JOURNAL OF LEARNING ANALYTICS

Volume 9(3), 50–74. https://doi.org/10.18608/jla.2022.7637

Understanding Student Navigation Patterns in Game-Based Learning

Min Liu[1], Ying Cai[2], Songhee Han[3], Peixia Shao[4]

Abstract

Research on learning analytics (LA) has focused mostly at the university level. LA research in the K–12 setting is needed. This study aimed to understand 4,115 middle school students' learning paths based on their behavioural patterns and the relationship with performance levels when they used a digital learning game as their science curriculum. The findings showed significant positive relationships between various tool uses and performance measures and varied tool use patterns at different problem-solving phases by high- and low-performing students. The results indicated that students who used tools appropriately and wisely, given the phase they were at, were more likely to succeed. The findings offered an insightful glimpse of learners' navigation patterns in relation to their performance and provided much-needed empirical evidence to support using analytics for game-based learning in K–12 education. The findings also revealed that log data cannot explain all learners' actions. Implications for both research and practice are discussed.

Notes for Practice

- Learning analytics research has mostly been focused at the university level and research on the topic is needed in K–12 education.
- This study examined navigation patterns using a large dataset of three million lines of log data by over four thousand middle school students as they used a digital learning game. The results showed close relationships between students' navigation patterns and their performances and offered detailed analyses of tool use patterns at different problem-solving phases by high- and low-performing students. The findings provided empirical evidence to support using analytics for game-based learning (GBL) in K–12 education while also revealing the weakness of relying only on the log data.
- Our findings suggest that combining log data with interview and classroom observation data may provide a more complete picture of learning processes, and incorporating analytics into a teacher dashboard is a good way to help K–12 teachers implement GBL in their classrooms.

Keywords

Learning analytics, game-based learning, log data, middle school students, science learning

Submitted: 29/11/2021 — **Accepted:** 18/07/2022 — **Published:** 13/12/2022

Corresponding author [1]Email: mliu@austin.utexas.edu Address: The University of Texas at Austin, Learning Technologies Program, Department of Curriculum & Instruction, 1912 Speedway Stop, D5700, Austin, Texas, 78712-1293, USA. ORCID ID: https://orcid.org/0000-0003-0735-1148

[2]Email: yingcai@utexas.edu Address: The University of Texas at Austin, Learning Technologies Program, Department of Curriculum & Instruction, 1912 Speedway Stop, D5700, Austin, Texas, 78712-1293, USA. ORCID ID: https://orcid.org/0000-0002-3611-630X

[3]Email: song9@utexas.edu Address: The University of Texas at Austin, Learning Technologies Program, Department of Curriculum & Instruction, 1912 Speedway Stop, D5700, Austin, Texas, 78712-1293, USA. ORCID ID: https://orcid.org/0000-0002-8984-7560

[4]Email: pxshao@utexas.edu The University of Texas at Austin, Learning Technologies Program, Department of Curriculum & Instruction, 1912 Speedway Stop, D5700, Austin, Texas, 78712-1293, USA.

1. Introduction

With advances in technology, collecting data in digital games is much easier, and data usually come in large quantities. An important aspect of collecting data in digital games is that dynamic, real-time data can be captured unobtrusively. Such data can provide an insightful glimpse into how learners navigate through the game and how their navigation patterns may relate to their performance and other factors (Hwang et al., 2017; Law & Lárusdóttir, 2015; Li et al., 2021; Rosenheck et al., 2021; Sutcliffe & Hart, 2017). Yet, processing such data and making sense of them is a challenge, given their massive nature (Linek et al., 2010). In recent years, we have seen growing interest in learning analytics (LA; Siemens, 2013) through collecting, analyzing, measuring, and reporting big data from online systems such as digital games used for educational purposes.

Researchers are particularly interested in using such analytics to measure the effectiveness of the games and gain a deeper understanding of student learning processes to make informed decisions regarding game design, interventions, and using games for teaching and learning purposes (Gursoy et al., 2017; Ifenthaler et al., 2019; L'Heureux et al., 2017; Loh & Sheng, 2015). However, given that much LA research has focused on the university level, research applied to game-based learning (GBL) in the K–12 setting is needed to enrich the field.

This study aims to explore the log data captured in a digital educational game designed for middle school science called Alien Rescue (https://alienrescue.education.utexas.edu). Over 4,000 sixth graders used this digital game for three weeks as their science curriculum. Mouse clicks from these students are captured, representing over 3 million lines of raw log data. These log files reflect the real-time actions of students as they play the game, offering an invaluable picture of student learning processes. We aim to explore these student navigation patterns and understand how they are associated with their performance.

2. Literature Review

2.1. Learning Analytics in GBL Environments

GBL refers to a learning environment where knowledge and skills can be acquired through game content and activities, including problem solving (Gee, 2008; Krath et al., 2021; Qian & Clark, 2016). Compared to traditional forms of teaching in non-game conditions, GBL can support meaningful and authentic learning in more profound ways (Boyle et al., 2014). It is also more effective in supporting content learning and skill development (Clark et al., 2016). Additionally, GBL provides an open learning environment where students can freely act and explore different paths toward achieving learning goals (Krath et al., 2021; Rapp, 2017) and learn specific pieces of knowledge from the game (Wrzesien & Raya, 2010). However, the implementation of GBL in classroom teaching still faces many challenges, one of which is identifying student learning progression (Squire, 2008). In particular, Gomez et al. (2021) reported that classroom teachers cannot clearly understand how students are progressing in a game or whether students have conquered obstacles or are struggling with problem solving. Hence, the difficulty of measuring learning performance achieved through GBL is one of the main barriers to its broad adoption within formal education (Hauge et al., 2014). This challenge highlights the need to utilize a method to understand student in-game learning processes that can help teachers better integrate GBL into the school curriculum (Park et al., 2019).

According to Berland et al. (2013), learning analytics can help gain a clearer idea of student actions in GBL. Collecting student log file data, such as time spent in a game and navigation patterns, and aggregating those data further prescribed by algorithms (Ifenthaler, 2015, 2017; Kim & Ifenthaler, 2019) can reveal hidden information concerning the interactions between student in-game behaviours and learning. This information can inform teachers of student learning processes through gaming (Cheng et al., 2017; Freire et al., 2016). In previous studies, researchers adopted different learning analytics techniques. With the help of these techniques, the relationship between learner behaviour patterns and learning performances was detected, allowing for a richer examination of learning processes instead of just outcomes (e.g., Blikstein et al., 2014).

2.2. Behaviour Patterns and Student Learning Performance

2.2.1. Data Mining Techniques to Analyze Learner Behaviour Patterns in GBL

Learning analytics enables researchers and educators to trace student navigations within a game as evidence of learning performance (e.g., Kang, Liu, & Qu, 2017; Kang, Liu, & Liu, 2017; Loh & Sheng, 2014). Advanced data analytics methodologies provide analysis techniques for GBL environments to better understand learner behaviours (Kim & Ifenthaler, 2019; Loh et al., 2015). For example, Alonso-Fernández et al. (2019) conducted a systematic literature review of the techniques applied to learning analytics collected from a GBL environment. The techniques in their study can be classified into three main categories: (a) supervised models (e.g., linear and logistic regression, decision trees), (b) unsupervised models (e.g., correlation, clustering), and (c) visualization techniques (e.g., display of gameplay pathways). They concluded that linear models were the most used methods for supervised models, while correlation and cluster analysis were the most widely employed unsupervised methods. They also noted that most of the focus was on displaying learner performance in visualization technique–oriented studies. Fu et al. (2014) suggested that unsupervised algorithms were more appropriate when student strategies in the game were unknown, and researchers had to identify them from student activity patterns. Additionally, visualization techniques can help uncover meaningful information. Other research used game-based behaviour detection methods (e.g., Rowe et al., 2017) to first derive the target behaviours or strategies using visualization methods, and then set them as the outcome variable in a predictive model (Baker & Siemens, 2014). These findings provide guidelines for this current study for selecting appropriate techniques to understand learner behaviours within the game.

Among the different visualization techniques, a body of work based on network analysis investigates learning pathways (e.g., Durand et al., 2013; Zhu et al., 2018) and learner pathways in a game (e.g., Martin et al., 2013). Ruipérez-Valiente et al. (2019) created a graph of the most typical quest pathways followed by students using Gephi, constructed by creating edges between quest completions. They utilized the thickness of the edge to show the frequency at which students followed each path and the node's size to indicate the node's centrality within the network. Furthermore, Vista et al. (2016) adopted sequence-based approaches to demonstrate learner pathways, which were visualized as directed graphs. In addition, complex network approaches, such as Epistemic Network Analysis (ENA), are also adopted by researchers to analyze learner behaviour. For example, Karumbaiah et al. (2019) used ENA to generate the description of student quitting behaviours in a physics educational game. Gomez et al. (2020) utilized sequence mining techniques to develop the sequences of actions for different students who solved the problems in the game. However, Karumbaiah et al. (2019) and Gomez et al. (2020) used a small sample of participants. Thus, more research is needed on using visualization techniques to clarify the effectiveness of pathway recommendations in GBL on learning performance using a wider range of students.

2.2.2. Relationship Between Behaviour Patterns and Learning Performance

Several studies have been conducted to investigate student behaviours in GBL to examine the relationship associated with learning performance, either in-game learning performance or external outcome measures. Previous studies found that the frequency and duration of interacting with the learning tools and materials were closely related to student learning performance in a game (e.g., Cheng et al., 2015, 2017; Rosenheck et al., 2021). For instance, Cheng et al. (2015) developed an educational game to facilitate seventh-grade student understanding of biological evolution and explored the correlations between game performance, concept learning, and in-game behaviours. The correlation analysis results indicated a significant positive relationship between students' in-game behaviours and their performance. The higher frequency and longer duration of reviewing the relevant information, the higher the game score, representing a better learning performance. In addition, Cheng et al. (2017) discovered that the combination of using tools in GBL and the frequency of tool use significantly affected learning performance. They also examined the association between the game completion status and in-game tool usage in a massively multiplayer online game, which was developed to teach students math and biology concepts by completing a set sequence of tasks. Their results showed that student decisions in choosing the key in-game tools and the high frequency of using them led to a high task completion rate. This finding is aligned with the study by Rosenheck et al. (2021), which investigated the effectiveness of the same game but focused on the actions of students within the game and whether these actions contributed to the success of the task. The results showed that students' choices of in-game tools and high-frequency tool use contributed to success on the question.

In addition, numerous scholars have compared behaviour patterns between the high achieving and low achieving groups of students to discover the relationship between students' in-game behaviour patterns and their performance. Li et al. (2021) designed a digital GBL system to familiarize ninth-grade students with different kinds of food additives. By comparing the behaviour patterns of high- and low-achieving students, they found that the former would actively identify and build connections between learning objectives and learning scenarios to achieve the goals, while the latter would feel lost in the game and were poor at identifying such connections. In addition, high achievers could build an effective connection between the learning tool and the learning materials to complete the learning task. However, low achievers could not build such an effective connection though they repeatedly checked the tool, much like their high-achieving peers. Moreover, Sun et al. (2021) found that student knowledge-construction behaviour patterns varied by learning performance in a mobile game. Their findings in learning sequence analysis for the high- and low-performing groups indicated that overall, the behaviours were very similar across the groups, but a big difference between these two groups was that the lower-performing group lacked a path to modify their knowledge effectively after they had incorrect answers. Additionally, the high-performing group tended to go back and review the relevant learning materials and actively test their understanding once they discovered that their solutions were incorrect, while their counterparts rarely studied or restudied the relevant learning material or modified their existing knowledge once they were in the same situation.

