Language Learning & Technology
ISSN 1094-3501
2023, Volume 27, Issue 1
pp. 1–26

ARTICLE

Applying educational data mining to explore individual experiences in digital games

Frederick J. Poole, Michigan State University

Jody Clarke-Midura, Utah State University

Abstract

Research involving digital games and language learning is rapidly growing. One advantage of using digital games to support language learning is the ability to collect data on students learning in real time. In this study, we use educational data mining methods to explore the relationship between in-game data and elementary students' Chinese language learning. Thirty-six students in the sixth grade played a digital game for eight 25-minute sessions as part of their Chinese Dual Language Immersion classroom instruction. We used classification and regression tree analyses and cluster analyses to explore how in-game indicators, such as battles, time spent reading a text, and the use of an in-game glossing tool are associated with language learning and change in affect. The results indicate that time on task and use of the glossing tool were the most important variables in determining language learning gains. We also identified four subgroups of gameplay styles. While there were no significant differences in learning or affective factors based on the subgroups, these gameplay styles allow for a more individualized approach to analyzing learning within digital environments.

Keywords: *Educational Data Mining, Game-based Language Learning, Chinese as a Foreign Language*

Language(s) Learned in This Study: *Chinese as a Foreign Language (Mandarin)*

APA Citation: Poole, F. J., & Clarke-Midura, J. (2023). Applying educational data mining to explore individual experiences in digital games. *Language Learning & Technology*, *27*(1), 1–26. http://hdl.handle.net/10125/73522

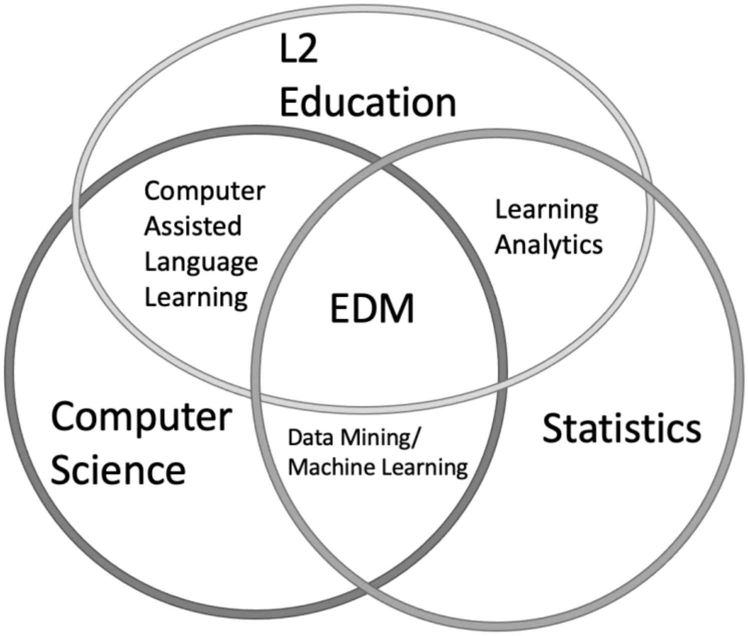
Introduction

Research investigating the application of digital games for foreign/second language (L2) learning and teaching has steadily been on the rise over the last 20 years (Cornillie et al., 2012; Poole & Clarke-Midura, 2020). One of the advantages of doing research with digital games is the ability to collect large amounts of data related to gameplay and learning. However, a recent review of digital games used for L2 learning noted that only 28.5% of studies in the review took advantage of data collected during gameplay (Poole & Clarke-Midura, 2020). Many of these studies used data in the form of answers to in-game quizzes (e.g., Erhel & Jamet, 2016; Ming et al., 2013) or to collect chat dialogue to be later analyzed manually (e.g., Bytheway, 2014; Rama et al., 2012). While these applications of game data can be beneficial to researchers, such data could be further explored using educational data mining (EDM) techniques.

Baker (2010) defines EDM as the "scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in" (pp. 112–113). In other words, EDM is not simply the analysis of large data sets, but additionally the exploration of *unique* data sets from *educational settings*. EDM makes use of computer science tools to wrangle and explore messy data sets and statistical techniques to discover potential patterns within those data (DeFreitas & Bernard, 2014). Figure 1, adapted from Romero and Ventura (2013), illustrates the multiple components associated with EDM.

Figure 1

Educational Data Mining: Adapted from Romero & Ventura (2013)



The application of EDM approaches has the potential to change how we explore learning and affective factors by expanding not only the type of data we collect but when and how we collect it. Such approaches can leverage user-generated data (e.g., choices, actions, and time spent on task) in digital environments like Learning Management Systems (LMS), online forums, and digital games as a proxy for behavioral and cognitive change (e.g., Gibson & Clarke-Midura, 2015; Gobert, et al. 2012).

While EDM is closely associated with Learning Analytics (LA), there are key differences. Both fields are concerned with using large, user-generated data to explore patterns within educational settings (Siemens, 2013). However, researchers have noted that EDM typically focuses on first using automated approaches to data analysis and then relying on human judgment to interpret findings, while LA focuses on using human judgment to first define a model that is then automatized across a data set (Baker & Inventado, 2014). The automated approaches of EDM are advantageous when exploring large data sets in which the relationships between variables are less clear (DeFreitas & Bernard, 2014).

In a recent call for a focus on EDM and LA approaches in L2 learning, Reinders (2018) argues that big data collected from LMS are not only valuable for L2 researchers but could also inform L2 teacher practices in the classroom. As a field, we need to understand what data is useful and associated with L2 learning. Currently, there is a dearth of studies exploring the use and application of big data for L2 learning (Reinders & Lan, 2021). The present study applies EDM approaches to data derived from a digital game designed to support Chinese as a foreign language learning in elementary classrooms. Specifically, this study explores how in-game behaviors are associated with gameplay styles and learning, as measured by a pre- and post-test assessment for vocabulary and reading comprehension.

Literature Review

Educational Data Mining

EDM applies data mining approaches and techniques to data derived from an educational context. Siemens and Baker (2012) identified five general categories of approaches applied in EDM: (a) prediction, (b) clustering, (c) relationship mining, (d) discovery with models, and (e) distillation of data for human judgment. In addition, EDM is concerned with extracting and processing raw data to make sense of a learning environment (Romero & Ventura, 2010). It involves messy data sets, so it is important to explicitly detail how such data is extracted and transformed before analysis. Regarding EDM and L2 studies, Godwin-Jones (2017) highlights that using data derived from digital environments is not new for L2 researchers but that it is the size of datasets being produced that is novel. He argues that this is largely due to the increased use of LMSs, personalized learning environments, and the creation of other digital software for learning. It is precisely these large data sets comprising user-generated data from digital environments that demand EDM approaches. This is because the data is not readily formatted for analysis, and potential relationships between variables within such data sets are unclear (Jacob et al., 2015).

In recent years, a few studies have started to apply these approaches to L2 contexts. In a study involving massive open online courses (MOOC), Martín-Monje et al. (2018) used classification trees to explore the relationship between Gardner's intelligence modalities and engagement with language MOOC objects (e.g., videos, and articles). The authors found that overall engagement with the course was associated with success and that a majority of the MOOC users were classified as ‘viewers’ (only watched course videos) rather than ‘all-rounders’ (those who completed all tasks). In another study exploring data from an LMS, Foung (2019) used classification decision trees to develop a model to predict English learners who were at risk of dropping out of a course. The author created models with approximately 90% accuracy in detecting at-risk students and argued that they benefit teachers and administrators by providing real-time analytics on learners. In a study that investigated student behavioral patterns in a digital reading environment, Lee et al. (2019) used clustering approaches to explore unique ways that students interacted with a textual glossing system. In an earlier study with the same data, the authors found that students using the glossing system reported the highest post-test scores on average compared to two other conditions (Lee et al., 2017). However, when using cluster analysis to explore potentially hidden patterns in the 2019 follow-up study, the authors found that L2 proficiency was a strong determinant of the effectiveness of the glossing system.

