Education Tech Research Dev
https://doi.org/10.1007/s11423-022-10138-4

RESEARCH ARTICLE

Does slow and steady win the race?: Clustering patterns of students' behaviors in an interactive online mathematics game

Ji-Eun Lee[1] · **Jenny Yun-Chen Chan**[1] · **Anthony Botelho**[2] · **Erin Ottmar**[1]

Accepted: 25 June 2022

Abstract

Online educational games have been widely used to support students' mathematics learning. However, their effects largely depend on student-related factors, the most prominent being their behavioral characteristics as they play the games. In this study, we applied a set of learning analytics methods (*k*-means clustering, data visualization) to clickstream data from an interactive online algebra game to unpack how middle-school students' ($N=227$) behavioral patterns (i.e., the number of problems completed, resetting problems, reattempting problems, pause time before first actions) correlated with their understanding of mathematical equivalence. The *k*-means cluster analysis identified four groups of students based on their behavioral patterns in the game: *fast progressors*, *intermediate progressors*, *slow progressors*, and *slow-steady progressors*. The results indicated that students in these clusters, with the exception of *slow progressors*, showed significant increases in their understanding of mathematical equivalence. In particular, *slow-steady progressors*, who reattempted the same problem more often than other students, showed the largest absolute learning gains, suggesting that behavioral engagement played a significant role in learning. With data visualizations, we presented evidence of variability in students' approaches to problem solving in the game, providing future directions for investigating how differences in student behaviors impact learning.

Keywords Mathematics education · Cluster analysis · Data visualization · Educational games · Middle-school mathematics

✉ Ji-Eun Lee
jlee13@wpi.edu

Jenny Yun-Chen Chan
jchan2@wpi.edu

Anthony Botelho
abotelho@coe.ufl.edu

Erin Ottmar
erottmar@wpi.edu

[1] Department of Social Science & Policy Studies, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA

[2] School of Teaching and Learning, University of Florida, 2821 Norman Hall, PO Box 117048, Gainesville, FL 32611, USA

Published online: 02 August 2022

Introduction

Elementary and middle-school students' struggle with mathematics learning has been a persistent problem in the United States. According to the results of the National Assessment of Educational Progress (Hussar et al., 2020), only 41% of fourth-graders and 34% of eighth-graders are proficient in mathematics, and the progress in students' mathematics performance has been stagnant over the past several years. More concerning, Hussar et al. (2020) reported that the gap between high-performers and low-performers has widened; while high-performing students' math achievement slightly increased or remained the same compared to the previous assessment, low-performers' math achievement decreased for both fourth- and eighth-graders.

Among mathematical topics, *algebra* is considered a gatekeeper in middle-school mathematics and students' success in high school (Bush & Karp, 2013; Knuth et al., 2006). In particular, understanding of mathematical equivalence, one of the core algebraic concepts, is critical for students' deeper understanding of algebra and their further study in mathematics. However, previous studies have shown that many middle-school students hold misconceptions and continue to struggle with understanding this concept (Alibali et al., 2007; Booth & Davenport, 2013; Knuth et al., 2006).

One possible instructional method to address these issues and support students' mathematics learning is the use of online educational games. Previous studies have reported that well-designed educational games are effective in enhancing students' math skills, performance (Es-Sajjade & Paas, 2020; Vanbecelaere et al., 2020), engagement (Chang et al., 2016; Deater-Deckard et al., 2014; Moon & Ke, 2020), and decreasing math anxiety (Vanbecelaere et al., 2021). In particular, research has revealed that students with lower math achievement benefited more from online educational games than students with higher math achievement, even after accounting for their pretest scores (Chang et al., 2015; Shin et al., 2012), suggesting that these games may be a potential tool to reduce achievement gaps.

However, due to the higher interactivity and larger flexibility of educational game environments (e.g., offering flexible options and choices for students) compared to other types of educational technologies, the effects of educational games largely depend on various factors, such as the unique features of the game environments, learners' behavioral patterns in games, personal traits, cognitive states, and affective states (Martin et al., 2015; Shute et al., 2015). Moreover, although it is important to provide personalized learning environments that consider students' individual differences (e.g., cognitive, affective, behavioral variables) in terms of instructional design, most studies in the field tended to focus on examining whether the games improve performance. Much uncertainty still exists about the relation between students' behaviors or actions in games and their learning outcomes, particularly in online mathematics game contexts (Deater-Deckard et al., 2014; Vandewaetere et al., 2011). Thus, more research is needed to understand how different student behaviors relate to learning outcomes, which in turn informs how educators and game designers can support learners with various needs, abilities, or skills.