2.3. Game-Based Learning Pathway Analysis

Since games have no limits or boundaries in playing, students in GBL environments tend to develop various learning paths — a series of actions taken to achieve the goal(s), including learning the content in the course of their actions (Feng & Yamada, 2021). Because of the diversity, many researchers posit that learning paths are one of the student attributes that can influence learning behaviour and performance (Shou et al., 2020; Williams & Rosenbaum, 2004).

According to different research goals, current GBL literature can be divided into two different types of learning path studies. On the one hand, some researchers focus on improving a GBL environment's adaptiveness by guiding a student to the personal ideal learning path based on their learning characteristics (e.g., Choi et al., 2020; Owen et al., 2019; Su, 2017). In this case, a recommendation system or detector aims to provide guidance during a student's in-game sessions based on an algorithm or model the researchers developed utilizing data mining techniques. For instance, Choi et al. (2020) developed an AI-based user interface allowing students to promptly decide whether certain recommendations met their needs or not by simply swiping and tapping. Similarly, Owen et al. (2019) built a detection model of unproductive persistence in a GBL environment to support varying student needs with just-in-time interventions for personalized learning experiences. On the other hand, others treat students' various learning paths as learning processes and focus more on analyzing the different paths that emerged from system-generated log data. This approach aims to better understand student learning behaviours than improve a GBL environment's adaptiveness (e.g., Kopeinik et al., 2012; Ruipérez-Valiente et al., 2019; Seng & Yatim, 2014). From this perspective, researchers tend to use visualization techniques (as discussed above) because student learning paths are a new type of data implicating learner characteristics rather than subsidiary data to improve a system's adaptiveness. This approach is applied to help teachers evaluate student learning performance in the GBL environment, which is a hard task due to the open-ended feature of a game environment (Feng & Yamada, 2021). For example, Ruipérez-Valiente et al. (2019) proposed three activity data metrics (i.e., quest progression linearity, quest event focus, and time per quest), showing the student path of completing tasks in the game. They presented multiple examples of how those metrics could both help teachers guide students to achieve their learning goals while also allowing students to reflect on their learning paths in the GBL context. This approach can unburden the teacher's role in supporting students in GBL and promote student self-regulation in their learning. However, previous studies rarely investigated student GBL paths in K–12 settings, especially not where GBL was used as a part of the school science curriculum, as in this study.

To this end, this study explores how middle school students navigate the GBL environment and what pathways they develop during their learning process. This study aims to understand students' learning pathways based on their behavioural patterns and the relationship with performance levels so that we can fill the knowledge gap by providing data-based evidence of the possible applications of LA techniques for GBL (Chaudy et al., 2014; Liu et al., 2017).

3. Research Questions

Our guiding research questions are these:

1. How do middle school students navigate a game-based learning environment?
2. What are the relationships between students' navigation patterns and their performance in the game?

By understanding the patterns and gaining insights into how student behaviours may affect their game performance, we hope to use evidence-based findings to inform the design of game-based learning and inform teachers as they incorporate digital games into their classrooms. The findings of this study contribute to much-needed research on GBL learning analytics in the K–12 setting.

4. Method

4.1. A Game-Based Learning Program

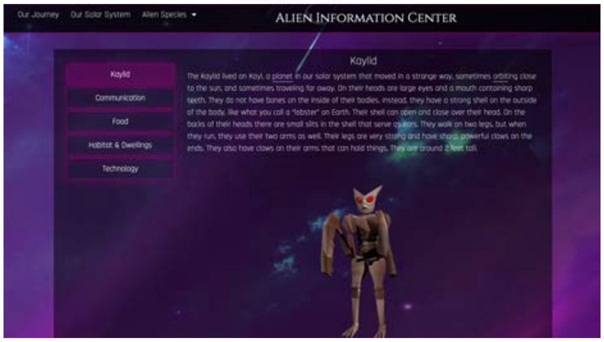
The data for this study came from the log files from 4,115 middle school students as they used an educational digital game called *Alien Rescue*. *Alien Rescue* is a space science game designed for 10- to 15-day classroom use of approximately one 45-minute session per day, delivered completely online. Students use the game to learn space science–related topics as schools cover that part of the curriculum. The goal of the game is to engage middle school students in inquiry-based learning as they solve the complex problem of finding suitable relocation sites within our solar system for six different alien species displaced from their home planets. *Alien Rescue* delivers a playful experience with an intentional problem-based narrative by combining game elements, play, and authenticity to engage and motivate students in learning science. The open scenario of *Alien Rescue* places students into the role of young scientists and asks them to join an urgent United Nations rescue mission to save the distressed aliens. This authentic scientific inquiry process is coupled with a playful fantasy experience (Lee & Liu, 2017) delivered through a 3D immersive, discovery, sensory-rich approach. The design aims to keep players motivated by incorporating such game attributes as challenge, control, fantasy, interaction, communication, mystery, role-play, representation, goals, sensory stimuli, and adaptation (Garris et al., 2002; Malone & Lepper, 1987; Wilson et al., 2009).

Alien Rescue is designed as an open environment. It encourages students to freely discover, explore, and acquire knowledge as they decide how to proceed in their problem-solving process. Students are challenged to set their own learning goals, work in collaboration, and control their own learning. To assist student problem solving, a set of nine multimedia-enriched tools is provided in the game (see Table 1 and Figure 1). Each built-in tool performs a specific function to assist students in their overall problem solving while the Communication Center functions as the home. While some tools (i.e., Solar System Database, Alien Information Center) are more critical during the initial problem-solving process, other tools (i.e., Probe Design Center and Mission Control Center) are essential later in the game. In addition, students are expected to use multiple tools in combination in different problem-solving phases. For example, when a student begins, they may want to spend more time studying the planets in the Solar System Database and alien characteristics in the Alien Information Center. Once a student has done this research and formed a hypothesis as to which planet(s) might serve as new alien homes, they may want to launch probes to get more information in Probe Design Center and Mission Control Center. Next, the student may want to access the Solar System Database and Alien Information Center again to see if their hypothesis is valid given the probe results. These tools are available through a two-layered interface (see Figures 1b and 1c). The first layer contains four centres accessed via arrows, while other tools are available in the toolbar (see Table 1). The two-layered tools are always available to students. How and when to use each tool during the entire inquiry process, however, are determined by the students, who are encouraged to be independent and self-directed in this open learning environment.

Table 1. Multimedia-Enriched Tools in *Alien Rescue*

Tool Name	Tool Functions
Available at tool bar	
Solar System Database	Provides information on selected planets and moons within our solar system. Data are intentionally *incomplete* to support the ill-structured nature of the problem-solving environment and foster the need for hypothesis testing.
Missions Database	Provides information on past NASA missions, including detailed descriptions of probes used on these missions.
Concept Database	Provides instructional modules on ten selected scientific concepts designed to facilitate conceptual understanding. This tool is for real-time learning. Students access this tool when they encounter an unfamiliar science concept needed for *Alien Rescue*.
Periodic Table	Provides a periodic table of the elements.
Spectra	Provides information to help students interpret spectra found in the Alien Information Center.
Notebook	Provides a notebook to store student notes about their research findings. This tool has a notes-comparison feature to facilitate comparing alien needs and planet requirements.
Available using 4 arrows	
Alien Information Center	Provides information via 3D imagery and text on the aliens' home planet, their journey, species characteristics, and habitat requirements.
Communication Center	Serves as a home for the program and the message tool. Students receive welcome messages when they log in, as well as when they submit their final answer to the problem using the message tool.
Probe Design Center	Provides information on scientific equipment used in both past and future probe missions. Students construct probes by deciding probe type, communication method, power source, and instruments.
Mission Control Center	Displays the data collected by the probes. Students analyze and interpret this data to develop a solution. Equipment malfunction may occur, and poor planning may lead to mission failure and budget waste.

ISSN 1929-7750 (online).

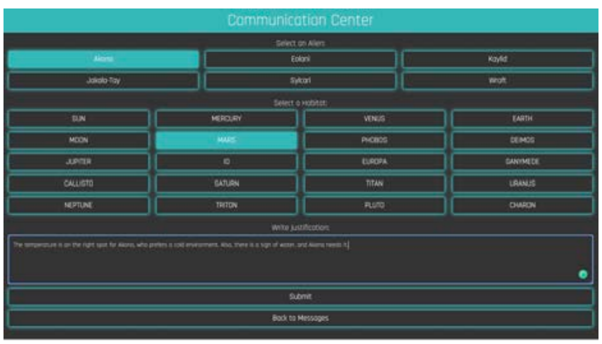


Alien Information Center

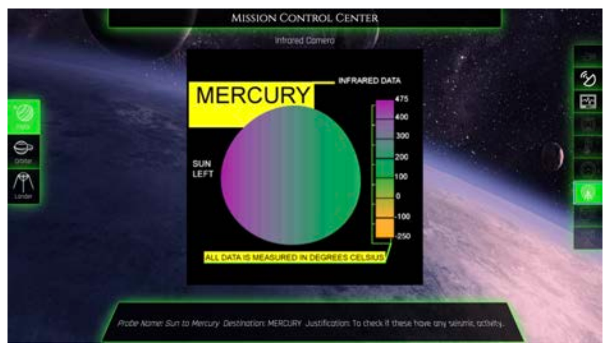
Solar System Database

Probe Design Center

Notebook



Communication Center Solution Form



Mission Control Center

Figure 1. Screenshot of *Alien Rescue*'s multimedia tools to support scientific inquiry.

4.2. Data Sources

The data sources for this study consisted of the log data of 4,115 sixth-grade students who used the game-based learning program from January to July 2019 as their curriculum unit on space science. The log files consist of over three million lines of raw log data reflecting each mouse click made as the student moved around the game and progressed from one part of the program to the next. These time- and date-stamped entries provide a picture of how each student used the nine tools to assist in their problem-solving process. The raw log data were cleaned and aggregated to calculate the frequency and duration of each of the nine tools. The usage of these nine tools is defined by the number of times a student accessed that tool (frequency) and the duration a student stays in that tool (duration in minutes). Average frequency and duration of tool use are used in the analyses.

How students solve the complex problem as presented by this GBL reflects their learning performance outcome. In discussing simulation-based assessment and advocating for using an evidence-centred assessment design, Mislevy (2011) stated, "An educational assessment is an evidentiary argument for reasoning what students say, do, or make in particular task situations as well as to generally claim different forms for different aspects of proficiency" (p. 10). This study has two types of evidence-centred assessments (performance measures). The primary one is the solution a student submits at the end of the gameplay. The student solution has two parts: 1) whether the submitted planet is an appropriate habitat for the alien to live on as measured by the solution success score (i.e., correct or incorrect), and 2) whether the student has provided a reasonable justification to show their rationale for the planet choice as measured by the solution justification score. The student justification score is graded on a 5-point rubric, with five being the highest (i.e., providing four or more specific reasons) and one being the lowest (i.e., providing no specific reasons). This justification score is as important as whether the planet choice is appropriate or inappropriate because of the emphasis on student rationale in the problem-solving process. It is possible for a student to still earn a few points if they choose an inappropriate planet but provide a good justification for the choice.

A secondary performance measure is whether a probe a student sent is successful or not. Sending probes is an important step during the later problem-solving process. A successful probe will return data that students need to test their hypothesis so that they can proceed with their problem solving. An unsuccessful probe will malfunction, and no data will be returned. While students can send as many probes as they want and submit as many solutions for any of the six alien species as they want, the probes sent and the solutions submitted may not be successful. Therefore, we calculated the probe success rate (percentage of probes submitted successfully) and the solution success rate (percentage of solutions submitted successfully) based on the number of probes sent and the number of solutions submitted as two additional performance measures.

4.3. Data Analysis

Student activity logs and performance scores were analyzed using both statistical and visualization techniques. To understand student navigation in the game, we first attempted to get an overall sense of how students used the tools in the game. In this step, a descriptive analysis of tool use was conducted regarding frequency and duration among all 4,115 students. Then, Spearman's correlation analysis was conducted to investigate any relationships between tool frequency, duration of use, solution success rate, and solution justification score, which were considered the primary performance measures.

