the linkages," *J. Higher Educ.*, vol. 47, no. 1, pp. 1–32, Feb. 2006.

[9] T. Dary, T. Pickeral, R. Shumer, and Williams, "Weaving student engagement into the core practices of schools: A national dropout prevention center/network," Clemson, SC: National Dropout Prevention Center/Network, Tech. Rep., 2016. [Online]. Available: www.dropoutprevention.org/resources/major-research-reports/studentengagement/student-engagement-2016-09.pdf

[10] R. Pekrun and L. Linnenbrink-Garcia, "Academic Emotions and Student Engagement," in *Handbook of Research on Student Engagement*, S. L. Christenson, A. L. Reschly, and C. Wylie, Eds. Boston, MA: Springer US, 2012, pp. 259–282.

[11] R. B. Miller, B. A. Greene, G. P. Montalvo, B. Ravindran, and J. D. Nichols, "Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability," *Contemporary Educ. Psychol.*, vol. 21, no. 4, pp. 388–422, Oct. 1996.

[12] K. Deater-Deckard, M. Chang, and M. E. Evans, "Engagement states and learning from educational games," *New. Dir. Child Adolesc. Dev.*, vol. 2013, no. 139, pp. 21–30, Spring 2013.

[13] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, Spring 2004.

[14] E. A. Skinner, T. A. Kindermann, J. P. Connell, and J. G. Wellborn, "Engagement and disaffection as organizational constructs in the dynamics of motivational development," in *Handbook of Motivation at School*, ser. Educational psychology handbook series, K. R. Wenzel and A. Wigfield, Eds. Routledge/Taylor & Francis Group, 2009, pp. 223–245.

[15] M.-T. Wang and S. C. Peck, "Adolescent educational success and mental health vary across school engagement profiles," *Develop. Psychol.*, vol. 49, no. 7, pp. 1266–1276, July 2013.

[16] A. Wigfield, J. S. Eccles, J. A. Fredricks, S. Simpkins, R. W. Roeser, and U. Schiefele, *Development of Achievement Motivation and Engagement*. American Cancer Society, 2015, ch. 16, pp. 1–44.

[17] H. M. Marks, "Student engagement in instructional activity: Patterns in the elementary, middle, and high school years," *Am. Educ. Res. J.*, vol. 37, no. 1, pp. 153–184, Mar. 2000.

[18] M. Csikszentmihalyi, "Toward a psychology of optimal experience," in *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Springer Netherlands, 2014, pp. 209–226.

[19] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing," *J. Exp. Social Psychol.*, vol. 45, no. 4, pp. 624–634, July 2009.

[20] E. N. Wiebe, A. Lamb, M. Hardy, and D. Sharek, "Measuring engagement in video game-based environments: Investigation of the User Engagement Scale," *Comput. Human Behav.*, vol. 32, pp. 123–132, Mar. 2014.

[21] D. K. Mayes and J. E. Cotton, "Measuring engagement in video games: A questionnaire," *Proc. Human Factors and Ergonom. Soc. Annu. Meeting*, vol. 45, no. 7, pp. 692–696, Oct. 2001.

[22] K. Kiili and T. Lainema, "Foundation for measuring engagement in educational games," *J. Interact. Learn. Res.*, vol. 19, no. 3, pp. 469–488, July 2008.

[23] P. Bouvier, E. Lavoué, and K. Sehaba, "Defining engagement and characterizing engaged-behaviors in digital gaming," *Simul. Gaming*, vol. 45, no. 4-5, pp. 491–507, Aug. 2014.

[24] V. Riemer and C. Schrader, "Impacts of behavioral engagement and self-monitoring on the development of mental models through serious games: Inferences from in-game measures," *Comput. Human Behav.*, vol. 64, pp. 264–273, Nov. 2016.

[25] A. R. Cano, B. Fernández-Manjón, and A. J. García-Tejedor, "Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities," *Brit. J. Educ. Technol.*, vol. 49, no. 4, pp. 659–672, Apr. 2018.