While the aforementioned studies have illustrated how such techniques can help uncover hidden relationships between learning and student behaviors in digital learning environments, EDM approaches are not without critique. One of the most obvious areas of concern is privacy and ethics (Baker & Inventado, 2016). By using EDM approaches, researchers can track student behavior unobtrusively, but this means that students may not always be aware that their data is being collected nor know how and why it is being used. Additionally, EDM researchers have noted that there is a potential for misinterpretation of results due to human judgment, inaccurate models, and/or over-analysis (Papamitsiou & Economides, 2014). Thus, while it is promising that data can potentially be used to identify students at risk or opportunities for learning, misclassification of a learner may have severe consequences. Similar to these concerns, in discussing the shortcomings of learning analytics, Youngs (2021) notes that while data can illuminate the presence or absence of an error or a skipped question, such data often cannot provide insight into why an error was made or a question was skipped. Researchers using EDM approaches must be cognizant of what can and cannot be inferred from the data. Finally, Fischer et al. (2020) point out that it is also important to examine potential bias in the algorithms used for analysis. In research, there must be transparency about the classification of students, the use of data, and the inferences that are made. As in any other approach, biases and alternate interpretations need to be checked.

L2 Gaming Research and Log Data

Data collected via digital games have been used in a variety of ways. The majority of studies examined chat logs to investigate the language used within a game (e.g., Bytheway, 2014; Rama et al., 2012). Some studies designed quiz-like tasks into games and exported the data/scores for analysis (e.g., Cornillie et al., 2012; Erhel & Jamet, 2016).

Other researchers have used in-game variables as indicators for the effect that playing a digital game has on learning outcomes and L2 proficiency. For example, in a study exploring the effect of time spent on reading a text, Collentine (2011) found that time spent on tasks was significantly associated with outcomes on a writing assessment after playing a digital game designed for L2 learners of Spanish. In a study investigating L2 learning in *Everquest II*, Rankin et al. (2006) found that a variable counting the number of messages produced was significantly associated with more advanced English speakers.

A few studies have applied EDM techniques to digital games used for L2 learning to explore how gameplay behavior is related to affect and learning. For example, Hwang et al. (2017) used a progressive sequential analysis to compare gameplay sequences of EFL learners with high and low L2 learning anxiety. They found that learners with higher levels of anxiety engaged in more complex forms of learning than less anxious students. They further found that high-anxiety learners reviewed more vocabulary and acquired more relevant knowledge before completing tasks. Hsiao et al. (2017) explored behavioral patterns around players interacting with new vocabulary words in a virtual environment. Players were told to learn 30 words while interacting with the environment. Through data visualization techniques, the authors found differences in learning strategies between high- and low-achieving learners. They concluded that low-achieving learners tended to click on words randomly or use the nearest-neighbor approach in which they clicked on vocabulary words that were close to each other. These EDM approaches allow researchers to explore relationships between individual gameplay styles and learning and affect.

Learning Gains in Digital Games

A recent review found that most studies exploring language learning within digital games tend to focus on vocabulary development (Poole & Clarke-Midura, 2020). Such studies tend to show that learning vocabulary via digital games is superior to non-game settings (e.g., Calvo-Ferrer, 2017; Wu & Huang, 2017). However, none of these studies consider how individual differences in gameplay style may impact the observed vocabulary learning gains. The review identified three studies that used digital games to develop reading comprehension. Two of the studies noted that students employed a variety of reading strategies when reading text via digital games (Dourda et al., 2014; Poole et al., 2018). The third study noted that students achieved higher reading comprehension scores after using a digital game in the school's computer lab in lieu of traditional instruction (Suh et al., 2010). However, similar to studies exploring vocabulary gains, these studies did not explore the potential relationship between individual gameplay styles and observed strategy use or learning gains. This is important because games are interactive and often allow players to engage with the material in a variety of ways. As noted in the previous section, EDM techniques may allow researchers to explore variation at the individual level.

Responding to previous calls, the present study builds on past research implementing data mining techniques in L2 contexts that use digital games. Specifically, we use EDM approaches to explore how in-game variables are associated with learning gains, and further, how individual gameplay styles may be associated with learning and/or change in affect. Our research questions are the following:

1. What, if any, in-game variables are associated with a change in vocabulary knowledge and reading comprehension, as measured by the pre- and post-assessments?
2. How are in-game playstyles associated with affective factors and/or vocabulary knowledge and reading comprehension gains, as measured by the pre- and post-assessments?

Methods

Materials

The digital game used in the present study was designed by the author to promote Chinese vocabulary learning and reading comprehension via an in-game glossing system and in-game tasks that are driven by text-based dialogue. In the game, students took on the role of an adventurer who embarks on a quest to save the last dragon in China. As students carried out quests, they engaged in battles (against baddies), collected items, and met non-player characters (NPC) who provided information, presented tasks, and directed students toward the last dragon (see Figure 2).

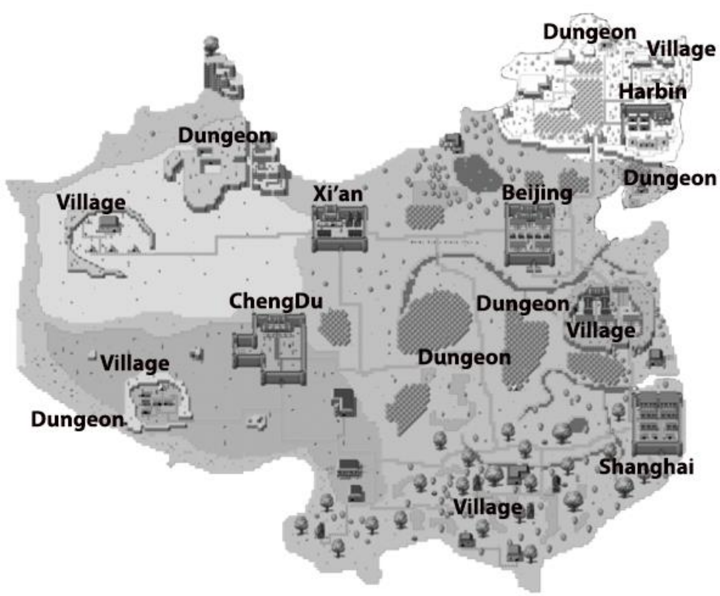
Figure 2

Example Quest

The game world consists of five major Chinese cities and several fictional villages and dungeons placed in proximity to the cities. The game world was designed to resemble the geographic shape of China with cities located in their approximate real-world locations (see Figure 3). A short walkthrough video of the game can be found via this link.

Figure 3

Overworld Map of Game



Participants and Setting

Data were collected from a Chinese Dual Language Immersion (DLI) public school across two DLI classrooms in the intermountain west region of the United States. Participants played the game in class as part of their Chinese instruction. We focused on sixth graders because the content of the game is more linguistically appropriate for older elementary students. For example, the game utilizes pinyin as a means of support for unknown words, and the participants had been using pinyin since third grade. They also had classroom laptops and familiarity with using digital environments as part of their DLI instruction. According to an in-class assessment based on the ACTFL scale, most of the participants were between novice-high and intermediate-low proficiency levels.