For performance-related behavioural variations, the sample was divided into three groups — high, medium, and low performers — based first on the students' solution performance, including solution success rates and solution justification scores, and then on their probe success rates. The high-performing group consisted of students who achieved scores above the 75th percentile in each of the three performance measures; the low-performing group had scores below the 25th percentile. Students whose solution success rates and probe success rates were between the 25th and 75th percentiles formed the medium-performing group. Next, the differences between these groups were analyzed using Mann-Whitney U tests.

To further investigate student success with sending probes, Spearman's correlation and regression analyses were conducted to detect if certain variables could predict student performance and the relationship between probe performance and other measures. Finally, pathway analyses were conducted to detect how students navigated the game. Gephi (Bastian et al., 2009), a network analysis tool, was used to visualize the pathways as a graph network. A weighted degree metric was used for edge-ranking size, and a directed graph was adopted to show the movement directions from one tool to another. Pathway analyses were conducted among those students playing the game for at least five or more days (n=4,022, see Table 2 as an example). Students' sequential data of the first 2-days, and the last 2-days, were extracted first, and then their navigation paths for the mid-days were divided into three phases based on the proposed five-phase problem-solving process using the same game with the same age population (Liu & Bera, 2005). During the first phase, students were expected to explore the environment and define the problem. In phase 2, they were to conduct research, collect information, and refine the problem. In phase 3, students continue to do research and generate hypotheses. Then in phase 4, students test their hypotheses and conduct further research. Finally, in phase 5, students finalized their solutions and wrote up their rationales. We created these five-phase navigation paths for students in different performing groups. The high- and low-performing groups were then compared across these five phases to detect any differences in navigation patterns in different phases.

Table 2. Sample Extracted Sequential Data from One Student

ID	Date	Location
6782	2/8/19	[3, 6, 3, 6, 3, 6, 3, 11, 10, 1, 10, 1, 3, 1, 10, 3, 10, 3, 7, 1, 2, 3, 11]
6782	2/12/19	[2, 3, 5, 2, 4, 9, 4, 2, 3, 2]
6782	2/13/19	[3, 10, 6, 3, 10]
6782	2/14/19	[3, 10]
6782	2/18/19	[3, 11, 8, 11, 8, 10, 1, 3, 1, 7, 1, 7, 3, 8]
6782	2/19/19	[3, 8, 10, 8, 1, 3, 1, 7, 3, 11, 8, 11, 8, 10]
6782	2/21/19	[3, 8, 10, 8, 10, 8, 10, 8, 10, 8]
6782	2/26/19	[2, 3, 8, 10, 8, 10, 8, 10, 8]
6782	2/27/19	[3, 8, 10, 8, 3, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 10, 8, 3, 8, 1, 7]

6782	2/28/19	[3, 10, 1, 3, 1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 3, 10, 1, 3, 1, 7, 1, 7, 1, 7]
6782	3/1/19	[3, 8, 3, 8, 3, 10, 11, 4, 3, 8, 2, 3, 8, 3, 10, 8, 1, 3, 1, 7, 1, 7, 3, 8, 3, 8, 1, 3, 8, 3, 8, 1, 7, 3, 10, 8, 10, 8, 10, 7]
6782	3/2/19	[3, 8, 10, 1, 3, 1, 7, 3, 10, 8, 10, 1, 3, 1, 7, 3, 8]
6782	3/3/19	[3, 8, 10, 2, 8, 2, 8, 7, 8, 7, 2, 8, 10, 2, 8, 10, 8, 10, 8, 2]
6782	3/4/19	[8, 7, 8, 3, 10, 8, 1, 3, 8, 10, 8, 4, 5, 7, 5, 7, 10, 1, 3, 1, 7, 3, 8, 3, 8, 3, 8, 7, 3, 8, 3, 8, 3, 8, 7, 3, 7, 3, 8, 3, 8, 10, 8]
6782	3/5/19	[6]
6782	3/6/19	[3, 6]

Note: The locations are as follows: 1) Probe Design Center; 2) Alien Information Center; 3) Tool Bar; 4) Spectra; 5) Periodic Table; 6) Communication Center; 7) Mission Control; 8) Notebook; 9) Concept Database; 10) Solar Database; 11) Mission Database.

5. Results

To address the two research questions — "How do middle school students navigate a game-based learning environment?" and "What are the relationships between students' navigation patterns and their performance in the game?" — we went through four steps to explore the data and discover any patterns. As explained in the data analysis section, we first looked at the overall tool usage patterns descriptively. We then examined in depth the student tool usage associated with their performance. Finally, we investigated navigation paths. The findings are explained below.

5.1. Step 1: Descriptive Analysis of Tool Use: Frequency and Duration

Table 3 demonstrates the average frequency and duration of use for each of the nine tools in the game by all 4,115 students. Frequency refers to the number of times a student accessed a tool, and duration is how long the student stayed within a tool. The descriptive analysis of tool use indicated that at least one student used all nine tools more than once within the game.

Table 3. Average Frequency and Duration of Tool Use for 4,115 Students

Tool name	Frequency*(clicks)*	Duration (*min*)
Alien Information Center	91.75	42.07
Solar System Database	153.2	56.35
Probe Design Center	178.6	24.23
Mission Control	81.76	16.74
Mission Database	5.25	0.9
Concept Database	6.44	1.72
Periodical Table	7.08	2.64
Spectra	13	5.51
Notebook	105.2	35.58

When comparing usage, the most frequently visited tool was the Probe Design Center, with an average of 178.60 clicks (see Table 3), followed by the Solar System Database (153.20 clicks) and Notebook (105.20 clicks). The Mission Database was the least frequently visited tool, with an average of only 5.25 clicks. Regarding duration, students spent the longest in the Solar System Database (56.35 minutes), followed by the Alien Information Center (42.07 minutes) and Notebook (35.58 minutes); they spent only 0.90 minutes in the Mission Database. Previous studies (Liu et al., 2009, 2016, 2019) revealed that the Alien Information Center and the Solar System Database contain essential content information for understanding the aliens and our solar system. Therefore, the high frequency and long duration of use of these tools aligns with their importance in achieving the game's goals. Additionally, the Notebook's high frequency and long duration of use indicates that students used the Notebook to take notes and generate solutions (Liu et al., 2016). The Probe Design Center's high frequency and long duration of use suggests that students attempted to launch probes and gather data to achieve the goal.

Overall behaviour patterns suggest that students generally spent more time and used essential tools, such as the Solar System Database and the Alien Information Center, more frequently in *Alien Rescue*. As well, the important role that the Notebook plays in facilitating student learning emphasizes the need for an embedded notebook in the game, showing the importance of using data to inform game design. Given these descriptive findings, we began to examine student tool usage and performance in the next step.

5.2. Step 2: Tool Use Associated with Solution Performance

As described in the data sources section, two primary performance measures for this study are based on whether students select appropriate planets for the aliens (solution success rate) and provide reasonable justifications to show the rationale for their choice (solution justification score). The relationship between tool use and these two solution performance measurements is presented separately below. In addition, we investigated the relationships between the solution justification score, solution success rate, and probe success rate (a secondary performance measure).

5.2.1. Individual Tool Use and Solution Success Rate

Spearman's rank correlations were conducted to determine the relationship between tool use and solution success rate. As seen in Table 4, the frequency of use for all nine tools by all 4,115 students had a statistically significant positive correlation with the solution success rate. Statistically significant positive correlations were also found between the duration of tool use and the solution success rate. These positive correlations inferred that the higher the frequency and the longer the duration of tool use, the higher the solution success rate. To correct the inflated Type I errors in conducting multiple comparison analyses, *q*-value (Benjamini & Hochberg, 1995) using the "qvalue" R-package (Dabney & Storey, 2004) was adopted as an adjusted *p*-value for statistical significance.

Table 4. Spearman's Rank Correlation Coefficients between Variables of All Students

Usage	Tool name	Solution success rate
Tool frequency	Solar System Database	0.1453**
	Mission Database	0.1844**
	Concept Database	0.2408**
	Periodic Table	0.2202**
	Spectra	0.2426**
	Notebook	0.3422**
	Alien Information Center	0.3543**
	Probe Design Center	0.1272**
	Mission Control Center	0.1670**
Tool duration	Solar System Database	0.0488**
	Mission Database	0.1646**
	Concept Database	0.2396**
	Periodic Table	0.1743**
	Spectra	0.1818**
	Notebook	0.3329**
	Alien Information Center	0.2484**
	Probe Design Center	0.0736**
	Mission Control Center	0.1142**

**$q < .01$.

Although statistically significant correlations were obtained between frequency and duration of all tool use, the relationships were found to be moderate or weak. For example, the duration of use of the Notebook ($r = 0.3329$, $q < .01$), the Alien Information Center ($r = 0.2484$, $q < .01$), and the Concept Database ($r = 0.2396$, $q < .01$) were the most well correlated with the solution success rate, although these correlations were weak to moderate. To find out if any of these correlations were significantly different from the others, Hittner et al.'s (2003) *Z* transformation approach using the "cocor" R-package (Diedenhofen & Musch, 2015) was used to compare the differences between the correlations. The results showed that the correlations between the solution success rate and duration of using these three tools (i.e., Notebook, Alien Information Center, and Concept Database), and the frequency of using the Notebook and the Alien Information Center were statistically significantly different from the correlations between the solution success rate and other tools. For instance, the *Z*-score correlation between Notebook use duration and solution success rate ($r = 0.3329$) compared to Spectra use duration and solution success rate ($r = 0.1818$) is –13.18 ($p < .01$), indicating a significant difference. That is, some correlations are stronger than others. Additionally, Probe Design Center use duration ($r = 0.0736$, $q < .01$) and frequency ($r = 0.1272$, $q < .01$) were weakly correlated with the solution success rate.

Not all 4,115 students submitted a solution at the end of the game. In this study, the data reflected the actual classroom use of the GBL program. This real-world application came with some challenges. One was that during the period of use of this GBL, some students were absent for various reasons (e.g., sick, field trips). Some students, especially those with special needs, were slower in figuring out the complex problem than others. Some were distracted by the fun features of the game, as revealed by the findings of this study. These students were therefore not able to submit a solution at the end. The learning process of being engaged in the problem solving, however, is as important to the teachers and designers of this GBL implementation as the student solutions. After filtering out those students who did not submit a solution from the data, a Spearman's rank correlation analysis was carried out with the 1,183 students who did submit at least one solution (see Table 5). Significant positive relationships were found between the solution success rate and the frequency and duration of use of the Solar System Database, Concept Database, Periodic Table, Spectra, Notebook, Mission Database, and Alien Information Center. While some probe-related tool use (e.g., duration of use of the Probe Design Center and Mission Control Center) were found to be very weakly though significantly correlated with the solution success rate, no statistically significant correlations were found between the solution success rate and other probe-related tool use, such as the frequency of use of the Probe Design Center and Mission Control Center. These findings seem to suggest that an association between launching probes and getting a successful solution was less clear.

It is worth noting that the highest correlation with the solution success rate was found with Notebook use. The correlation analyses showed a significant moderate correlation between the frequency of Notebook use ($r = 0.313$, $q < .01$) and the duration of Notebook use with the success rate ($r = 0.3187$, $q < .01$). Comparisons also revealed that correlations between the use of the Notebook and the solution success rate were statistically significantly different from those between other tools with the solution success rate. For example, $z = -9.84$ ($p < .01$) in the differences between the correlation of Alien Information Center frequency of use and the solution success rate ($r = 0.2545$) and the correlation of Notebook frequency of use and the solution success rate ($r = 0.313$). These findings show the importance of using the Notebook within the game to support information processing, which helped students determine the appropriate solutions.