[26] N. Yee, "The gamer motivation profile: What we learned from 250,000 gamers," in *Proc. Annu. Symp. Compt.-Human Interact. Play*. ACM, 2016, pp. 2–2.

[27] L. E. Nacke, M. N. Grimshaw, and C. A. Lindley, "More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game," *Interact. Comput.*, vol. 22, no. 5, pp. 336–343, Sept. 2010.

[28] F. Ke, K. Xie, and Y. Xie, "Game-based learning engagement: A theory-and data-driven exploration," *Brit. J. Educ. Technol.*, vol. 47, no. 6, pp. 1183–1201, Nov. 2016.

[29] H. Jang, E. J. Kim, and J. Reeve, "Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model," *Learn. Instruction*, vol. 43, pp. 27–38, June 2016.

[30] S. Järvelä, H. Järvenoja, J. Malmberg, J. Isohätälä, and M. Sobocinski, "How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement?" *Learn. Instruction*, vol. 43, pp. 39–51, June 2016.

[31] "The Radix Endeavor Game – Web Site," http://www.radixendeavor.org/, accessed: 2020-01-01.

[32] "The Radix Endeavor – YouTube Channel," http://www.youtube.com/user/radixendeavor, accessed: 2020-01-01.

[33] National Research Council, *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press, 2013. [Online]. Available: https://www.nap.edu/catalog/18290/next-generation-science-standards-for-states-by-states

[34] National Governors Association Center for Best Practices, Council of Chief State School Officers, *Common Core State Standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010. [Online]. Available: http://www.corestandards.org/

[35] J. Clarke-Midura, L. Rosenheck, and J. Groff, "Designing games for learning and assessment: the Radix Endeavor," in *Proc. 14th Int. Conf. Interact. Des. Children*, ser. IDC'15. New York, NY, USA: ACM, 2015, pp. 343–346.

[36] J. A. Ruiprez-Valiente, P. J. Muoz-Merino, C. D. Kloos, K. Niemann, M. Scheffel, and M. Wolpers, "Analyzing the impact of using optional activities in self-regulated learning," *IEEE Trans. Learn. Technol.*, vol. 9, no. 3, pp. 231–243, July 2016.

[37] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-Means clustering," *Int. J. Adv. Res. Comput. Sci.*, vol. 1, no. 6, pp. 90–95, Nov. 2013.

[38] G. Alexandron, J. A. Ruipérez-Valiente, Z. Chen, P. J. Muñoz-Merino, and D. E. Pritchard, "Copying@Scale: Using harvesting accounts for collecting correct answers in a MOOC," *Comput. Educ.*, vol. 108, pp. 96–114, May 2017.

[39] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner, "Off-task behavior in the cognitive tutor classroom: When students game the system," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, ser. CHI'04. New York, NY, USA: ACM, 2004, pp. 383–390.

[40] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. Delgado Kloos, "Detecting and clustering students by their gamification behavior with badges: A case study in engineering education," *Int. J. Eng. Educ.*, vol. 33, no. 2-B, pp. 816–830, Mar./Apr. 2017.

[41] M. Dye, C. S. Green, and D. Bavelier, "Increasing speed of processing with action video games," *Curr. Dir. Psychol. Sci.*, vol. 18, pp. 321–326, May 2009.

[42] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit MUDs," *J. MUD Res.*, vol. 1, no. 1, p. 19, 1996.

[43] B. Manero, J. Torrente, C. Fernndez-Vara, and B. Fernndez-Manjn, "Investigating the impact of gaming habits, gender, and age on the effectiveness of an educational video game: An exploratory study," *IEEE Trans. Learn. Technol.*, vol. 10, no. 2, pp. 236–246, Apr. 2016.

[44] C. Pruett, "Hot failure: Tuning gameplay with simple player metrics," *Game Developer Magazine*, vol. 19, no. 9, Dec. 2010.