In recent years, advances in learning analytics have enabled researchers to look at student actions and behavioral patterns in educational technologies at a more fine-grained level. By leveraging log data collected in these computer-based tools, learning analytics have provided information and insights into students' in-game behaviors, the effectiveness of educational games, as well as improvement and validation of game design elements (Alonso-Fernandez et al., 2019; Cano et al., 2018). For instance, our previous work has focused on the design and evaluation of an algebraic learning game called *From Here to*

There! (FH2T) that implements perceptual learning theories into math problem solving. Classroom studies have revealed that the game is effective in improving students' mathematical understanding as well as decreasing mathematical errors (Chan et al., 2022a; Hulse et al., 2019; Ottmar et al., 2015). With the advancement of learning analytic techniques, we can examine the ways in which students play the game and the types of behavioral patterns that lead to an increase in their learning outcomes. By leveraging log data collected in FH2T, this study aims to unpack how students' different behavioral patterns influence their mathematical understanding through a person-centered approach that explores subgroups of students characterized by different types of behavioral patterns in the game. Specifically, the study addresses the following research questions:

1. How many different clusters emerge based on students' behavioral patterns in the game?
2. To what extent do the different clusters of students' behavioral patterns in the game predict changes in their understanding of mathematical equivalence?
3. How do the students' problem-solving processes and solution strategies vary across the clusters of students' behavioral patterns?

Background literature

Student behavioral patterns in online educational games

Online educational games have been found to enhance learners' engagement and motivation, which in certain contexts have also led to increases in their problem-solving skills and learning (Huang et al., 2020; Karakoç et al., 2020). The same results have also been shown in mathematics learning contexts. Specifically, previous studies have reported that core features of educational games (e.g., goal setting, ongoing feedback, authentic problem solving, challenges) led to increases in learners' math performance (Es-Sajjade & Paas, 2020; Tokac et al., 2019; Vanbecelaere et al., 2020) and engagement (Chang et al., 2016; Deater-Deckard et al., 2014; Moon & Ke, 2020).

Although online educational games can positively contribute to students' learning, studies have demonstrated that various moderators influence the effects of educational games, including: (1) game design characteristics (e.g., types of tasks, presence of scaffolding or helps, control choices), (2) study or research characteristics (e.g., intervention duration, learning context), and (3) learner related factors such as their personal traits (e.g., gender), cognitive states (e.g., prior knowledge), or affective states (e.g., engagement) (Clark et al., 2016; Shute et al., 2015; Vanbecelaere et al., 2021). More importantly, as educational games provide more interactive options for learners than standard testing or other educational technologies, learners subsequently exhibit larger variations in their behaviors while interacting with educational games (Kerr, 2015). Thus, how learners behave in games may play a significant role in their learning during or after the gameplay.

A number of studies have examined what behaviors lead to positive learning outcomes in educational games. First, some studies have found that in-game progress is positively associated with performance on a posttest (Martin et al., 2015; Shute et al., 2015). For example, Shute et al. (2015) investigated the relations among middle-school students' prior knowledge, persistence (i.e., a performance-based measure), in-game progress (i.e., measured by the number of trophies awarded in the game), and learning outcome (i.e., understanding of physics; measured by a posttest) after playing an online physics game. The results indicated that students' prior knowledge and in-game progress significantly predicted their physics understanding after the gameplay.

Another important behavioral factor found to be predictive of learning outcomes is students' propensity to retry problems. For instance, one study (Chen et al., 2020) examined which behavioral features best predicted students' performance in a digital game-based assessment. The findings showed that the number of students' retry attempts on tasks were the most influential predictors for their performance among 27 behavioral features included in the prediction model. In other words, students who reattempted tasks more frequently performed better on these tasks. Shute and Ventura (2013) noted that measures such as the number of attempts to solve a problem and the number of failures or retries before success in games might be indicators of students' persistence or conscientiousness. In addition, Strmečki et al. (2015) claimed that gamified online learning environments should provide learners opportunities to fail or attempt multiple times so that they could learn from their mistakes or previous experiences.