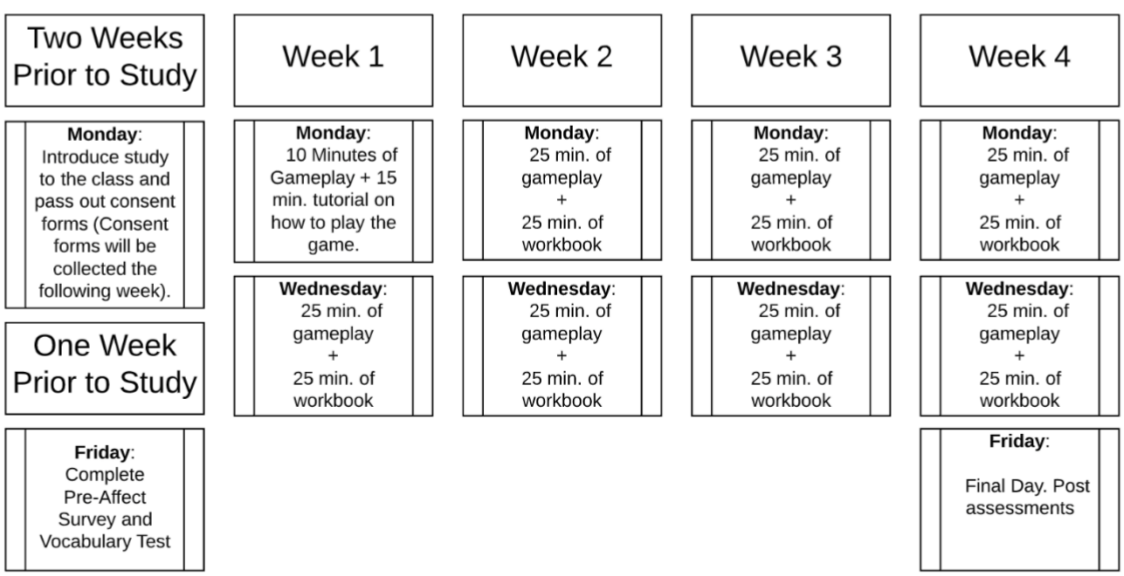
The two classrooms had 19 and 21 students respectively. The students' age range was between 10 and 12 years (M = 11.05), and they were evenly distributed by gender (17 boys and 19 girls). All procedures involving participants complied with the provisions of Utah State University's internal review board and were approved by them. The consent of parents and 36 participants in the study was obtained; however, four students (one in Class A and three in Class B) did not return the consent form. Students were informed that their data would be collected and analyzed after gameplay. Since this was part of classroom instruction, students without consent were still able to play the game, but we did not collect their data. Two students in Class B did not take the pre-assessments and one student in both classes did not take the post-assessments. Thus, there was only complete data for 32 students. Of the 32 students, 30 spoke English as their first language, and two spoke Chinese as their first language.

Procedures

Three days prior to gameplay, the pre-assessment was administered with a randomized version of the vocabulary and reading comprehension assessments via paper. Participants played the game for 50 minutes (two 25-minute sessions) and completed supplemental materials for 50 minutes (two 25-minute sessions) per week over 4 weeks (3 hours of gameplay). Two days following the final gameplay session, participants completed the post-assessment. See Figure 4 for an overview of the study procedures.

Figure 4

Study Procedures



Data Collection

Data for the present study came from four sources: pre- and post-affect surveys (See Appendix A), pre- and post-vocabulary assessments (See Appendix B), a pre- and post-reading comprehension assessment (See Appendix C), and log files that captured in-game actions, texts read, and choices made.

Affective Survey

The pre-affect survey consisted of background information, two questions on gaming background, and seven items regarding Chinese reading anxiety using an 8-point Likert scale. Reading anxiety was added to the affect survey because research indicates that L2 anxiety can influence in-game behaviors (Hwang et al., 2017). The seven items for Chinese reading anxiety were adapted from Saito et al. (1999). The post-affect survey comprised the same seven items concerning Chinese reading anxiety, and ten items about the gaming experience adapted from De Grove et al. (2010). For this analysis, only positive affect, which explores student enjoyment and reading anxiety measures, was used.

Vocabulary Assessment

The pre- and post-vocabulary assessment consisted of 45 words found in the game. The words were selected based on a corpus of gameplay data (created prior to the study) that identified the most frequent vocabulary used in the game. Only words that were glossed were added to the list. Students were prompted to enter the pinyin and the English version of each word in character form. Each item was scored up to 2 points (1 for correct pinyin, 1 for correct English). Half points were awarded if only part of the pinyin or English was correct.

Reading Comprehension Assessment

The external reading comprehension assessment was adapted from the Youth Chinese Test (YCT), an official Chinese proficiency assessment developed by the Confucius Institute, and consisted of ten items. Although the format of the reading comprehension assessment was adapted to reflect the Chinese YCT test, the content was adapted to reflect text that the learners might see in the game. Finally, gain scores for both vocabulary learning and reading comprehension were calculated by subtracting pre-scores from post-scores.

Log Data

Five variables were collected from the in-game log files: (a) *Text exposure* counted the total number of texts that a player was exposed to while playing the game; (b) *Battles* counted the total number of battles that a player engaged in; (c) *Menu-on* counted the total number of times a player accessed the in-game menu to look at a skill, card, or item that was acquired; (d) *Average time reading* tracked how long students spent reading each in-game text in seconds, on average; and (e) *Look up* tracked how many times players used the glossing system in the game to look up words. These variables were selected because (a) they could be extracted from the log data files, and (b) they represented an in-game action that may be associated with in-game text engagement.

Data Analysis

Baker et al. (2021) suggest that perspectives on EDM approaches can be classified by one of (or a combination of) four classic schools of thought: (a) entitative, (b) ontological, (c) existentialist, and (d) essentialist. The authors state that early EDM researchers have mostly followed the entitative school of thought, which is associated with reductionist approaches. Reductionist approaches attempt to break down complex phenomena into smaller components, and then explore the relationship between the identified components. In the present study, we take a reductionist approach by using EDM techniques to identify gameplay patterns (components) that may be associated with learning and/or affect change. In the following sections, we detail how data was collected, processed, and analyzed.

Log Data Processing

In-game data came from log files that captured and timestamped players' actions, texts read, and choices. These data were saved to a JSON file after each session's gameplay and then automatically uploaded to a Mongo database. After the study was completed, all JSON files were pulled from the Mongo database and stored in one master JSON file. This file was then wrangled into a single data frame using primarily *tidyverse* (Wickham, 2017). To extract this data and wrangle it into a data frame, the *knowledge discovery in databases* process was used. Knowledge discovery refers to the "nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (Frawley et al., 1992, p. 58). This is an iterative process that involves five broad phases: (a) selection, (b) pre-processing, (c) transformation, (d) data mining, and finally (e) interpretation/evaluation (Fayyad et al., 1996).

The selection phase refers to the process of identifying data of interest based on the target domain. We adapted and added a JavaScript plugin to the digital game to track in-game actions with timestamps. The pre-processing phase refers to data cleaning and removing missing or irrelevant items from the dataset. Our pre-processing phase primarily focused on transforming data from long format to wide format. When the data was collected, every action was stored in a separate row. Thus, some data needed to be collapsed around events or actions, or removed due to redundancy. For example, *reading a text* occurred across several rows of data. We collapsed actions of reading texts and created a time variable that captured the time spent reading the text. The original data contained 520,812 rows of data. Cleaning and collapsing resulted in 19,781 rows of data (cases). We then summarized the data by learners. Finally, the last two phases, data mining and interpretation, were concerned with choosing data analysis techniques and interpreting the analyses. These will be discussed further in the sections below.

Classification and Regression Tree Analysis

The first research question explores which in-game variables, if any, are important indicators for L2 learning. Learning was defined as gains in vocabulary knowledge and reading comprehension scores as measured by the pre- and post-assessments. We used classification and regression tree (CART) analysis to answer the question. CART analysis is an approach often used in data mining or machine learning, and it employs an algorithm that chooses variables that will reduce the sum of squared errors in a regression model each time a partition is made (a variable is selected). Variables that have little or no effect on the outcome variable are not selected by the algorithm in the analysis. This makes the CART analysis ideal for data exploration that seeks to understand which variables are the most important indicators of a target variable (Song & Lu, 2015).

CART analysis is an ideal approach for analyzing log data. First, variables from in-game data are on very different scales and are likely not normally distributed. Second, the autonomy that students are allotted within the game and the vast number of ways the game can be played make for rich and varied data. For example, some students may interact with 700 passages and read them at quick speeds, while others may read 300 passages at slower speeds. Finally, CART analysis is good at dealing with outliers and missing data, making it ideal for analyzing potentially messy log data (Mendez et al., 2008; Song & Lu, 2015).

The first step in setting up a CART analysis is to identify the outcome variable. In the present study, the two outcome variables were vocabulary learning and reading comprehension gains. The next step is to enter variables that are relevant to the outcome variable. Table 1 explains the five variables, collected via log data, that were entered into the model.