Table 5. Spearman's Rank Correlation Coefficients for Students Who Submitted at Least One Solution

Usage	Tool name	Solution success rate
Tool frequency	Solar System Database	0.2588**
	Mission Database	0.0457*
	Concept Database	0.1872**
	Periodic Table	0.2021**
	Spectra	0.2646**
	Notebook	0.313**
	Alien Information Center	0.2545**
	Probe Design Center	0.0004
	Mission Control Center	0.0208
Tool duration	Solar System Database	0.2147**
	Mission Database	0.0455*
	Concept Database	0.229**
	Periodic Table	0.1669**
	Spectra	0.2358**
	Notebook	0.3187**
	Alien Information Center	0.1498**
	Probe Design Center	0.0203*
	Mission Control Center	0.0477*

**$q < .01$, *$q < .05$

Tool Use by High, Medium, and Low Solution Rate Groups: Further analysis was carried out to determine whether tool use varied among groups with different success rates. A one-tailed Mann-Whitney U test was conducted to detect if the frequency and duration of tool use were different across the high ($n = 325$), medium ($n = 394$), and low ($n = 464$) solution rate groups (see Table 6). The 25th percentile of the solution success rate score was 0%, while the 75th percentile was 100%, so we used those as the cut-off scores for the low, medium and high groups. A one-tailed test was adopted because we aimed to detect if there was a difference between groups in a specific direction and hypothesized that the higher performance groups would perform better than the lower performance groups.

Table 6. Mann-Whitney U tests of Tool Usage for Students Who Submitted at Least One Solution

	Solution Success Rate								
	High vs. Low groups			High vs. Medium groups			Medium vs. Low groups		
	Median			Median			Median		
Variable	High (*n*=325)	Low (*n*=464)	Mann Whitney U	High (*n*=325)	Medium (*n*=394)	Mann Whitney U	Medium (*n*=394)	Low (*n*=464)	Mann Whitney U
Tool Frequency									
Solar System Database	176	91	49662**	176	170.5	63663	170.5	91	63660**
Mission Database	4	3	71805	4	4	68812	4	3	80232**
Concept Database	6	2	58056**	6	6	61810	6	2	72796**
Periodic Table	8	4	56866**	8	8	67764	8	4	64331**
Spectra	15	6	50050**	15	16	65854	16	6	59140**
Notebook	176	9	47126**	176	169	65861	169	9	55200**
Alien Information Center	126	86.5	50716**	126	156	75592	156	86.5	47968**
Probe Design Center	165	154	75214	165	210.5	73106	210.5	154	78644**
Mission Control Center	65	57.5	73154	65	87	74014	87	57.5	75764**
Tool Duration									
Solar System Database	52.92	25.03	53898**	52.92	46.41	57162**	46.41	25.03	74672**
Mission Database	0.22	0.22	72706	0.22	0.38	68572	0.38	0.22	81540**
Concept Database	1.1	0.13	53570**	1.1	0.63	57038**	0.63	0.13	71915**
Periodic Table	1.73	0.39	60046**	1.73	1.68	65579	1.68	0.39	70642**
Spectra	5.28	1.73	52493**	5.28	5.08	63848	5.08	1.73	65246**
Notebook	68.27	0.78	46803**	68.27	48.46	62743	48.46	0.78	57375**
Alien Information Center	48.4	41.15	60936**	48.4	52.23	68776	52.23	41.15	67883**
Probe Design Center	18.93	18.18	72222	18.93	20.32	65022	20.32	18.18	86470
Mission Control Center	10.92	8.93	70110*	10.92	13.5	67572	13.5	8.93	80423**

**q<.01, *q<.05, one-tailed.

Regarding tool frequency, the high solution rate group used significantly more tools than the low group, given their use of the Solar System Database (U = 49,662, q < .01), Concept Database (U = 58,056, q < .01), Periodic Table (U = 6,866, q < .01), Spectra (U = 50,050, q < .01), Notebook (U = 47,126, q < .01), and Alien Information Center (U = 50,716, q <.01). However, no significant differences were found regarding the Mission Database, Probe Design Center, and Mission Control Center. Moreover, the medium group had a significantly higher frequency of use for all nine tools than the low group. No significant differences were found between the high and medium groups' frequency of tool use.

Concerning the duration of tool use, there were significant differences between the high and low groups for the following tools: Solar System Database (U = 53,898, q < .01), Concept Database (U = 53,570, q < .01), Periodic Table (U = 60,046,

$q < .01$), Spectra ($U = 52{,}493$, $q < .01$), Notebook ($U = 46{,}803$, $q < .01$), Alien Information Center ($U = 60{,}936$, $q < .01$), and Mission Control Center ($U = 70{,}110$, $q < .01$). Similar findings were discovered between the medium and low groups. The medium group spent significantly more time in the Solar System Database ($U = 74672$, $q < .01$), Mission Database ($U = 81540$, $q < .01$), Concept Database ($U = 71915$, $q < .01$), Periodic Table ($U = 70642$, $q < .01$), Spectra ($U = 65246$, $q < .01$), Notebook ($U = 57375$, $q < .01$), Alien Information Center ($U = 67883$, $q < .01$), and Mission Control Center ($U = 80423$, $q < .01$). These results indicate that the high and medium groups spent significantly more time using tools than the low group. Few significant differences were found between the high and medium groups.

5.2.2. Individual Tool Use and Solution Justification Score

The relationship between the solution justification score and tool use was also examined. The solution justification score represents the rationale a student provides for their solution. As shown in Table 7, Spearman's rank correlation found that the solution justification score was positively significantly related to the frequency and duration of use for most of the nine tools, except the Probe Design Center and Mission Control Center. Interestingly, the solution justification score was significantly negatively correlated with frequency of use of the Probe Design Center ($r = -0.0560$, $p < .01$); there was no significant relationship with its duration of use. For the Mission Control Center, its frequency of use was also significantly negatively correlated with the solution justification score ($r = -0.0891$, $p < .01$); no significant relationship was found between duration of use and the solution justification score. These negative correlations indicate that the more students accessed the Probe Design Center and Mission Control Center, the lower their solution justification scores, further confirming the previous finding of a possible disconnect between using probe-related tools and solution (justification) scores.

Among all the variables, the strongest correlation with the solution justification score was with the Notebook; the coefficient with frequency was 0.4273 ($q < .01$) and 0.4253 ($q < .01$) with duration, both indicating a moderate to strong effect size. The Z transformation results comparing the coefficients of the frequency and duration of Notebook use with the correlation coefficients of other tools also demonstrated that the correlation with solution justification score was significantly different from other tools. For instance, the z coefficient of Notebook duration and solution justification score ($r = 0.4253$), and Concept Database duration and solution justification score ($r = 0.2625$) is 10.80 ($p < .01$), indicating its statistical significance. This finding further emphasized the importance of the Notebook in facilitating student problem solving. In addition, the solution justification score was moderately correlated with the solution success rate score ($r = 0.3596$, $q < .01$) and weakly correlated with the probe success rate score ($r = 0.1023$, $q < .01$), two other performance measures. This indicates that students with a stronger rationale for selecting a planet for an alien also got higher solution success rate and probe success rate scores.

Table 7. Spearman's Rank Correlation Coefficients for Solution Justification Score and Other Variables

	Variable	Solution Justification Score
Tool frequency	Solar System Database	0.2984**
	Mission Database	0.1504**
	Concept Database	0.2049**
	Periodic Table	0.2194**
	Spectra	0.2432**
	Notebook	0.4273**
	Alien Information Center	0.1322**
	Probe Design Center	-0.0560**
	Mission Control Center	-0.0891**
Tool duration	Solar System Database	0.2515**
	Mission Database	0.1638**
	Concept Database	0.2625**
	Periodic Table	0.1754**
	Spectra	0.2095**
	Notebook	0.4253**
	Alien Information Center	0.1367**
	Probe Design Center	0.0218
	Mission Control Center	-0.0150
Other Performance Indicators	Probe Success Rate	0.1023**
	Solution Success Rate	0.3596**

**$q < .01$

Solution Justification Score by High, Medium, and Low Solution Rate and Probe Success Rate Groups: To find out the solution justification scores for different solution rate groups and probe success rate groups, we conducted one-tailed Mann-Whitney U tests. Significant differences were found across the groups (see Table 8). The high solution success rate group had a significantly higher solution justification score than the low group ($U = 528214$, $q < .01$) and medium group ($U = 1222674$, $q < .01$). Additionally, the medium group received significantly higher solution justification scores than the low group ($U = 995744$, $q < .01$). Regarding the probe success rate groups, among the students who submitted at least one solution, the 25th percentile of the probe success rate score was 50%, while the 75th percentile was 100%, so we used these as the cut-off scores for the low, medium, and high groups. Statistically significant results were also found between the high probe success rate group and the low group ($U = 789221$, $q < .01$); and between the high and medium groups ($U = 962152$, $q < .01$) (see Table 9). No significant difference was found between the medium and low groups. That is, students in the high probe success rate group received a higher solution justification score than their peers in either the medium or low groups. Figures 2 and 3 demonstrated these differences in score distribution.

Table 8. Mann-Whitney U Tests on Solution Justification Scores of Students Who Submitted at Least One Solution

Solution Success Rate									
	High vs. Low groups			High vs. Medium groups			Medium vs. Low groups		
	High (n=325)	Low (n=464)		High (n=325)	Medium (n=394)		Medium (n=394)	Low (n=464)	
Variable	Median		Mann Whitney U	Median		Mann Whitney U	Median		Mann Whitney U
Solution Justification Score	3	1	528214**	3	3	1222674**	3	1	995744**

**q<.01, one-tailed

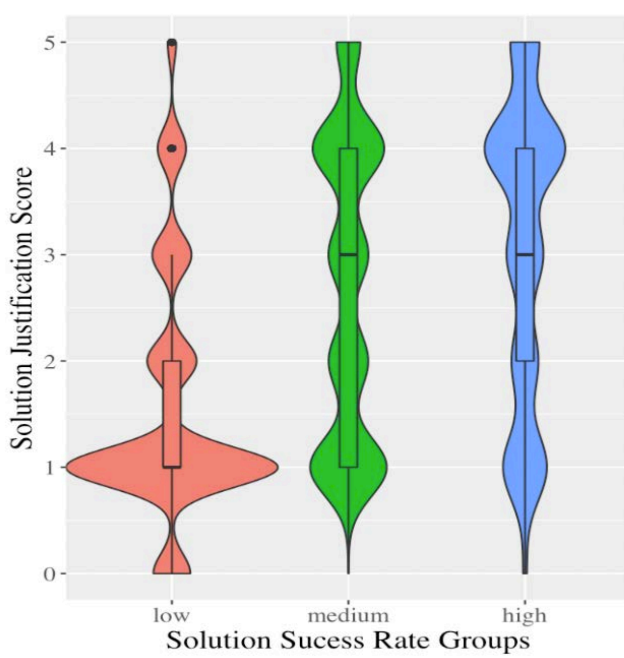


Figure 2. Distribution of the solution justification scores of students who submitted at least one solution in the low-, medium-, and high-solution success rate groups.