[45] E. Andersen, "Optimization of online games through telemetry and large-scale experimentation," University of Washington Center for Game Science, 2012. [Online]. Available: https://www.gdcvault.com/play/1016149/Optimization-of-Online-Games-through

[46] D. Charles and M. Black, "Dynamic player modeling: A framework for player-centered digital games," in *Proc. Int. Conf. Compt. Games: Artif. Intell., Des. and Educ.*, 2004, pp. 29–35.

[47] M. Freire, Á. Serrano-Laguna, B. M. Iglesias, I. Martínez-Ortiz, P. Moreno-Ger, and B. Fernández-Manjón, *Game Learning Analytics: Learning Analytics for Serious Games*. Springer International Publishing, 2016, pp. 1–29.

[48] R. J. Mislevy, S. Corrigan, A. Oranje, K. DiCerbo, M. I. Bauer, A. von Davier, and M. John, *Psychometrics and game-based assessment*. Routledge New York, Aug 2016, pp. 23–48.

[49] Y. J. Kim, J. A. Ruipérez-Valiente, T. Philip, R. Louisa, and E. Klopfer, "Towards a process to integrate learning analytics and evidence-centered design for game-based assessment," in *Companion Proc. 9th Int. Learn. Analytics and Knowledge Conf.* ACM, 2019, pp. 204–205.

[50] I. Jivet, M. Scheffel, H. Drachsler, and M. Specht, "Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice," in *Data Driven Approaches in Digital Education*, É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, and M. Pérez-Sanagustín, Eds. Springer International Publishing, 2017, pp. 82–96.

[51] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, J. A. Gascón-Pinedo, and C. D. Kloos, "Scaling to massiveness with ANALYSE: A learning analytics tool for Open edX," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 6, pp. 909–914, Dec. 2016.

[52] S. Huynh, S. Kim, J. Ko, R. K. Balan, and Y. Lee, "EngageMon: Multi-modal engagement sensing for mobile games," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018.

[53] A. Psaltis, K. Kaza, K. Stefanidis, S. Thermos, K. C. Apostolakis, K. Dimitropoulos, and P. Daras, "Multimodal affective state recognition in serious games applications," in *IEEE Int. Conf. Imag. Syst. Techn.* IEEE, Oct. 2016, pp. 435–439.

[54] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. D. Kloos, "Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators," *Expert Syst.*, vol. 35, no. 6, p. e12298, Dec. 2018.

**José A. Ruipérez-Valiente** (M'20) completed his B.Eng. and M.Eng. in Telecommunications at Universidad Católica of 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 at Institute IMDEA Networks in the area of learning analytics. Currently he is a postdoctoral associate at the CMS/W department at MIT where he is part of the Teaching Systems Lab and the Playful Journey Lab, while also collaborating with the Education Arcade.

**Matthew Gaydos** is a design researcher at the Playful Journey Lab. He completed his Ph.D. from the University of Wisconsin - Madison where he worked on developing and studying digital educational games for middle school. He has worked as a Learning Scientist at the National Institute of Education in Singapore, as a policy intern at the U.S. Office of Science and Technology - Policy, and as an Assistant Professor at the State University of New York - Korea. He currently works as a design researcher at MIT where he is studying and designing games for play-based learning from a Learning Sciences perspective.

**Louisa Rosenheck** is a designer and researcher of educational technologies in the MIT Education Arcade. 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.

**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 learning environments.

**Eric Klopfer** is Professor and Director of the Scheller Teacher Education Program and The Education Arcade at MIT. He is also a co-faculty director for MIT's J-WEL World Education Lab. He is the co-author of the books, "Adventures in Modeling," "The More We Know," and the recently released "Resonant Games," as well as author of "Augmented Learning." His lab has produced software (from casual mobile games to the MMOG *The Radix Endeavor*) and platforms (including StarLogo Nova and Taleblazer) used by millions of people, as well as online courses that have reached hundreds of thousands. Klopfer is also the co-founder and past President of the non-profit Learning Games Network.