Table 1

Variables Used in CART Analysis

Variable	Description
Text	The total number of texts read by a student
Menu On	The total number of times a student accessed the menu
Battle Start	The total number of times a student engaged in battle
AvTime	The average amount of time a learner spent reading a passage in the game (measured in seconds)
LookUpCount	The total number of times that a learner looked up a word while playing

The algorithm cycles through the predictors and chooses the predictor that will best parse the selected outcome and the cutoff point for that variable that best reduces the sum of squared errors. The purpose is to create the most similar groups according to the outcome variable. This process is repeated until the minimum sum of squared errors is reached. The cutoff point is called the complexity parameter and controls how large the tree grows (Song & Lu, 2015). A cutoff point of .01 is often the default. This would dictate that once the splitting of the outcome variables fails to reduce the sum of squared errors by at least .01, then the model stops. The goal is to find a model that is neither too complex (i.e., overfit) nor too simple. To determine the ideal complexity parameter, a series of trees are built from the smallest to the largest, then the cross-validated error for each tree is examined (Therneau & Atkinson, 1997). Once the cross-validated error begins to increase (instead of decrease) as a result of adding a split, the decision tree should stop building (Williams, 2011). Thus, the goal is to identify the lowest cross-validated error and the complexity parameter that is associated with the lowest cross-validated error. In the present study, a complexity parameter of 0.06 was chosen for the vocabulary learning gains CART analysis, based on examining cross-validated errors from the default model. A complexity parameter of 0.11 was used for the reading comprehension gains CART analysis using the same approach. All analyses were conducted in *R*, using the *Rpart* package (Therneau & Atkinson, 2019).

Cluster Analysis

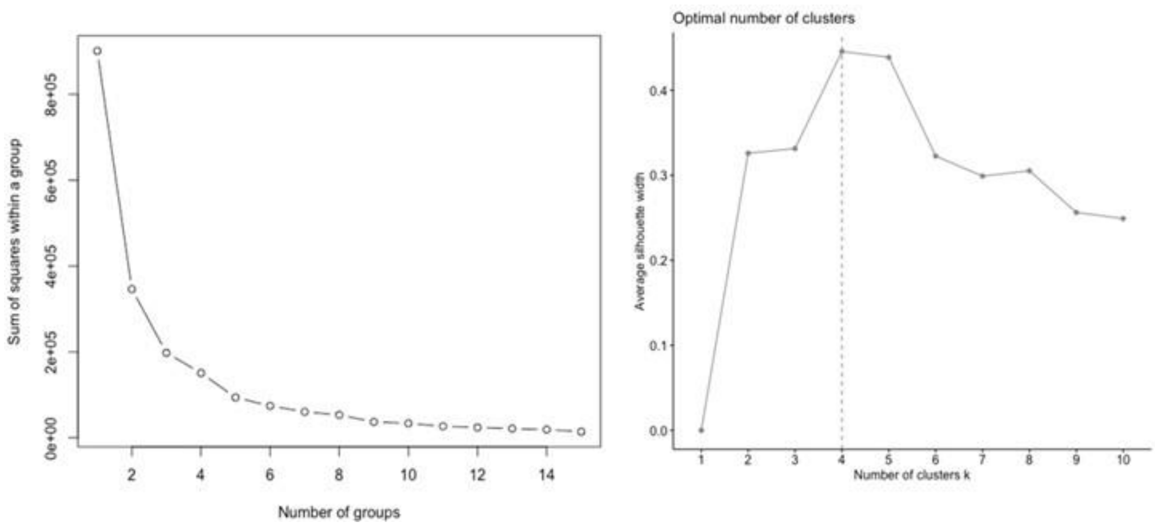
The second research question explores in-game behaviors and their potential relationship to learning and proficiency. Cluster analysis can be used to explore data or to test hypotheses about data structures (Huberty et al., 2005) and is useful for exploring potential groups within a set of participants who share similar characteristics (Warschauer et al., 2019). Further, cluster analysis is useful to explore potential patterns or groupings that are not easily seen from other analytic approaches. Therefore, the present study applied cluster analysis as an exploratory approach to analyzing student gameplay data, including participants' gameplay styles. We identified distinct in-game behaviors such as time spent on text, the number of battles engaged in, and the use of the glossing tool, and explored their relationship to vocabulary gains and affective factors related to L2 learning (i.e., reading anxiety and game enjoyment).

All variables were normalized before beginning the analysis to prevent any inflated differences from variation in scales. Formann (1984) suggests that the minimal sample size for cluster analyses is no less than 2^k, in which k refers to the number of variables included in the analysis. The preferred amount would be 5×2^k. For the present study, five variables ($2^5=32$) are the maximum that should be used, with fewer variables being ideal. Thus, four of the variables from Table 1 that were shown to be important in the CART analysis were added to the cluster analysis: the number of look-ups, average time spent on reading the text, the total number of battles engaged in, and the total number of texts read.

The next decision to make when doing a cluster analysis is the algorithm to use. We ran a K-means cluster analysis. The k-means algorithm is a type of centroid clustering that searches for center points of clusters in a way that minimizes the distance of each member of each cluster. K-means clustering also involves indicating how distance is measured (Attewell et al., 2015). When using a k-means approach, the number of clusters must first be identified. One approach is to try a different number of defined clusters and then use a silhouette plot to compare which number of clusters bests divides the data into meaningful groups. However, the elbow method can also be run to determine the optimal number of clusters. Figure 5 (left) illustrates the total sum of squares that is accounted for by each number of clusters. The point at which the line begins to level out is when adding more clusters no longer increases the sum of squares accounted for by the analysis. In this figure, either 3 or 4 clusters appear to be ideal. Figure 5 (right) provides a similar analysis but uses the average silhouette width, a measure of how similar an item is to its cluster. The higher the average similarity score for each item in a cluster, the better. After four clusters, the average silhouette score begins to drop suggesting that a 4-cluster solution is ideal.

Figure 5

Elbow (left) and Silhouette Width (right) Methods of Cluster Analysis



After the clusters were identified, a unique identifier was assigned to each cluster and descriptive statistics were run for each group to identify what makes them unique. Further, linear regressions were used to explore if these groups differed significantly in terms of change in learning or affect.

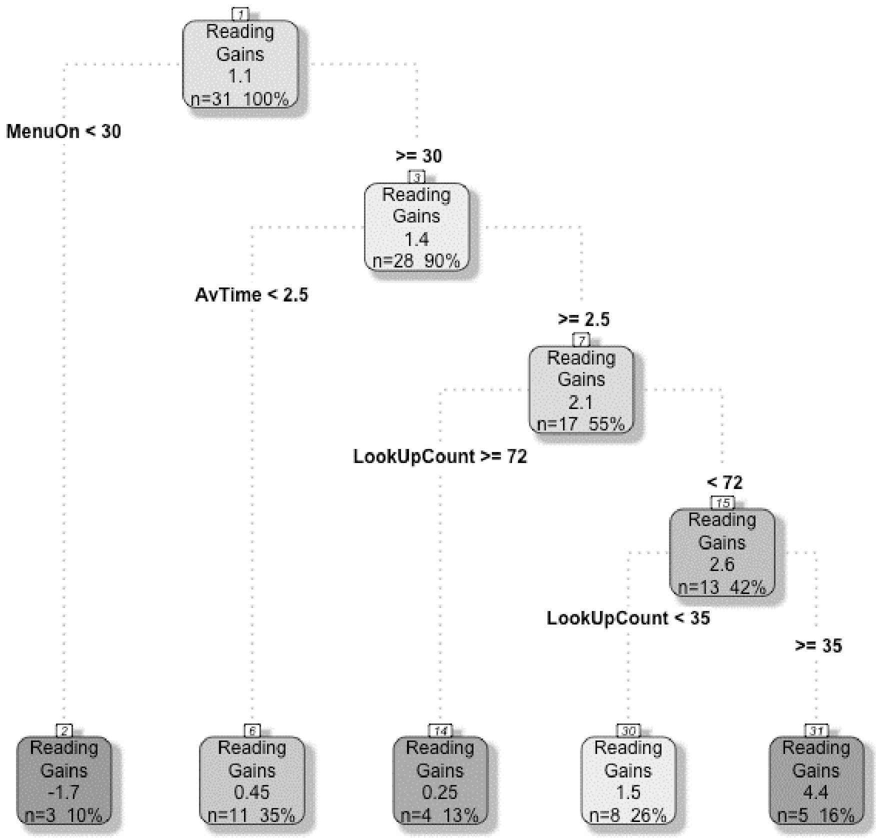
Results

Research Question 1

CART analysis was used to explore variables that were potentially important for predicting reading comprehension gains. Figure 6 shows the results of this analysis. The first node shows that for 31 students, there was an average reading gain of 1.1 points per student. The first variable that splits how students performed on the reading gains assessment is *Menu-On*. This variable indicates how many times a player opened their menu. When opening the in-game menu, players have access to information about items they have picked up, skills they have acquired, and baddies they have fought. Players who did this more than 30 times (n = 28) reported an average reading gain of 1.4, compared to those who accessed the menu less than 30 times (n = 3) and had an average 1.7 decrease in reading scores.