Table 9. Mann-Whitney U Tests on Probe Success Rates of Students Who Submitted at Least One Solution

	Probe Success Rate								
	High vs. Low groups			High vs. Medium groups			Medium vs. Low groups		
	High (*n*=414)	Low (*n*=409)		High (*n*=414)	Medium (*n*=360)		Medium (*n*=360)	Low (*n*=409)	
Variable	Median		Mann Whitney U	Median		Mann Whitney U	Median		Mann Whitney U
Solution Justification Score	3	2	789221**	3	2	962152**	2	2	817930

**q<.01, one-tailed

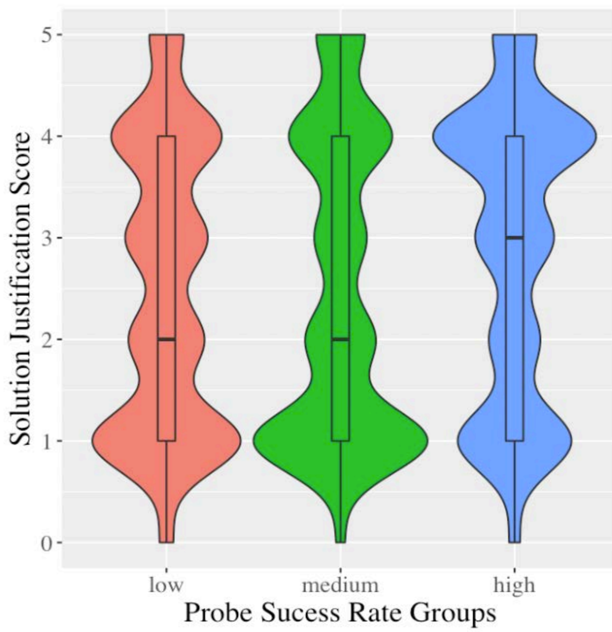


Figure 3. Distribution of the solution justification scores of students who submitted at least one solution in the low-, medium-, and high-probe success rate groups.

To summarize, the analyses from Step 2 showed significant positive relationships between the solution success rate and individual tool use by all students, indicating that the higher the frequency and the longer the duration of tool use, the higher the solution success rate would be. Examining individual tool use by students who submitted at least one solution showed significant positive associations between the solution success rate with most of the tools, except for some probe-related tools (i.e., the frequency of Probe Design Center and Mission Control Center use). The analyses of tool use by high, medium, and low solution rate groups further supported this finding by showing that the high and medium solution rate groups used significantly more tools and stayed longer in those tools than the low group. However, no significant differences were found between the high and low groups for two probe-related tools (i.e., Mission Database and Probe Design Center).

Analyzing the relationship between the solution justification score and tool usage showed significant positive associations for most of the tools, but very weak significant negative correlations or no correlations for two probe-related tools (i.e., Probe Design Center and Mission Control Center). Further analyses of the solution justification solution scores by high, medium, and low solution success rate and probe success rate groups showed that the high and medium solution success rate group had a significantly higher solution justification score than the low group. The high probe success rate group also had a significantly higher solution justification score than the low and medium probe success rate groups.

5.3. Step 3: Examining Probe Performance in Relation to Other Measures

The above results clearly demonstrate strong positive associations between tool usage for most tools and student performance — more and longer tool usage led to better performance, as shown in the final performance measure of both the solution success rate and the solution justification score. However, the association between usage of the three probe-related tools and student performance was less clear. Results show a very weak but negative correlation between the frequency of using two probe-related tools and the solution justification score and no correlation for the duration of use (Table 7). This finding is unexpected, since we would expect the opposite — that using more probe-related tools would lead to a higher solution justification score. Given this finding, we further examined probe tool usage in relation to other performance measures.

As with the solution success rate and solution justification scores, when students launch a probe, they are also asked to provide a rationale and what data it is to gather. The probe justification score is graded on a 5-point rubric, based on the scientific argumentation framework developed by Mao et al. (2018), with one being random/no justification, four providing a specific inquiry to send a probe, and five not only indicating a specific inquiry but also providing reasoning. That is, not all probe justifications are equal — some are better than others and some are meaningless. We wanted to find out if better probe justifications (i.e., those that achieved four or above) were associated with higher probe success and solution success rate scores. Spearman's rank correlations between the use of three probe-related tools and the percentages of reasoning and specific inquiry were conducted. In the analysis, the percent of reasoning referred to the percentage of reasoning probes out of the total number of probes the student launched, and the percent of specific inquiry was the percentage of specific inquiry probes out of total probes. The results showed significant positive relationships between these two higher probe justification scores and the use of probe-related tools (see Table 10). In addition, regression analyses conducted with the percent of reasoning and specific inquiry as the predictor and solution success rate ($Ys'=a+bX$) and probe success rate ($Y_P'=a+bX$) as the dependent variables showed a significant positive relationship between the percent of reasoning and specific inquiry and probe success rate ($\beta = 0.56$, $t = 42.88$, $p < .001$) and solution success rate ($\beta = 0.26$, $t = 8.01$, $p < .001$). The percent of reasoning and specific inquiry explained a significant amount of the variance in the probe success rate ($R^2 = 0.31$, $F(1,4113) = 1839$, $p < .001$) and solution success rate ($R^2 = 0.06$, $F(1,1181) = 64.2$, $p < .001$). Finally, comparing the probe justification scores by the high, medium, and low probe success rate groups found significant differences between the high and medium groups and the low group (see Table 11). In short, these results indicate that those students with higher probe justification scores also had high probe success and higher success rate scores and that they used the three probe-related tools more.

Table 10. Spearman's Rank Correlations Between Probe-Related Tools, Reasoning, and Specific Inquiry

Variable		Percent of Reasoning	Percent of Specific Inquiry
Tool Frequency	Mission Control	0.2049**	0.4299**
	Mission Database	0.1128**	0.0542**
	Probe Design Center	0.1810**	0.4649**
Tool Duration	Mission Control	0.1445**	0.5125**
	Mission Database	0.1101**	0.0554**
	Probe Design Center	0.1663**	0.5008**

**q<.01

Table 11. Mann-Whitney U tests on Probe Justification of Students Who Submitted at Least One Solution

	Probe Success Rate								
	High vs. Low groups			High vs. Medium groups			Medium vs. Low groups		
	Median			Median			Median		
Variable	High (n=414)	Low (n=409)	Mann Whitney U	High (n=414)	Medium (n=360)	Mann Whitney U	Medium (n=360)	Low (n=409)	Mann Whitney U
Percent of Reasoning	14.15%	7.54%	777063**	14.15%	9.18%	809256	9.18%	7.54%	982894**
Percent of Specific inquiry	45.46%	16.61%	964378**	45.46%	26.93%	1024256**	26.93%	16.61%	1053894**
Percent of vague inquiry	9.63%	7.66%	720411**	9.63%	13.34%	615957	13.34%	7.66%	1098715**
Percent of vague claim	12.54%	11.17%	710910*	12.54%	22.00%	1111864**	22.00%	11.17%	1131726**
Percent of random input	18.22%	10.84%	796076**	18.22%	28.54%	613190**	28.54%	10.84%	1189887**

**q < .01

5.4. Step 4: Navigation Pathway Analysis

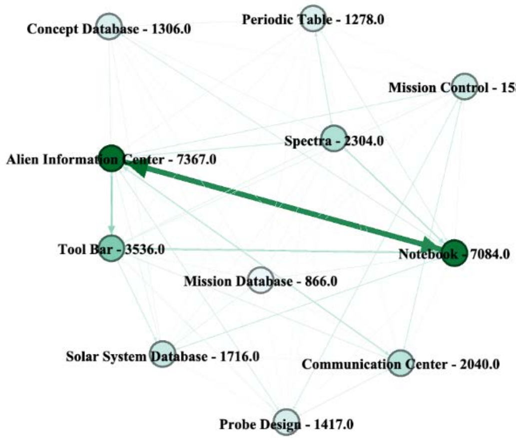
Previous research using the same game has shown that student problem-solving processes can be divided into different stages as students explore and identify the problem, perform needed research, generate and test the hypothesis, and then submit solutions (Liu & Bera, 2005). To better understand the relationships between students' navigation patterns and their performances, the entire path in the game was divided into five phases (see data analysis section), and a comparison of the navigation patterns was conducted between the high-performing and low-performing groups based on their solution success rate. Figure 4 presents the navigation paths by both groups in each of the five phases. It is necessary to point out that the thickness of a line was decided upon the weighted degree of edge ranking size within the group, which means that the different levels of thickness presented in the high-performing group (e.g., Figure 4a) and low-performing group (e.g., Figure 4b) should not be directly compared.

In the first phase, students in the high solution success rate group had a focused back-and-forth path between the Alien Information Center and the Notebook, as shown by the thickest line, indicating that they visited the Alien Information Center to get information and then used the Notebook for note-taking; on the other hand, students in the low solution success rate group explored more tools without focusing on the essential ones. Specifically, they accessed tools the Communication Center, Probe Design Center, and Mission Control Center quite often, indicating that they started to launch probes at the beginning stage. Mission Control and the Probe Design Center are the two most fun tools needed later in the problem-solving process (Liu et al., 2011; Liu & Bera, 2005). The focus for this phase should be on gathering information, as done by the high solution success rate group. Spending time using other tools more relevant to later phases, as done by the low solution success rate group, is not productive.

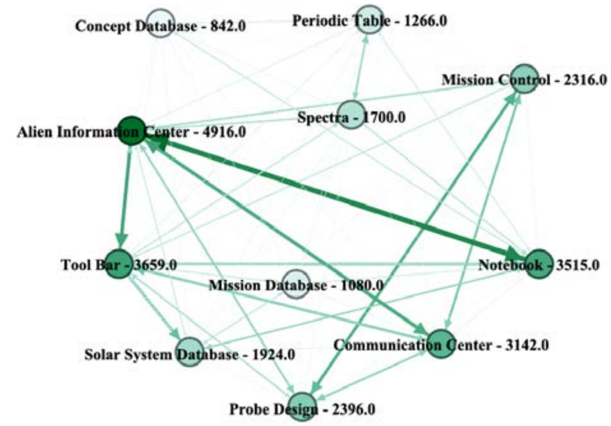
In the second phase, the thickest line in Figure 4c indicates that the high solution success rate group had a back-and-forth path between the Notebook and Alien Information Center, and the Notebook and Solar System Database. The Notebook was at the centre of the path, with the highest weighted degree (5010.0), which represents the central role of the Notebook in processing information in problem solving. The low solution success rate group also visited the Alien Information Center, Tool Bar, Solar System Database, and Notebook; however, their Notebook use had a weighted degree of 3348.0 — less than the Alien Information Center (4196) and Tool Bar (3851). This suggests that both groups were doing research using the two essential tools relevant for this phase — the Alien Information Center and Solar System Database; however, the high solution success rate group seemed to use the Notebook more frequently to process information from their research. While both groups started to launch probes, as shown by the path pattern between the Probe Design Center and Mission Control Center, this path was more frequently visited by the low solution success rate group during this phase. For the third and fourth phases, there were no apparent differences between the high and low groups. It is possible, since both phases three and four refer to generating hypotheses, testing hypotheses, and conducting further research, that all students performed similar actions, and that dividing usage into two separate phases may not be necessary.

In the last phase, the high solution success rate group focused more on the Probe Design Center and Mission Control Center, indicating that they were launching probes to test hypotheses, which is appropriate for this phase. Their paths showed the use of the Solar System Database and Notebook, suggesting that they most likely revisited them to check information after they received the data from launched probes, an indication of effective tool use appropriate for the phase (Liu & Bera, 2005). By contrast, the low group visited all tools and seemed to lack focus on what to pay attention to in this final phase of problem solving. This finding is significant since the participating schools used the game for a specific period of up to 10–12 days. If students were short of time, they would not be able to finish their problem-solving process and would not submit appropriate solutions.

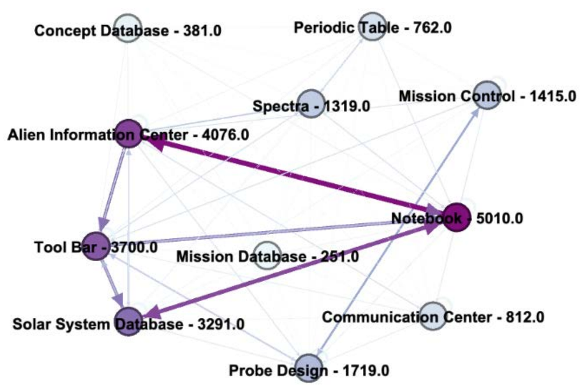
The results from the pathway analysis in Step 4 provided additional evidence that the high solution success rate group was using the tools more appropriately for each problem-solving phase. As a result, we conclude that they used the tools more productively than the low solution success group.