In previous work on FH2T, it has been found that the students' in-game progress (i.e., the number of problems completed in the game) predicts their posttest performance (Hulse et al., 2019). Further, pause time before problem solving (i.e., the amount of time the students pause before they initiate the first actions; Chan et al., 2022b) were positively associated with students' performance within an online algebra learning game. While the prior studies were able to identify certain measures correlated with student performance, they did not explore how these measures were related to each other to describe patterns of behavior. The current work, therefore, expands upon these prior studies to explore the relation between emerging patterns of behavior (i.e., in-game progress, pause time before problem solving, the number of retries, the number of reattempts) and demonstrated mathematical understanding.

Understanding of mathematical equivalence

Performance on algebra is an important predictor of later success in upper-level mathematics (National Mathematics Advisory Panel, 2008). An important aspect of algebra is understanding *mathematical equivalence*—having the knowledge that two sets are the same in quantity and can be interchangeable. However, elementary students struggle with this concept (Rittle-Johnson et al., 2011), and almost half of the middle-school students have misconceptions about the equals sign (Knuth et al., 2006). Given the importance of understanding mathematical equivalence in upper-level mathematics (Fyfe et al., 2018), it is crucial to identify interventions that promote these equivalence skills.

Several studies have tested interventions for improving students' understanding of equivalence in elementary school (Alibali et al., 2018; Blanton et al., 2015; McNeil et al., 2012). However, few studies were designed to promote students' understanding of equivalence at the middle-school level (McNeil, 2008), and even fewer studies leveraged technologies or educational games to improve middle-school students' understanding of equivalence. One example of such studies used spreadsheets, such as Excel, to emphasize the numerical meaning of equivalent expressions among seventh-grade students (Tabach & Friedlander, 2008).

The online learning game, FH2T, aims to improve students' algebra performance through their understanding of mathematical equivalence. In this game, students can dynamically transform mathematical expressions (e.g., $2+3$) into perceptually different but mathematically equivalent states (e.g., $6-1$), providing students an interactive experience with algebraic transformations. Prior work has found that playing FH2T improved middle-school students' understanding of mathematical equivalence (Chan et al., 2022a). The current study expands this work by investigating how students' behaviors and engagement within the game are related to their improved understanding of mathematical equivalence.

Students' use of solution strategies

Understanding and evaluating multiple solution strategies (i.e., ways to solve a problem) then selecting the most efficient strategy are core competencies in algebra (Lynch & Star, 2014; Star & Rittle-Johnson, 2008). For example, in a problem "$4(2+x)=12$", the equation can be solved by using a three-step standard strategy: (a) distribute 4 into the parentheses, (b) subtract 8 from both sides, and (c) divide both sides by 4. Alternatively, the equation can be solved with a two-step efficient strategy: (a) divide both sides by 4, and (b) subtract 2 from both sides. Although the second strategy is more efficient in terms of the steps compared to the first strategy, it requires students to notice that 12 is divisible by 4. Students applying the standard distribution procedure may not notice the arithmetic shortcut in the problem. Previous studies have found that students' broad knowledge about solution strategies was positively associated with their gains in conceptual knowledge, flexibility, as well as further learning in mathematics (Heinze et al., 2009; Rittle-Johnson & Star, 2007).

To better understand and support students' efficient algebraic problem solving, researchers have examined how factors such as mathematical knowledge influence students' strategy selection and efficiency of problem solving (Newton et al., 2020; Star & Rittle-Johnson, 2008). Some studies have found that students' use of solution strategies is influenced by their understanding of two core algebraic concepts: equivalence and variables (Bush & Karp, 2013; Knuth et al., 2005). However, as previously noted, many middle-school students have misconceptions about these topics (Stephens et al., 2013). Relatedly, students' prior knowledge in various solution strategies (i.e., equation solving procedures) also contributes to their efficiency of problem solving (Khng & Lee, 2009; Rittle-Johnson et al., 2009). However, to the best of our knowledge, no studies to date have examined how students' behavioral characteristics are associated with their mathematical problem-solving processes. Thus, this study investigates associations between students' different behavioral patterns in the game and their use of solution strategies when solving algebraic problems.