Figure 6

Reading Comprehension Gains CART Analysis

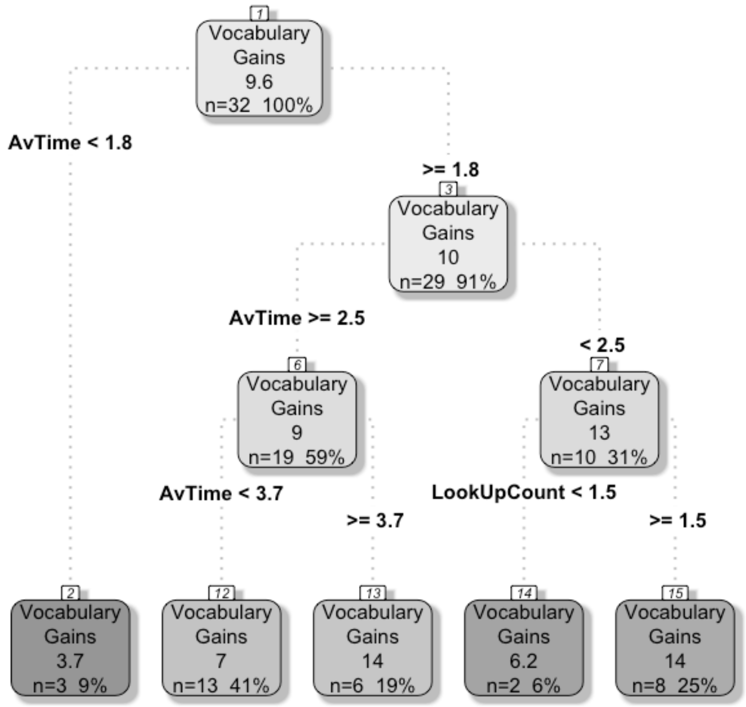


Although this split is first, it should be interpreted conservatively given that only three students were in the less-than-30 node. The second split is on *AvTime*. This variable indicates the average amount of time spent reading in-game texts. Students who spent more than 2.5 seconds on average reading a text (n = 17) reported a 2.1 gain in their reading scores on average, compared to those who spent less than 2.5 seconds (n = 11) and reported only a 0.41 increase in their reading scores on average. *LookUpCount*, the number of times a student used the glossing tool, was used to split the last two nodes. Students who spent more than 2.5 seconds on average reading a text and used the glossing tool less than 72 times but more than 35 times (n = 5) reported a gain of 4.4 points on average on the reading assessment. This suggests that students who used the menu options, spent a longer time reading the text, and used the glossing function reported higher reading gains. Note that to build this decision tree, one student whose high reading gains were strongly influencing the decision tree was removed.

The next CART analysis explored variables that were associated with vocabulary gains (see Figure 7). *AvTime*, or the average amount of time in seconds spent reading a text, was used to determine most of the splits, with the *LookUpCount* variable deciding one split. While vocabulary gains were reported for groups in all final nodes, those who spent longer average times reading the text reported higher vocabulary gains. Similarly, those who used the glossing tool (n = 8) reported higher vocabulary gains on average (M = 14).

Figure 7

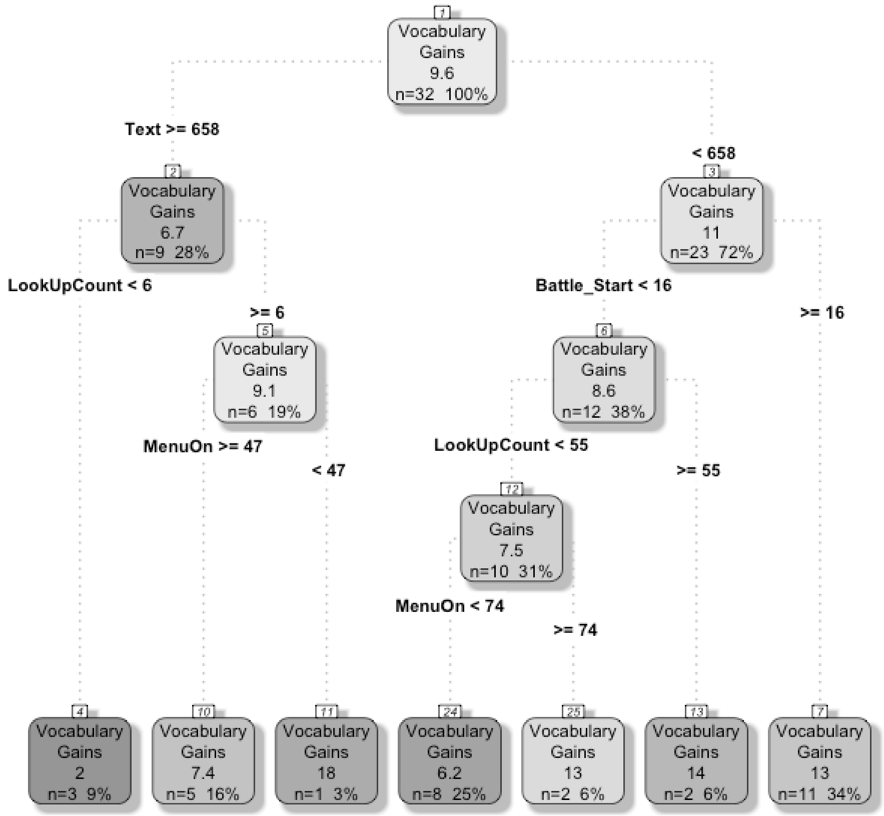
Vocabulary Gains CART Analysis



Because these vocabulary gains were primarily split depending on the variable *AvTime*, this variable was removed, and another CART analysis was created to explore other potential in-game indicators that were important for vocabulary learning. In this second CART analysis for vocabulary gains, with *AvTime* removed, the tree was first split based on *Texts* read (see Figure 8). Students exposed to less than 658 texts (n = 23) reported an average gain of 11 points on the vocabulary assessments. Students exposed to more than 658 texts (n = 9) reported a 6.7-point increase in the vocabulary assessments on average. Those who read fewer texts but engaged in more than 16 battles (n = 11) reported a 13-point gain on the vocabulary assessments on average. Finally, in two separate splits with *LookUpCount*, students who used the glossing tool had higher gains on the vocabulary assessment than students who did not on average. In these two CART analyses, no cases were removed.

Figure 8

Vocabulary Gains CART Analysis without Average Time



Research Question 2

A cluster analysis using four variables (*time, look-up, texts*, and *number of battles*) was run to answer RQ2. Table 2 shows the four clusters identified as a result of the analysis.

Table 2

Cluster Group Averages

Cluster group	*n*	Average Time (seconds)	Average number of Look-Ups	Average number of Texts	Average number of Battles
Battle Group	12	2.32 (-0.65)	21.91 (-0.43)	677.5 (0.82)	29 (0.98)
Not Interested	9	2.37 (-0.59)	18.44(-0.54)	465.55(-0.54)	11.89(-0.72)
Close Reading	10	3.94(1.40)	38.7 (0.04)	434.1 (-0.74)	14.2 (-0.49)
Vocabulary Check	5	2.71(-0.16)	105.20 (1.94)	624.00 (0.47)	17.80 (-0.13)
Total	36	2.83	37.27	549.47	19.1

Before describing the unique aspects of each group, it is important to note that boys and girls were distributed evenly across all four groups (see Table 3). Also of note is the distribution of clusters by class. This is important because the *Battle Group* consisted almost completely of students from Class A, which may suggest that battling was a focus of Class A and that playstyle was associated with the classroom environment and culture.