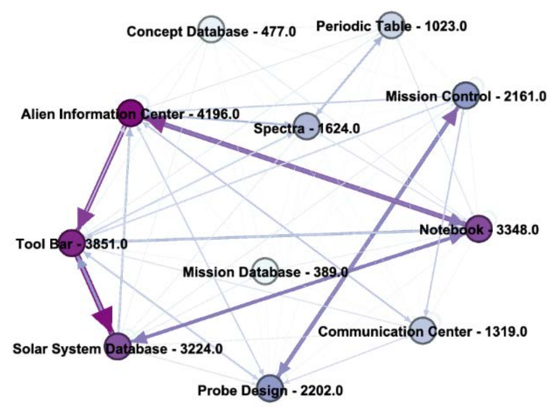
(a) 1st Phase High Solution Success Rate Group



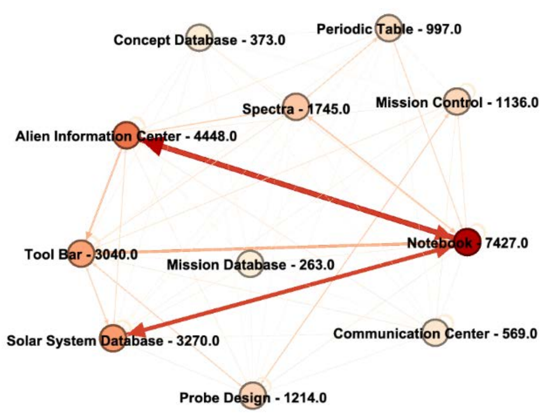
(b) 1st Phase Low Solution Success Rate Group



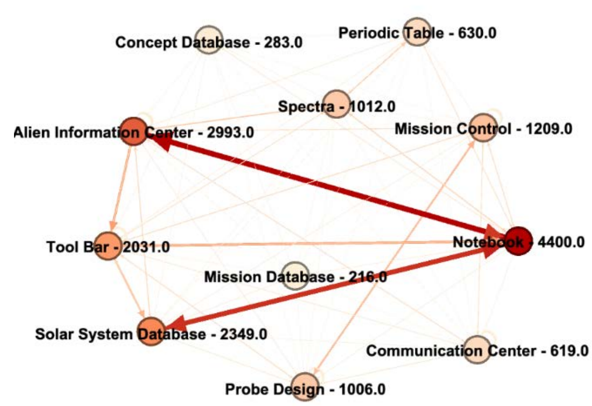
(c) 2nd Phase High Solution Success Rate Group



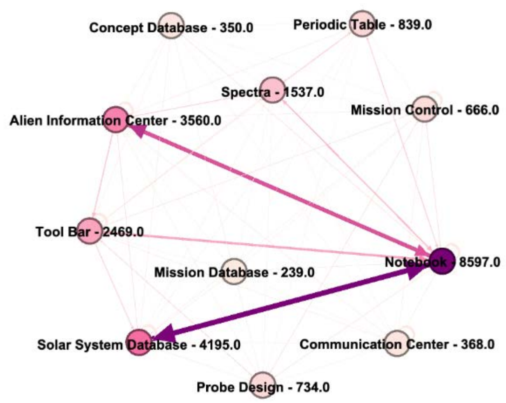
(d) 2nd Phase Low Solution Success Rate Group



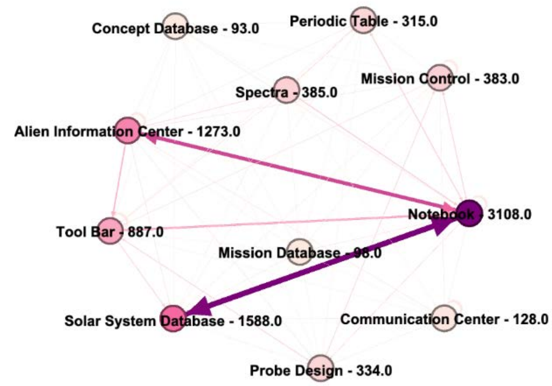
(e) 3rd Phase High Solution Success Rate Group



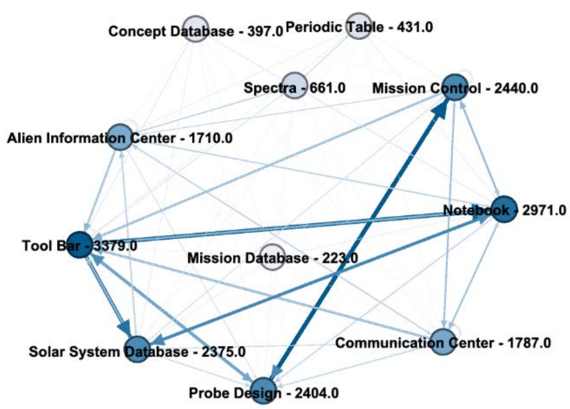
(f) 3rd Phase Low Solution Success Rate Group



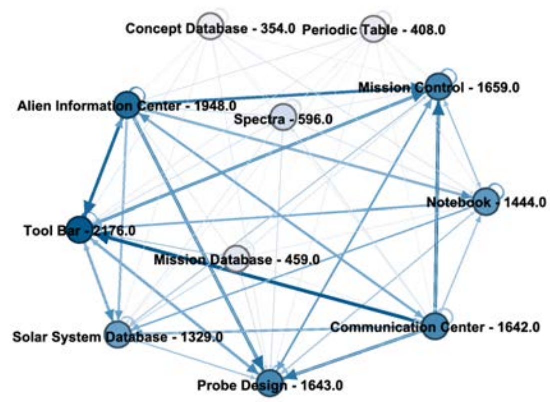
(g) 4th Phase High Solution Success Rate Group



(h) 4th Phase Low Solution Success Rate Group



(i) 5th Phase High Solution Success Rate Group



(j) 5th Phase Low Solution Success Rate Group

Figure 4. Comparison of tool usage in five phases by the high and low solution success rate groups.

6. Discussion

The goal of this study was to explore a large set of log data (3 million lines) by 4115 students captured in a digital educational game designed for middle school science and to understand student navigation patterns in relation to their performance. Our two research questions were these: 1) How do middle school students navigate a game-based learning environment? and 2) What are the relationships between students' navigation patterns and their performance in the game? Given our findings, we will discuss three main areas.

6.1. The Connection Between Individual Tool Use and Student Performance

The findings from the four-step data analysis demonstrated the navigation patterns of middle school students in a game-based learning environment. The descriptive analysis of tool use showed that, in general, students spent more time and used essential tools more often in the game than non-essential tools. Previous studies (Liu et al., 2009, 2016, 2019) identified that four tools (e.g., Alien Information Center, Solar System Database, Probe Design Center, Mission Control Center) are essential in the game in that they contain information students must use to solve the problems and find the solutions. Our current findings regarding the overall patterns of essential tool usage in the game aligned with the previous research and emphasized the critical role of those essential tools in facilitating the learning process.

In addition, results showed that the Notebook not only had high frequency, long duration usage, it also significantly positively correlated with solution success. The Notebook had the highest correlation with the solution success rate for students who submitted at least one solution (see Table 5). The Notebook was also the tool most strongly correlated with the solution justification score (see Table 7). In a previous study (Liu et al., 2019), the importance of using the embedded Notebook was proposed given its features, designed specifically for this game, in supporting learners' problem solving. Teachers were also encouraged to ask students to use this Notebook instead of a generic notetaking tool. Since the previous study, teachers were provided with professional training on the importance of the Notebook and how to utilize it effectively. The findings of this study provide empirical evidence to support the necessity of embedding the Notebook as an information processing tool in the game, and show that using an embedded Notebook in the system for note taking, especially when it provides features designed specifically for the game (Liu et al., 2019), is more convenient and relevant than other generic notebook tools outside the system.

Regarding the relationship between tool usage and solution success, the strongest correlations were found between the solution success rate and usage of the Alien Information Center, Concept Database, and Notebook. Additionally, the navigation patterns of the high-performing group showed strong connections between essential tools and the Notebook (e.g., Figure 4a, Figure 4c), also illustrating that more frequent use of essential tools and the Notebook helped students find solutions in the game. Previous literature (Hauge et al., 2014; Park et al., 2019) revealed that it was challenging for teachers to understand student in-game learning processes. The findings of this study reveal that students who use tools appropriately and wisely are more likely to succeed. The overall relationship between tool usage and learning performance could help classroom teachers better monitor student learning progress and provide intervention or scaffolding if necessary.

In terms of tool use among different solution success rate groups, the findings show that the high solution rate group used tools more frequently and spent more time using them than the low solution rate group. This finding indicated that the higher frequency and longer duration of using most tools led to a higher solution success rate. These findings are consistent with the results in the previous studies that the higher frequency and longer duration of using learning tools and materials in games contributes to successful game performance or learning achievement (Cheng et al., 2015, 2017; Rosenheck et al., 2021). In addition, the comparison of probe-related tool use among different probe success rate groups in the current study showed that students who had higher scores used more probe-related tools than lower performers. Students who chose and used key tools related to the tasks were more likely to have a higher success rate in completing the tasks. In *Alien Rescue*, while all tools are available through a two-layered interface (see Figure 1), the findings highlight the importance of strategic tool use when students are given the same amount of time to use the game. Previous literature has discovered that learners' ability to choose the proper tools and locations in the game was related to target performance (Cheng et al., 2017; Snow et al., 2015). In addition, high-performing learners could detect cues that are not as obvious to non-experts when solving problems in games (Loh & Sheng, 2015). Based on the findings of this study, it is possible for teachers to provide scaffolding, based upon the analytics, to those low-performing learners since they need more guidance on their path to a successful solution in games — a direction we are currently undertaking (Liu et al., 2022). It is also valuable to provide guidance or scaffolding features in games such as hints or prompts to assist student learning (Oren et al., 2021).

6.2. Learning Pathway Analysis in Five Phases

Overall, student strategies in selecting tools per phase aligned with the problem-solving process outlined in previous research using this game (Liu & Bera, 2005). However, student navigation patterns between high- and low-performing groups showed some differences in the first two phases. Specifically, both groups' paths were focused on the Alien Information Centre and Notebook tools in the first phase, suggesting that they gathered information about the aliens and took notes of what they thought important. In the second and third phases, both groups also showed increased traffic between the Solar System Database and the Notebook while continuing with the higher traffic between the Alien Information Center and the Notebook, implying that they were researching to solve the problem using more tools to look for suitable habitats for the aliens based on their notes saved in the previous phase. Nevertheless, the high solution success rate group focused more frequently on exploring essential tools, such as the Alien Information Center and Notebook, to get information and note-taking in phase one, and then started to send probes in phase two. On the other hand, the low solution success rate group focused more on tools not relevant for the first phase but more appropriate for later phases, such as probe-related tools, by sending probes earlier than the high solution success rate group.

In the fourth phase, the weighted traffic between the Solar System Database and the Notebook for both groups indicated that they were testing their hypotheses of suitable habitats in the solar system. In the fifth phase, the two groups showed a distinct difference in their dominant path-taking strategies. The high solution success group mainly used the Probe Design Center and Mission Control Center, suggesting that they were finalizing their solutions, while the low solution success group's focus was disoriented in using multiple tools, implying that they lacked focus and were not near finalizing their solutions.