Methods

Participants

The participants of this study were 227 students from six middle schools located in the Southeastern United States. They were drawn from a larger randomized controlled study that investigated the efficacy of two educational technologies—the online algebraic learning game (i.e., FH2T) and traditional online problem sets—in Fall 2019 (Chan et al., 2022a). The initial sample consisted of 348 students who played the game (i.e., students in the FH2T condition), but we excluded 121 students who completed less than 50% of items on either the pretest or posttest for more accurate estimates of their understanding and performance. This resulted in a final sample of 227 students. Among the 227 students (55.9% male, 44.1% female), most of them (95.6%) were in sixth grade, and 4.4% were in seventh grade. Regarding students' instruction levels, 84.6% of the students were in advanced math classes, 9.3% were in support classes, and the remaining 6.2% were in on-level classes.

Materials: From Here to There! (FH2T)

FH2T (https://graspablemath.com/projects/fh2t) is a game-based dynamic algebraic notation tool developed based on theories of perceptual learning and embodied cognition to improve students' algebraic understanding (Hulse et al., 2019; Ottmar et al., 2015). One of the core features of the game is that numbers and mathematical symbols are reified as movable physical objects on the screen. In this way, students can tap, touch, and move numbers and symbols in object-like ways, which in turn provides opportunities for them to identify the underlying structure of algebraic expressions and realize that mathematical transformations are not static, re-copying of lines but dynamic.

In each problem in the game, students' task is to transform a mathematical expression from the starting state (*here*) to the mathematically equivalent goal state (*there*) using dynamic gesture-actions (e.g., moving, tapping). Figure 1 represents a sample problem with a series of actions in FH2T. As shown in Fig. 1a, the objective of this problem is to transform the starting expression (i.e., 9*4) into the goal state in a white box (i.e., 3*6*2). Students must find a path between two mathematically equivalent expressions and transform the starting state into the perceptually different goal using gesture-actions. If a student completes a problem in the most efficient way, in other words, with the fewest steps possible to reach the goal state (e.g., three steps: 9*4 → 3*3*4 → 3*3*2*2 → 3*6*2; Fig. 1b through 1h), then three clovers are given (see Fig. 1i). The number of clovers given is deducted if a student exceeds the fewest steps possible to reach the goal state.

Another important feature of the game is that it provides opportunities to *reset* or *reattempt* the problems as many times as the student would like. If a student clicks the restart

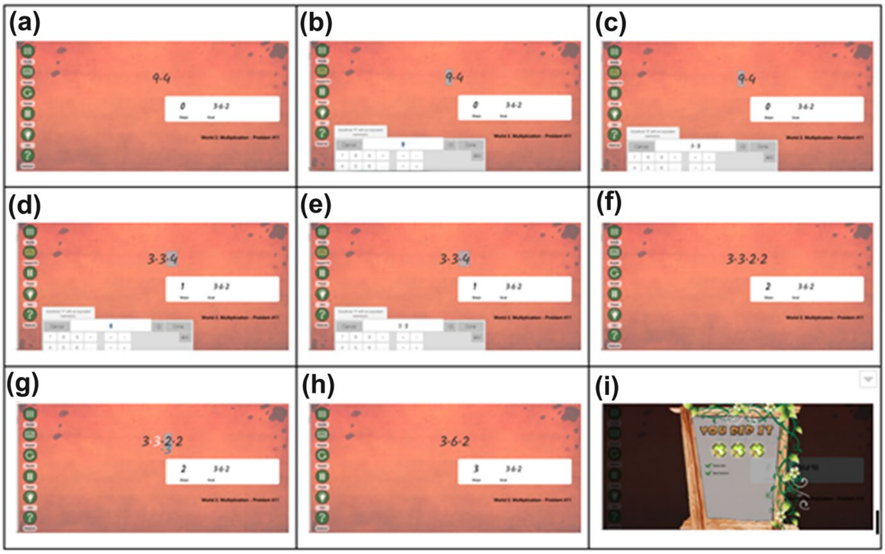

Fig. 1 A sample problem (a) and a potential solution involving three steps (factor 9 into 3*3 [b, c]; factor 4 into 2*2 [d, e]; multiply 3 and 2 [g]) to reach the goal (h) and receive three clovers (i)

button on the left side of the screen, the mathematical expression and the number of steps made reset to the initial state so that the student can restart the problem from the beginning. Students can also revisit the problem they completed and try to solve the problem again (i.e., reattempt) in a more efficient way. The game consists of 252 problems organized into 14 "worlds" that cover various mathematical concepts (e.g., addition, multiplication, fraction, division) arranged in the order of difficulty. Students can advance to the next world when they complete 14 consecutive problems in each preceding world.