Table 3

Cluster Group by Gender and Class

Cluster	Boys	Girls	Total	Class A	Class B	Total
Battle Group	5	7	12	11	1	12
Not Interested	5	4	9	3	6	9
Close Reading	5	5	10	3	7	10
Vocabulary Check	2	3	5	3	2	5
Total	17	19	36	20	16	36

The *Battle Group* contained 12 students and had the largest average number of battles and texts. A one-way ANOVA and Tukey Post-Hoc test confirmed that the *Battle Group* engaged in significantly more battles than any other group ($p < .001$) and was exposed to significantly more texts than both the *Close Reading* and *Not Interested* groups ($p < .001$). This group also had the lowest average reading times and the lowest number of words looked up while reading. Table 4 shows that this group had the highest average scores on the positive experience construct, suggesting that they enjoyed the game. They also had higher reading gains than two of the other three groups.

The *Not Interested* cluster contained nine students. They had the second lowest average read time and number of look-ups, but the number of battles and texts read were also comparatively low, suggesting that this group did not engage with the text or the battles as much as other groups. Table 4 shows this group started with the highest anxiety scores on average, and the lowest vocabulary knowledge and reading comprehension scores, which suggests they did not engage with the game as much. However, this group reported a relatively high score on the positive experience construct and vocabulary knowledge gains.

The *Close Reading* cluster contained 10 students. This group spent the highest average amount of time on reading, almost two standard deviations more than the total average and significantly higher than any other group ($p < .001$). They also had almost twice as many vocabulary lookups on average than those in the *Battle* and *Not Interested* groups. Table 4 shows this group had the lowest positive experience score on average, but their average was still high (6 out of 8 on the Likert Scale). This group had the highest reading gains and vocabulary gains, on average despite starting with the highest reading comprehension scores and the second highest vocabulary knowledge scores on average.

Finally, the *Vocabulary Check* group had five students. This group looked up words almost two standard deviations more than the entire group average and significantly more than the other three groups ($p < .001$). This group had the second-highest average amount of time spent on text, and the number of texts read.

After identifying the clusters, we explored differences between the clusters on the four outcome variables: positive experience, change in anxiety, vocabulary gains, and reading gains. We conducted a separate linear regression for each outcome (See Table 5).

Table 4

Affect and Knowledge Score Averages by Cluster Groups

	Pre-vocabulary		Pre-reading comprehension		Pre-reading anxiety		Vocabulary gains		Reading gains		Positive experience	
Cluster Groups	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Battle Group	17.83	14.30	3.50	2.75	4.14	1.64	8.63	5.73	1.41	2.97	7.25	0.81
Close Reading	16.94	11.15	3.40	1.92	4.30	0.92	11.33	5.37	1.44	2.01	6.35	1.67
Not Interest-ed	15.31	21.36	2.63	3.20	4.67	0.83	9.08	5.42	0.43	2.48	6.64	1.41
Vocabu-lary Check	20.60	9.19	3.20	2.49	4.11	1.44	9.70	6.02	1.00	1.00	7.00	0.94

Table 5

Linear Regressions Comparing Outcomes by Cluster

	Reading gain	**Vocabulary gain**	**Anxiety gain**	**Positive experience**
Close Reading	0.028	2.708	0.234	-0.900
	(-2.064, 2.119)	(-2.147, 7.563)	(-0.540, 1.008)	(-1.963, 0.163)
Not Interested Group	-0.250	0.458	-0.488	-0.607
	(-2.622, 2.122)	(-5.047, 5.963)	(-1.366, 0.390)	(-1.788, 0.574)
Vocabulary Check	-0.417	1.075	-0.493	-0.250
	(-2.941, 2.108)	(-4.786, 6.936)	(-1.427, 0.442)	(-1.571, 1.071)
Constant	1.417	8.625^{*}	-0.107	7.250^{*}
	(0.047, 2.786)	(5.447, 11.803)	(-0.614, 0.400)	(6.533, 7.967)
Observations	32	32	32	34
R^2	0.005	0.043	0.111	0.091
Adjusted R^2	-0.101	-0.059	0.016	-0.0001
Residual Std. Error	2.420 (*df* = 28)	5.617 (*df* = 28)	0.896 (*df* = 28)	1.267 (*df* = 30)
F Statistic	0.051 (*df* = 3; 28)	0.423 (*df* = 3; 28)	1.163 (*df* = 3; 28)	0.999 (*df* = 3; 30)

Note. Battle Group is the Reference Group; $^*p < .001$

No significant differences between these groups were found. The implications of these findings will be further explored in the discussion and conclusion sections.

Discussion

Past studies on games and L2 learning have explored how high- and low-proficiency students engage with games (e.g., Rankin et al., 2006), how lexical complexity within games is associated with language production (e.g., Collentine, 2011), and how gameplay is related to affect (e.g., Hwang et al., 2017). However, one size does not fit all in education. More research is needed to explore how players engage with digital games in unique ways and how individual playstyles impact learning. In the present study, we explored gameplay patterns and types of actions that students engage in (e.g., reading, battling) and how they are related to learning and affective outcomes.

To answer the first research question, CART analyses were used to identify variables that were important to both vocabulary learning and reading comprehension gains. In both analyses, the average amount of time a student spent reading the text and the number of times the glossing tool was used to look up vocabulary were important variables in accounting for variance in learning gains. This may be explained by Laufer and Hultsijn's (2001) involvement load hypothesis, which argues that learning and vocabulary development are dependent on the amount of involvement of the learner. Hill and Laufer (2003) point out there is a difference between time-on-task and time-on-target forms—a distinction that could not be made in the present study because it is unclear whether the average time reading a text is associated with time spent processing vocabulary within the text. However, in a follow-up analysis, we found that average time on task was significantly correlated with the rate of glossing tool usage ($R = 0.34$, $p = .04$), suggesting time spent reading the task also involved looking up vocabulary words. These findings extend previous findings conducted in laboratory settings and illustrate how EDM approaches can broaden how research is conducted in language learning classrooms.

To answer the second research question, we used cluster analysis to group players into four groups that engaged with the game in unique ways. The first group identified was the *Battle Group*. This group was unique because they had significantly more battles than other groups and more text exposure. The *Battle Group* reported the highest average of game enjoyment and the lowest average of pre-reading anxiety scores compared to other groups. They also had the largest membership ($N = 12$).

The second group identified was the *Close Reading*. This group had the longest average reading times and the lowest average number of texts. Unlike the *Battle Group*, the *Close Reading* group may not have been as interested in fighting. It is important to note that no significant differences in learning were identified between these groups and both groups reported high levels of enjoyment with the game. Past research has indicated that games can promote learner autonomy (Ryan, et al., 2006) and that perception of autonomy is associated with enjoyment (Huang et al., 2019). This may explain why two groups of players who played in unique ways both reported high levels of enjoyment after playing the game. More importantly, this study illustrates how EDM approaches can be used to identify how player autonomy is manifested in an educational game.

The next group was the *Vocabulary Check* group. This group of five members had the highest average of vocabulary word look-ups and the highest pre-vocabulary scores. This is similar to earlier studies on glossing tools that found students with higher proficiency levels benefited more from glossing tools (Lee et al., 2019). This is also in line with the findings of Hsiao et al.'s (2017) study which examined patterns in how learners explored a virtual world. They noted that students with higher L2 proficiencies were more purposeful in the types of vocabulary they clicked on to learn.

The last group was deemed the *Not Interested* group. The three aforementioned groups all had an area in which they excelled or set them apart from the other groups (e.g., battling, reading, vocabulary). The *Not Interested* group had the second shortest reading time average, suggesting they did not engage in the reading. They also reported the second-lowest look-up average and text-exposure average. This again suggests they did not engage in reading the text as much as their other classmates. While they did engage in more battles than two other groups, they battled less than half as much as the *Battle Group*. The identification of this group may have the most significant implications for further research. Similar to approaches used in Foung (2019), EDM techniques could be used to identify these students during gameplay so they could receive additional instructor support.