Our findings suggest how selecting proper tools per phase contributes to student achievement in this problem-based learning, which is aligned with what Li et al. (2021) found in their study, as well as in previous research using the same game (Liu & Bera, 2005). Furthermore, the analysis of visualized learning paths confirms that students generally show a similar learning path in terms of sequencing the tools regardless of their learning performance levels, as Sun et al. (2021) reported, except for the duration of essential tool use, especially in the initial two phases of the game. Thus, the findings of this study enrich the GBL literature by providing more detailed descriptions of how students take their learning paths differently and the impact of path-taking strategy on performance.

6.3. Disconnect Between Probe Design Success and Solution

In the above sections, we have discussed student navigation patterns and their relationship to student performance. The two performance measures in this game are described in the "Method" section above. The primary one is the solution a student submits at the end of the gameplay; the secondary one is whether a probe a student sent is successful or not. Examining the log data in detail, the data analysis showed several interesting findings in using probe-related tools (i.e., Probe Design Center, Mission Database, Mission Control Center). First, there was no significant difference between the high and low solution success rate groups in their frequency and duration of tool use in terms of these probe-related tools (Table 6). Because these tools are essential (especially in phase 4) to solve the problem since they contain key information needed for hypothesis testing, it is not surprising that no significance was found between the two groups. Second, the results revealed an unexpected finding — the use of some probe-related tools was not correlated with the solution success rate (Table 5) or the solution justification score (Table 7). We would have thought that higher use of these tools would lead to a more successful solution, which is not the case, as revealed by the log data. Moreover, a negative correlation was found between the frequency of Probe Design Center tool use and the solution justification score. These findings seem to indicate that using the probe-related tools, while aligned with probe rate success performance, did not contribute to the solution success rate. Why?

One possible explanation is that probe-related tools (especially the Probe Design Center and Mission Control Center) have been found to be the most fun to use in previous research (Kang, Liu, & Qu, 2017; Kang, Liu, & Liu, 2017; Liu et al., 2016) because those tools allow students to work as scientists to design probes and thus experience the authenticity of learning that middle school students do not have an opportunity to experience otherwise. Therefore, this finding may explain why both high-performing and low-performing groups use these probe-related fun tools in high frequency and long duration, with no significant group differences. The probe justification scores — an indicator of whether a probe should be sent or not — have five levels of specificity (see Table 11) from high (more desired) to low (not desired). A detailed examination of these five levels shows that students with a high probe success rate sent more reasoning and specific inquiry justifications than the low-performing group. At the same time, they also sent more vague justifications and random input in probe design. Providing high levels of reasoning and specific inquiry justifications by the high-performing group showed that students understood the problems well and sent probes for testing their specific hypotheses. These two higher-level types of probe justifications could predict the success of probes and then contribute to getting the solutions. But why did the high-performing learners also send vague and random input justifications (lower-level types of probe justifications), showing that student justifications lacked meaningful purpose? It is quite possible that these students sent probes just for fun, especially during the first phase, rather than sending probes for testing hypotheses. This may explain why the frequency of using probe-related tools was not correlated with the solution success rate shown in Table 5. That is, the probe-related tools were not only used by students for testing hypotheses but also for entertainment purposes. The high frequency use of those probe tools did not provide an accurate picture of whether students were staying on task within the game. As a result, a disconnect between probe success and solution success was revealed.

Another possible explanation is related to the ambiguity of log data. Although they provide objective information and large quantities of behavioural characteristics about each user, sometimes it is difficult to precisely determine why a user takes a certain action in the problem-solving process (Horn et al., 2016; Linek et al., 2010). In support of previous learning analytics literature indicating this difficulty, we believe that more data from different sources other than log data is necessary to understanding student learning processes based on less guessing and more evidence (Gauthier et al., 2015).

6.4. Limitations

It is necessary to point out the limitations of this study. In our data analysis, we only investigated if the percentage of reasoning and specific inquiry probes students sent could predict their probes and solution success rates; other variables were not included in the model. The low R-square values of the two regression models suggest that other variables could be included, which can be done in our future studies. The analyses of this study were based on a large set of log data only. While log data provide an objective picture of what happened — especially in offering a glimpse into real-time actions during the problem-solving process — they may not reveal a complete picture of why certain events happen the way they do. Therefore, we advocate using multiple data sources (see Implications and Conclusion below). In addition, the research context for this study occurred in middle school classrooms in the U.S. Many constraints are associated with this real-world application, such as the attrition rate. While this could be partly explained by student absences, slow progress by students with special needs, and distraction by the fun features of the game, there could be other reasons, such as lack of engagement, losing concentration, or simply finding the game too hard and needing more scaffolding from teachers (Liu et al., 2011). Our hope is that by reporting how

LA is used in real-world settings, more effective games can be designed by incorporating LA to support teaching and learning in K–12 to achieve the ultimate educational goal.

7. Implications and Conclusion

Our findings showed significant positive relationships between various tool use and performance measures; and varied tool use patterns at different problem-solving phases by high- and low-performing students. Importantly, these findings revealed that students who used tools appropriately and wisely were more likely to succeed, as shown by the high-performance group. In addition, detailed multiple analyses showed an unexpected disconnect between the two performance measures.

The findings of this study have implications for both research and the design of digital learning environments. The learning performance measured showed two types of evidence-centred assessments based on the log data, which is the focus of this study. In future research, we plan to add external assessments, such as the science knowledge test given by the schools before and after students have completed this curriculum unit. Such an external measure could help establish a relationship between in-game performance and external learning measures. While analytics using log data provide dynamic, real-time, detailed information about learner actions unobtrusively and objectively, sometimes such data cannot provide a full picture of what is happening and why. To this end, we suggest that future studies use other data (e.g., surveys, interviews, classroom observations) in combination with log data to grasp a fuller picture of student learning in a classroom setting. For example, observing and interviewing students can help us understand why some students do not submit solutions to this game.

An important benefit of using analytics is to gain a better understanding of student learning processes to inform design and create interventions (Ifenthaler et al., 2019; Loh & Sheng, 2015). Yet it is challenging for teachers to know how their students use the game (Hauge et al., 2014; Park et al., 2019). To address this challenge and support K–12 teachers in implementing GBL more effectively, we have begun to incorporate analytics into a newly created teacher dashboard (Liu et al., 2022) to allow teachers not only to view their students' actions in real-time but also to provide their scaffolding based on analytics (van Leeuwen et al., 2021). In short, learning analytics offers new opportunities for GBL researchers and designers to gain more insights into learners' problem-solving processes. Research on GBL analytics will help guide more effective game design to support teaching and learning in the K–12 setting.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

References

Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education, 141*. https://doi.org/10.1016/j.compedu.2019.103612

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*, 2nd ed. (pp. 253–274). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media* (ICWSM '09) 17–20 May 2009, San Jose, CA, USA (pp. 361–362). AAAI Press. https://ojs.aaai.org/index.php/ICWSM/article/view/13937/13786

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *Journal of the Learning Sciences, 22*(4), 564–599. https://doi.org/10.1080/10508406.2013.836655

Blikstein, P., Worsley, M., Piech, C., Sahami, M., Cooper, S., & Koller, D. (2014). Programming pluralism: Using learning analytics to detect patterns in the learning of computer programming. *Journal of the Learning Sciences*, *23*(4), 561–599. https://doi.org/10.1080/10508406.2014.954750

Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A., & Van Rosmalen, P. (2014). A narrative literature review of games, animations and simulations to teach research methods and statistics. *Computers & Education*, *74*, 1–14. https://doi.org/10.1016/j.compedu.2014.01.004

Chaudy, Y., Connolly, T., & Hainey, T. (2014). Learning analytics in serious games: A review of the literature. *Proceedings of European Conference in the Applications of Enabling Technologies* (ECAET 2014), 20–21 November 2014, Glasgow, UK. https://research-portal.uws.ac.uk/en/publications/learning-analytics-in-serious-games-a-review-of-the-literature

Cheng, M. T., Rosenheck, L., Lin, C. Y., & Klopfer, E. (2017). Analyzing gameplay data to inform feedback loops in the Radix Endeavor. *Computers & Education, 111*, 60–73. https://doi.org/10.1016/j.compedu.2017.03.015

Cheng, M. T., Lin, Y. W., & She, H. C. (2015). Learning through playing Virtual Age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18–29. https://doi.org/10.1016/j.compedu.2015.03.007

Choi, Y., Na, Y., Yoon, Y., Shin, J., Bae, C., Suh, H., Kim, B., & Heo, J. (2020). Choose your own question: Encouraging self-personalization in learning path construction. https://doi.org/10.48550/arXiv.2005.03818

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Dabney, A., & Storey, J. D. (2004). QVALUE: The Manual Version 1.0. University of Washington, Department of Biostatistics. http://www.bioconductor.org/packages/2.13/bioc/vignettes/qvalue/inst/doc/manual.pdf

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PloS One*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Durand, G., Belacel, N., & LaPlante, F. (2013). Graph theory based model for learning path recommendation. *Information Sciences, 251*, 10–21. https://doi.org/10.1016/j.ins.2013.04.017

Feng, X., & Yamada, M. (2021). An analytical approach for detecting and explaining the learning path patterns of an informal learning game. *Educational Technology & Society, 24*(1), 176–190. https://www.jstor.org/stable/26977866

Freire, M., Serrano-Laguna, Á., Manero, B., Martínez-Ortiz, I., Moreno-Ger, P., Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. In M. Spector, B. Lockee, & M. Childress (Eds.), *Learning, design, and technology* (pp. 1–29). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-319-17727-4_21-1

Fu, J., Zapata, D., & Mavronikolas, E. (2014). Statistical methods for assessments in simulations and serious games. *ETS Research Report Series, 2014*(2), 1–17. https://doi.org/10.1002/ets2.12011

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Gauthier, A., Corrin, M., & Jenkinson, J. (2015). Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. *Computers & Education*, *87*, 24–34. https://doi.org/10.1016/j.compedu.2015.03.017

Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (pp. 21–40). MIT Press. https://doi.org/10.1080/13691180802552890

Gomez, M. J., Ruipérez-Valiente, J. A., Martínez, P. A., & Kim, Y. J. (2021). Applying learning analytics to detect sequences of actions and common errors in a geometry game. *Sensors*, *21*(4), Article 1025. https://doi.org/10.3390/s21041025

Gomez, M. J., Ruipérez-Valiente, J. A., Martinez, P. A., & Kim, Y. J. (2020). Exploring the affordances of sequence mining in educational games. *Proceedings of the 8th International Conference on Technological Ecosystems for Enhancing Multiculturality* (TEEM '20), 21–23 October, Salamanca, Spain (pp. 648–654). ACM Press. https://doi.org/10.1145/3434780.3436562

Gursoy, M. E., Inan, A., Nergiz, M. E., & Saygin, Y. (2017). Privacy-preserving learning analytics: Challenges and techniques. *IEEE Transactions on Learning Technologies*, *10*(1), 68–81. https://doi.org/10.1109/TLT.2016.2607747

Hauge, J. B., Boyle, E., Mayer, I., Nadolski, R., Riedel, J. C., Moreno-Ger, P., Bellotti, F., Lim, T., & Ritchie, J. (2014). Study design and data gathering guide for serious games' evaluation. In T. Connolly, T. Hainey, E. Boyle, G. Baxter, & P. Moreno-Ger (Eds.), *Psychology, pedagogy, and assessment in serious games* (pp. 394–419). IGI Global. http://doi:10.4018/978-1-4666-4773-2.ch018

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Horn, B., Hoover, A. K., Barnes, J., Folajimi, Y., Smith, G., & Harteveld, C. (2016). Opening the black box of play: Strategy analysis of an educational game. *Proceedings of the Annual Symposium on Computer–Human Interaction in Play* (CHI PLAY '16), 16–19 October 2016, Austin, Texas, USA (pp. 142–153). ACM Press. https://doi.org/10.1145/2967934.2968109

Hwang, G. J., Chu, H. C., & Yin, C. (2017). Objectives, methodologies and research issues of learning analytics. *Interactive Learning Environments*, *25*(2), 143–146. https://doi.org/10.1080/10494820.2017.1287338

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Sage.