Procedure

First, the students took a pretest using an online assessment system for 45 minutes. After completing the pretest, the students played the game using a device over four half-hour sessions (over 4 weeks) during their regular math classes. They played the game individually at their own pace, resulting in each student completing a different number of problems in the game ($M = 104.7$, $SD = 31.96$, $Min. = 31$, $Max. = 173$) by the end of the intervention. After the intervention, the students took a posttest using the same online assessment system for 40 minutes. The posttest consisted of the isomorphic items that mirrored the pretest. As students solved problems in FH2T, the system automatically recorded detailed log files of all students' touch- or mouse-based actions with timestamps, including measures such as the number of steps taken in solving each problem, the number of resets, the number of attempts, and all mathematical expressions made by the students.

Measures

Measures of student behavioral patterns

While the game logs recorded several different measures describing student interactions in the game, four key variables were selected for this study based on the review of previous literature. First, the *number of problems completed* refers to the distinct number of problems a student completed in the game during the intervention. For instance, even if a student solved the same problem more than once, it was counted as completion of one problem. Second, the *average number of resets* is computed by dividing the total number of resets by the total number of problems attempted. Resets in the game refer to clicking the reset button to set the expression back to the initial state and restart the problem *during* solving. Unlike resets, attempts refer to revisiting and trying the same problem again *after* completing the problem. The *proportion of reattempts* is calculated by dividing the total number of reattempts by the distinct number of problems completed during the intervention. For example, if a student solved 71 distinct problems during the intervention and reattempted some of the problems a total of 16 times, the proportion of reattempts for this student would be 0.23 (i.e., $16 \div 71$). Lastly, the *pause time* (seconds) represented the amount of time students spent (i.e., paused) before taking the very first interaction (e.g., step, error) on a problem in their first attempt to reach the goal state. We computed the average pause time by dividing the total amount of time students spent before taking the first action by the total number of problems attempted during the intervention.

Note that the four variables of student behavioral patterns were normalized using min–max scaling because the number of problems completed (Min = 31.00, Max = 173.00) and average pause time (Min = 6.49, Max = 106.12) had much larger scales than the average number of resets (Min = 0.11, Max = 2.56) and the proportion of reattempts (Min = 0.00, Max = 0.68). Normalizing variables can prevent the variables with a larger scale from overly influencing cluster analyses, as demonstrated in Kassambara (2017).

Understanding of mathematical equivalence

Students' understanding of mathematical equivalence was measured with six items selected from two established measures (Rittle-Johnson et al., 2011; Star et al., 2015) at pretest and posttest. The six items assessed students' ability to balance two sides of an equation (e.g., 898 + 13 = 896 + __), perform operations on both sides, define the equal sign, and identify equivalent expressions [e.g., (n + 3) + (n + 3) + (n + 3) + (n + 3) is equivalent to 4(n + 3)]. The questions were presented one at a time, and students entered their answers via the keyboard or selected a response option using a mouse on each question. Posttest items were similar to those of the pretest but with different numbers in the questions and the response options. Each item was scored as correct (1) or incorrect (0), and the total score out of 6 on the posttest was included as the outcome, and the pretest score was included as a covariate in the analyses. The Kuder–Richardson-20 coefficients of these six items were 0.63 at pretest and 0.64 at posttest, indicating acceptable reliability.

Productivity of students' solution strategies

We measured the productivity of students' solution strategies to explore qualitative differences in student behavior and qualitative differences in students' problem-solving processes in the game. Here, productivity refers to whether or not a student makes an appropriate mathematical transformation that brings the student closer to the target goal state of the selected problems. In particular, we focused on students' first mathematical transformation (hereafter, first step) to measure the productivity of their solution strategies as our previous research (Lee et al., 2022) suggested that students' first steps might influence their subsequent transformations as well as overall efficiency of problem solving.