Finally, the identification of these unique groups is an important finding as it highlights the need for more individualized analyses that capture individual experiences. This is the primary advantage of using EDM approaches to explore L2 learning environments.

Limitations

The present study has a few limitations; thus, the findings reported should be interpreted conservatively. First, although gameplay provided the several thousand data points ideal for data mining approaches, the total number of participants in this study is small. To further confirm some of the trends associated with learning in this study, research with a larger sample size is needed. Second, there was no relative qualitative data associated with the analysis in this study. While affect data via surveys were collected, video data or interview data that support the conjectures made about the groups identified in the cluster analysis would have been ideal. Future studies should employ mixed-method approaches to confirm and contextualize patterns found in the data.

Conclusion

In this study, we implemented EDM approaches to identify in-game behaviors that were associated with learning gains as well as four unique groups who played the game in different ways. Research using EDM approaches in L2 gaming contexts is rare and perhaps even non-existent at the elementary level. As noted by other researchers, applying such approaches to explore L2 learning in digital environments may be beneficial to educators (Godwin-Jones, 2017; Reinders, 2018). Further, collecting and analyzing data from digital games using EDM approaches may allow educators to visualize and capture learning as it happens in context. Providing teachers with visuals similar to the ones in this study can also be beneficial. Teachers may want to strategically organize group work to leverage the unique playstyles identified in the gaming experience and/or may be able to design follow-up activities that utilize the unique gaming experiences of students. These findings may also be beneficial to future game designers by providing insight into the role of time spent reading a text, use of the menu, text exposure, and glossing tools. Specifically, we found that reading gains on average were higher for those who spent more time reading the text and using the glossing tool. However, before such data and analytics are provided to educators, future research should explore teacher perceptions and beliefs about data and visualizations derived from in-class activities. Finally, this study took a reductionist approach to analyzing gameplay data. Future studies should explore other EDM perspectives (e.g., ontological, see Baker et al., 2021). For example, they could explore how language use in and around games in the classroom influences learning and gameplay. Such studies have the potential to provide more contextualization of language learning data from games.

References

Attewell, P., Monaghan, D., & Kwong, D. (2015). *Data mining for the social sciences: An introduction*. University of California Press. http://dx.doi.org/10.1525/9780520960596

Baker, R. (2010). Data mining for education. *International Encyclopedia of Education*, *7*, 112–118.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 61–75). Springer. http://dx.doi.org/10.1007/978-1-4614-3305-7_4

Baker, S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and innovation in digital learning* (pp. 83–98). http://dx.doi.org/10.15215/aupress/9781771991490.01

Baker, R. S., Gašević, D., & Karumbaiah, S. (2021). Four paradigms in learning analytics: Why paradigm convergence matters. *Computers and Education: Artificial Intelligence*, *2*, 100021. http://dx.doi.org/10.1016/j.caeai.2021.100021

Bytheway, J. (2014). In-game culture affects learners' use of vocabulary learning strategies in massively multiplayer online role-playing games. *International Journal of Computer-Assisted Language Learning and Teaching*, *4*(4), 1–13. http://dx.doi.org/10.4018/ijcallt.2014100101

Calvo-Ferrer, J. R. (2017). Educational games as stand-alone learning tools and their motivational effect on L2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology*, *48*(2), 264–278. https://doi.org/10.1111/bjet.12387

Collentine, K. (2011). Learner autonomy in a task-based 3D world and production. *Language Learning & Technology*, *15*(3), 50–67. http://dx.doi.org/10125/44262

Cornillie, F., Clarebout, G., & Desmet, P. (2012). Between learning and playing? Exploring learners' perceptions of corrective feedback in an immersive game for English pragmatics. *ReCALL*, *24*(3), 257–278. http://dx.doi.org/10.1017/S0958344012000146

De Grove, F., Van Looy, J., & Courtois, C. (2010). Towards a serious game experience model: Validation, extension and adaptation of the GEQ for use in an educational context. In L. Calvi, K. C. M. Nuijten, & H. Bouwknegt (Eds.), *Playability and player experience* (Vol. 10, pp. 47–61). Breda University of Applied Sciences.

DeFreitas, K., & Bernard, M. (2014). A framework for flexible educational data mining. In R. Stahlblock, G. Weiss, M. Abou-Nasr, & H. Arabnia (Eds.), *Proceedings of the 2014 International Conference on Data Mining* (pp. 176–180). CSREA Press. http://worldcomp-proceedings.com/proc/proc2014/dmin/DMIN_Papers.pdf

Dourda, K., Bratitsis, T., Griva, E., & Papadopoulou, P. (2014). Content and language integrated learning through an online game in primary school: A case study. *Electronic Journal of e-Learning*, *12*(3), 243–258.

Erhel, S., & Jamet, E. (2016). The effects of goal-oriented instructions in digital game-based learning. *Interactive Learning Environments*, *24*(8), 1744–1757. http://dx.doi.org/10.1080/10494820.2015.1041409

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3), 37–54.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130–160. https://doi.org/10.3102/0091732X20903304

Formann, A. K. (1984). *Die latent-class-analyse: Einführung in die theorie und anwendung*. Beltz.

Foung, D. (2019). Making good suggestions in analytics-based early alert systems: Shaping minds and changing behaviours. *Journal of Applied Research in Higher Education*, *12*(1), 109–123. https://doi.org/10.1108/JARHE-12-2018-0264

Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, *13*(3), 57–70.

Gibson, D., & Clarke-Midura, J. (2015). Some psychometric and design implications of game-based learning analytics. In P. Isaías, J. Spector, D. Ifenthaler, & D. Sampson (Eds.), *E-learning systems, environments and approaches* (pp. 247–261). Springer. http://dx.doi.org/10.1007/978-3-319-05825-2_17

Gobert, J. D., Sao Pedro, M. A., Baker, R. S., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 104–143.

Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. *Language Learning & Technology*, *21*(1), 4–15. https://dx.doi.org/10125/44592

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL*, *41*(2), 87–106. http://dx.doi.org/10.1515/iral.2003.007

Hsiao, I. Y., Lan, Y. J., & Kao C. L., & Li, P. (2017). Visualization analytics for second language vocabulary learning in virtual worlds. *Educational Technology & Society*, *20*(2), 161–175.

Huang, Y. C., Backman, S. J., Backman, K. F., McGuire, F. A., & Moore, D. W. (2019). An investigation of motivation and experience in virtual learning environments: A self-determination theory. *Education and Information Technologies*, *24*(1), 591–611. https://doi.org/10.1007/s10639-018-9784-5

Huberty, C. J., Jordan, E. M., & Brandt, W. C. (2005). Cluster analysis in higher education research. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 437–457). Springer. http://dx.doi.org/10.1007/1-4020-3279-X_8

Hwang, G. J., Hsu, T. C., Lai, C. L., & Hsueh, C. J. (2017). Interaction of problem-based gaming and learning anxiety in language students' English listening performance and progressive behavioral patterns. *Computers & Education*, *106*, 26–42. http://dx.doi.org/10.1016/j.compedu.2016.11.010

Jacob, J., Jha, K., Kotak, P., & Puthran, S. (2015). Educational data mining techniques and their applications. In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)* (pp. 1344–1348). IEEE. http://dx.doi.org/10.1109/ICGCIoT.2015.7380675

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. http://dx.doi.org/10.1093/applin/22.1.1

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, *21*(2), 32–51. https://dx.doi.org/10125/44610

Lee, H., Warschauer, M., & Lee, J. H. (2019). Advancing CALL research via data mining techniques: Unearthing hidden groups of learners in a corpus-based L2 vocabulary learning experiment. *ReCALL*, *31*(2), 135–149. http://dx.doi.org/10.1017/S0958344018000162

Martín-Monje, E., Castrillo, M. D., & Mañana-Rodríguez, J. (2018). Understanding online interaction in language MOOCs through learning analytics. *Computer Assisted Language Learning*, *31*(3), 251–272. http://dx.doi.org/10.1080/09588221.2017.1378237

Mendez, G., Buskirk, T. D., Lohr, S., & Haag, S. (2008). Factors associated with persistence in science and engineering majors: An exploratory study using classification trees and random forests. *Journal of Engineering Education*, *97*(1), 57–70. http://dx.doi.org/10.1002/j.2168-9830.2008.tb00954.x

Ming, Y., Ruan, Q., & Gao, G. (2013). A Mandarin edutainment system integrated virtual learning environments. *Speech Communication*, *55*(1), 71–83. http://dx.doi.org/10.1016/j.specom.2012.06.007

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society.*, *17*(4), 49–64.