Ifenthaler, D. (2017). Designing effective digital learning environments: Toward learning analytics design. *Technology, Knowledge and Learning*, *22*(3), 401–404. https://doi.org/10.1007/s10758-017-9333-0

Ifenthaler, D., Mah, D. K., & Yau, Y. K. (2019). Utilising learning analytics for study success: Reflections on current empirical findings. In D. Ifenthaler, D. K. Mah, & Y. K. Yau (Eds.), *Utilizing learning analytics to support study success* (pp. 27–36). Springer. https://doi.org/10.1007/978-3-319-64792-0_2

Karumbaiah, S., Baker, R. S., Barany, A., & Shute, V. (2019). Using epistemic networks with automated codes to understand why players quit levels in a learning game. In B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Eds.), *Advances in quantitative ethnography* (Vol. 1112, pp. 106–116). Springer. https://doi.org/10.1007/978-3-030-33232-7_9

Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior, 72*, 757–770. https:// doi.org/10.1016/j.chb.2016.09.062

Kang, J., Liu, S., & Liu, M. (2017). Tracking students' activities in serious games. In F. Q. Lai., & J. D. Lehman (Eds.). *Learning and knowledge analytics in open education* (pp. 125–137). Springer. https://doi.org/10.1007/978-3-319-38956-1_10

Kim, Y. J., & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. In D. Ifenthaler & Y. Kim (Eds.), *Game-based assessment revisited: Advances in game-based learning* (pp. 3–11). Springer. https://doi.org/10.1007/978-3-030-15569-8_1

Kopeinik, S., Nussbaumer, A., Bedek, M., & Albert, D. (2012). Using CbKST for learning path recommendation in game-based learning. *Proceedings of the 20th International Conference on Computers in Education* (ICCE 2012), 26–30 November 2012, Singapore (pp. 26–30). Asia-Pacific Society for Computers in Education. https://www.researchgate.net/publication/274256220_Using_CbKST_for_Learning_Path_Recommendation_in_Game-based_Learning

Krath, J., Schürmann, L., & von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, *125*, Article 106963. https://doi.org/10.1016/ j.chb.2021.106963

L'Heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. M. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, *5*, 7776–7797. https://doi.org/10.1109/ACCESS.2017.2696365

Law, E. L. C., & Lárusdóttir, M. K. (2015). Whose experience do we care about? Analysis of the fitness of scrum and kanban to user experience. *International Journal of Human–Computer Interaction*, *31*(9), 584–602. https://doi.org/10.1080/10447318.2015.1065693

Lee, J., & Liu, M. (2017). Design of fantasy and their effect on learning and engagement. In R. Z. Zheng & M. K. Gardner (Eds.), *Handbook of research on serious games for educational applications* (pp. 200–219). IGI Global. https://doi.org/10.4018/978-1-5225-0513-6.ch009

Li, F. Y., Hwang, G. J., Chen, P. Y., & Lin, Y. J. (2021). Effects of a concept mapping-based two-tier test strategy on students' digital game-based learning performances and behavioural patterns. *Computers & Education*, *173*, 104293. https://doi.org/10.1016/j.compedu.2021.104293

Linek, S., Öttl, G., & Albert, D. (2010). Non-invasive data tracking in educational games: Combination of logfiles and natural language processing. In L. Gómez Chova, D. Martí Belenguer, & I. Candel Torres (Eds.), *Proceedings of the 4th International Technology, Education and Development Conference* (INTED2010), 8–10 March 2010, Valencia, Spain (pp. 148–149). International Association of Technology, Education and Development. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.972.1032&rep=rep1&type=pdf

Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research and Development*, *53*(1), 5–21. https://doi.org/10.1007/BF02504854

Liu, M., Horton, L., Corliss, S. B., Svinicki, M. D., Bogard, T., Kim, J., & Chang, M. (2009). Students' problem solving as mediated by their cognitive tool use: A study of tool use patterns. *Journal of Educational Computing Research*, *40*(1), 111–139. https://doi.org/10.2190/EC.40.1.e

Liu, M., Li, C., & Pan, Z. (2022). Creating an interactive dashboard to support middle school teachers' implementation of a technology-supported problem-based learning program. *International Journal of Designs for Learning*, *13*(1). 1–18. https://doi.org/10.14434/ijdl.v13i1.31243

Liu, M., Horton, L., Olmanson, J., & Toprac, P. (2011). A study of learning and motivation in a new media enriched environment for middle school science. *Educational Technology Research and Development*, *59*(2), 249–266. https://doi.org/10.1007/s11423-011-9192-7

Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning analytics as an assessment tool in serious games: A review of literature. In M. Ma & A. Oikonomou (Eds.), *Serious games and edutainment applications*. Springer. https://doi.org/10.1007/978-3-319-51645-5_24

Liu, M., Lee, J., Kang, J., & Liu, S. (2016). What we can learn from the data: A multiple-case study examining behavior patterns by students with different characteristics in using a serious game. *Technology, Knowledge and Learning*, *21*(1), 33–57. https://doi.org/10.1007/s10758-015-9263-7

Liu, M., Li, C., Pan, Z., & Pan, X. (2019). Mining big data to help make informed decisions for designing effective digital educational games. *Interactive Learning Environments*, 1–21. https://doi:10.1080/10494820.2019.1639061

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In C. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics: Advances in game-based learning* (pp. 3–29). Springer. https://doi.org/10.1007/978-3-319-05834-4_1

Loh, C. S., & Sheng, Y. (2014). Maximum similarity index (MSI): A metric to differentiate the performance of novices vs. multiple-experts in serious games. *Computers in Human Behavior, 39*, 322–330. https://doi.org/10.1016/j.chb.2014.07.022

Loh, C. S., & Sheng, Y. (2015). Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies*, *20*(1), 5–19. https://doi.org/10.1007/s10639-013-9263-y

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction* (vol. 3, pp. 223–253). Lawrence Erlbaum. https://ocw.metu.edu.tr/pluginfile.php/2340/mod_resource/content/0/ceit706/week3/MakingLearningFun-ATaxonomyOfIntrinsicMotivationsForLearning.pdf

Mao, L., Liu, O. L., Roohr, K., Belur, V., Mulholland, M., Lee, H. S., & Pallant, A. (2018). Validation of automated scoring for a formative assessment that employs scientific argumentation. *Educational Assessment, 23*(2), 121–138. https://doi.org/10.1080/10627197.2018.1427570

Martin, T., Aghababyan, A., Pfaffman, J., Olsen, J., Baker, S., Janisiewicz, P., Phillips, R., & Smith, C. P. (2013). Nanogenetic learning analytics: Illuminating student learning pathways in an online fraction game. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 165–169). ACM Press. https://doi.org/10.1145/2460296.2460328

Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment* (No. 800). National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://files.eric.ed.gov/fulltext/ED522835.pdf

Oren, M., Pedersen, S., & Butler-Purry, K. L. (2021). Teaching digital circuit design with a 3-D video game: The impact of using in-game tools on students' performance. *IEEE Transactions on Education, 64*(1), 24–31. https://doi.org/10.1109/TE.2020.3000955

Owen, V. E., Roy, M.-H., Thai, K. P., Burnett, V., Jacobs, D., Keylor, E., & Baker, R. S. (2019). Detecting wheel-spinning and productive persistence in educational games. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (EDM2019), 2–5 July 2019, Montréal, Quebec, Canada (pp. 378–383). International Educational Data Mining Society. https://files.eric.ed.gov/fulltext/ED599202.pdf

Park, J., Kim, S., Kim, A., & Mun, Y. Y. (2019). Learning to be better at the game: Performance vs. completion contingent reward for game-based learning. *Computers & Education*, *139*, 1–15. https://doi.org/10.1016/j.compedu.2019.04.016

Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior, 63*, 50–58. https://doi.org/10.1016/j.chb.2016.05.023

Rapp, A. (2017). Designing interactive systems through a game lens: An ethnographic approach. *Computers in Human Behavior*, *71*, 455–468. https://doi.org/10.1016/j.chb.2015.02.048

Rosenheck, L., Cheng, M. T., Lin, C. Y., & Klopfer, E. (2021). Approaches to illuminate content-specific gameplay decisions using open-ended game data. *Educational Technology Research and Development*, *69*(2), 1135–1154. https://doi.org/10.1007/s11423-021-09989-0

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., Brown, R. A., & Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*, *76*, 617–630. https://doi.org/10.1016/j.chb.2017.03.043

Ruipérez-Valiente, J. A., Rosenheck, L., & Kim, Y. J. (2019). What does exploration look like? Painting a picture of learning pathways using learning analytics. In D. Ifenthaler & Y. Kim (Eds.), *Game-based assessment revisited: Advances in game-based learning* (pp. 281–300). Springer. https://doi.org/10.1007/978-3-030-15569-8_14

Seng, W. Y., & Yatim, M. H. M. (2014). Computer game as learning and teaching tool for object oriented programming in higher education institution. *Procedia-Social and Behavioral Sciences, 123*, 215–224. https://doi.org/10.1016/j.sbspro.2014.01.1417

Shou, Z., Lu, X., Wu, Z., Yuan, H., Zhang, H., & Lai, J. (2020). On learning path planning algorithm based on collaborative analysis of learning behavior. *IEEE Access, 8*, 119863–119879. https://doi.org/10.1109/ACCESS.2020.3005793

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

Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter? Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education*, *82*, 378–392. https://doi.org/10.1016/j.compedu.2014.12.011

Squire, K. D. (2008). Video game–based learning: An emerging paradigm for instruction. *Performance Improvement Quarterly*, *21*(2), 7–36. https://doi.org/10.1002/piq.20020

Su, C. (2017). Designing and developing a novel hybrid adaptive learning path recommendation system (ALPRS) for gamification mathematics geometry course. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(6), 2275–2298. https://doi.org/10.12973/eurasia.2017.01225a

Sun, Z., Lin, C. H., Lv, K., & Song, J. (2021). Knowledge-construction behaviors in a mobile learning environment: A lag-sequential analysis of group differences. *Educational Technology Research and Development, 69*(2), 533–551. https://doi.org/10.1007/s11423-021-09938-x

Sutcliffe, A., & Hart, J. (2017). Analyzing the role of interactivity in user experience. *International Journal of Human–Computer Interaction*, *33*(3), 229–240. https://doi.org/10.1080/10447318.2016.1239797

van Leeuwen, A., Knoop-van Campen, C. A. N., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K–12. *Journal of Learning Analytics*, *8*(2), 6–21. https://doi.org/10.18608/jla.2021.7325

Vista, A., Awwal, N., & Care, E. (2016). Sequential actions as markers of behavioural and cognitive processes: Extracting empirical pathways from data streams of complex tasks. *Computers & Education*, *92*, 15–36. https://doi.org/10.1016/j.compedu.2015.10.009

Williams, J., & Rosenbaum, S. (2004). Learning paths: Increase profits by reducing the time it takes employees to get up-to-speed. John Wiley & Sons.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*(2), 217–266. https://doi.org/10.1177/1046878108321866

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education*, *55*(1), 178–187. https://doi.org/10.1016/j.compedu.2010.01.003

Zhu, H., Tian, F., Wu, K., Shah, N., Chen, Y., Ni, Y., Zhang, X., Chao, K., & Zheng, Q. (2018). A multi-constraint learning path recommendation algorithm based on knowledge map. *Knowledge-Based Systems, 143*, 102–114. https://doi.org/10.1016/j.knosys.2017.12.011