Poole, F. J., & Clarke-Midura, J. (2020). A systematic review of digital games in second language learning studies. *International Journal of Game-Based Learning*, *10*(3), 1–15. http://dx.doi.org/10.4018/IJGBL.2020070101

Poole, F., Franco, J., & Clarke-Midura, J. (2018). Developing a personalized, educational gaming experience for young Chinese DLI learners: A design-based approach. In R. Zheng (Ed.), *Digital technologies and instructional design for personalized learning* (pp. 253–274). IGI Global http://dx.doi.org/10.4018/978-1-5225-3940-7.ch012

Rama, P. S., Black, R. W., van Es, E., & Warschauer, M. (2012). Affordances for second language learning in World of Warcraft. *ReCALL*, *24*(3), 322–338. http://dx.doi.org/10.1017/S0958344012000171

Rankin, Y., Gold, R., & Gooch, B. (2006). 3D role-playing games as language learning tools. In E. Croller & L. Szirmay-Kalos (Eds.), *Proceedings of EuroGraphics 2006* (pp. 33–38). The Eurographics Association.

Reinders, H. (2018). Learning analytics for language learning and teaching. *JALT CALL Journal*, *14*(1), 77–86. http://dx.doi.org/10.29140/jaltcall.v14n1.225

Reinders, H., & Lan, Y. J. (2021). Big data in language education and research. *Language Learning & Technology*, *25*(1), 1–3. http://hdl.handle.net/10125/44746

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618. http://dx.doi.org/10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. http://dx.doi.org/10.1002/widm.1075

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*, 344–360. http://dx.doi.org/10.1007/s11031-006-9051-8

Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, *83*(2), 202–218. http://dx.doi.org/10.1111/0026-7902.00016

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. http://dx.doi.org/10.1177/0002764213498851

Siemens, G., & Baker, R. S. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In S. Shum, D. Gasevic, & R. Ferguson (Eds.). *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). ACM. http://dx.doi.org/10.1145/2330601.2330661

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. https://dx.doi.org/10.11919%2Fj.issn.1002-0829.215044

Suh, S., Kim, S. W., & Kim, N. J. (2010). Effectiveness of MMORPG-based instruction in elementary English education in Korea. *Journal of Computer Assisted Learning*, *26*(5), 370–378. http://dx.doi.org/10.1111/j.1365-2729.2010.00353.x

Therneau, T. M., and Atkinson, E. J. (1997). *An introduction to recursive partitioning using the rpart routine* (Technical Report 61). Mayo Clinic. http://www.mayo.edu/hsr/techrpt/61.pdf

Therneau, T., & Atkinson, B. (2019*). Rpart: Recursive portioning and regression tree* (R package version 4.1-15). https://cran.r-project.org/package=rpart

Warschauer, M., Yim, S., Lee, H., & Zheng, B. (2019). Recent contributions of data mining to language learning research. *Annual Review of Applied Linguistics*, *39*, 93–112. http://dx.doi.org/10.1017/S0267190519000023

Wickham, H. (2017). *tidyverse* (R package version []). https://tidyverse.org .

Williams, G. (2011). *Data mining with rattle and r: The art of excavating data for knowledge discovery*. Springer Science & Business Media.

Wu, T. T., & Huang, Y. M. (2017). A mobile game-based English vocabulary practice system based on portfolio analysis. *Journal of Educational Technology & Society*, *20*(2), 265–277. https://www.jstor.org/stable/10.2307/90002180

Youngs, B. L. (2021). Item-level learning analytics: Ensuring quality in an online French course. *Language Learning & Technology*, *25*(1), 73–91. https://hdl.handle.net/10125/73426

Appendix A. Post Survey

First Name	
Last Name	

How well do you the following things in Chinese?

Check one for each skill.

	0 - Not at all	1	2	3	4	5	6	7	8 – I am the best in class
Speak									
Listen									
Write									
Read									

Answer the following questions by checking one box for each item.

1 = I completely disagree. ----- 8 = I completely Agree

Item	1	2	3	4	5	6	7	8

I enjoy reading Chinese								
I would be happy to learn to speak Chinese rather than having to learn to read as well.								
I felt irritated while playing the game.								
I felt bored playing the game.								
I get upset when I'm not sure whether I understand what I am reading in Chinese.								
The story of the game interested me.								
I put in a lot of effort while playing the game.								
I felt happy playing the game.								
I am worried about all the new symbols I have to learn in order to read Chinese								
The game was challenging								
It bothers me to encounter words I can't pronounce while reading Chinese								
I forgot everything around me while playing the game.								
The hardest part of learning Chinese is learning to read.								
I feel confident when I am reading in Chinese								
The game was boring.								
I felt frustrated while playing the game.								
I was totally absorbed in the game.								

Appendix B. Vocabulary Assessment

Chinese Character	I know the answer	I think I know the answer	I'm just guessing	Pinyin	English
蚂蚁					
蝙蝠					
欢迎					
龙					

鹰					
摘					
熊					
布					
找					
木头					
野猪					
狼					
宝剑					
棉花					
召唤					
蛇					
告诉					
砂石					
害怕					
折					
香蕉					
蝎子					
战斗					
兵马俑					
买菜					
石头					
无聊					

地震仪					
上海					
需要					
磁铁					
鸵鸟					
兔子					
北京					
船					
蜜蜂					
听说					
狐狸					
帮忙					
甲虫					
武器					
指南针					
知道					
喜欢					
猫头鹰					

Appendix C. Reading Assessment

1.虽然他喜欢红色的花但是他摘了蓝色的花。 ★他现在有什么颜色的花？ A – 蓝色 B – 黄色	6. 磁铁可以吸引铁的东西。 ★磁铁不可以吸引哪个东西？ A – 钥匙 B – 书

C – 红色	C – 白板
2. 兵马俑很古老，所以人们需要保护他们。 ★为什么兵马俑很古老？ A – 兵马俑太胖了 B – 是中国古代的人 C –因为生病了	7. 蜜蜂只能和小的动物战斗。 ★蜜蜂会和哪个动物战斗? A – 野猪 B – 老鼠 C – 狼
3. 今天的天气很不舒服。 ★今天的天气怎么样？ A – 不是很冷不是很热 B – 天气很不错! C –太热!	8. 我听说中国的龙可以召唤暴雨， 而西方的龙才可以喷火。 ★中国的龙…? A – 可以喷火 B – 可以让下雨 C – 是不好的
4. 这个指南针是指向西边。 ★指南针应该指哪边？ A –北边 B – 东边 C – 西边	9. 狼喜欢吃蔬菜。 ★狼会吃哪个？ A – 胡萝卜 B – 苹果 C – 米饭
5. 其实人们不应该害怕猫头鹰，应该害怕狼。 ★人们应该害怕哪个动物？ A – 狼 B – 猫头鹰 C – 野猪	10. 李老师需要你帮忙。 ★李老师可能要你做什么？ A – 现在回家 B – 和朋友说话 C – 给校长一张纸

About the Authors

Frederick Poole is an Assistant Professor in the Master of Arts in Foreign Language Teaching program at Michigan State University. His research explores the implementation of technology for enhancing second language literacy skills. He has a master's degree in second language teaching and a doctorate in instructional technology and learning sciences.

E-mail: poolefre@msu.edu

Jody Clarke-Midura is an Associate Professor of Instructional Technology and Learning Sciences at Utah State University. Her research investigates learning and assessment with digital media, toys, and games. She holds master's and doctorate degrees from the Harvard Graduate School of Education.

E-mail: jody.clarke@usu.edu