Among the problems solved by the students, we selected one problem (labeled "problem A") from the multiplication topic because this problem showed the highest average number of resets and a higher proportion of reattempts compared to other problems. The students' task for this problem was to transform the start state (4*6*c*24*16) into the mathematically equivalent goal state (96*96*c) using gesture-actions learned in the game. Table 1 shows examples of productive and non-productive first steps.

As shown in Table 1, we coded the action of multiplying 6 and 16 to transform "4*6*c*24*16" into "4*c*24*96" as a productive first step because the student made a 96, which is a number in the goal state of the problem (i.e., 96*96*c). Contrary to this, transforming the start state into 4*6*c*384 by multiplying 24 and 16 was coded as a non-productive first step because this action did not bring the student closer to the goal state of the problem in a productive way. Two researchers hand-coded students' first steps as productive (1) or non-productive (0) and had a perfect agreement for this problem (i.e., There were no instances where the coders disagreed on the label of productivity). Finally, Table 2 summarizes the variables used in the study, including operational definitions of each variable and how we measure them.

Table 1 Examples of productive and non-productive first steps for problem A

Problem	Productivity	Actions taken	Expressions after First steps
Start state: 4*6*c*24*16 Goal state: 96*96*c	Productive first steps (1)	Multiplying 6 and 16	4*c*24*96
			4*96*c*24
		Multiplying 4 and 24	6*c*96*16
			96*6*c*16
	Non-productive first steps (0)	Multiplying 24 and 16	4*6*c*384
		Multiplying 6 and 24	4*144*c*16
		Multiplying 4 and 16	64*6*c*24
		Factoring 6 into 2 and 3	4*2*3*c*24*16

Data analyses

To address research question 1, we performed a *k*-means clustering analysis, one of the most commonly used unsupervised machine learning algorithms, to find groups of observations that had similar characteristics. The method works by comparing the distances between data points within the feature space with a specified *k* number of central points, or "centroids," that are iteratively calculated from the observed data. It is a particularly useful technique when researchers seek to gain a better understanding of how students in the sample are alike based on variables of interest in order to identify students' profiles that are grounded in their activities or behaviors (Antonenko et al., 2012; Martin et al., 2015). The analysis was performed using the "cluster" package in R (Maechler et al., 2014).

For cluster identification, we used the Euclidean distance measure with the Hartigan–Wong algorithm, which defines the total within-cluster variation as the sum of squared distances between items and the corresponding centroid (Boehmke & Greenwell, 2019). In order to validate the optimal number of clusters (i.e., cluster validation), we implemented the elbow method using the "fviz_nbclust" function in the "factoextra" R package (Kassambara & Mundt, 2016), which identifies *k* as the point where the reduction in the total within-cluster sum of squares drops significantly and produces an angle (an elbow point) in the graph (Antonenko et al., 2012). In other words, *k* is selected as the number of clusters that best describes the data such that adding more clusters would not provide meaningful gains in explaining the variance among and between groups of data points.

To examine whether each cluster of students, as identified by the behavioral patterns, significantly improved on their understanding of mathematical equivalence from pretest to posttest (research question 2), we first considered paired t-tests. However, the results of the Shapiro–Wilk tests indicated that the null hypothesis of normal distributions was rejected. Thus, we conducted Wilcoxon signed-rank tests and also computed absolute learning gain and normalized learning gain scores. Absolute learning gain scores were computed by taking the difference between posttest score and pretest score (i.e., posttest score − pretest score). Normalized learning scores were calculated by taking the ratio of the actual learning gain and the maximum possible gain (i.e., (posttest score − pretest score)/(6 − pretest score); 6 was the maximum possible score on pretest; Hake, 1998). Further, we conducted linear regression analyses to examine how much of the variance in students' later understanding of mathematical equivalence (i.e., posttest score) was explained by their behavioral patterns within the game (i.e., the results of the cluster analysis).

Table 2 Summary of the variables and measures included in the study

Variables	Operational definitions	Measures
Behavioral patterns (independent variables for clustering)		
Number of problems completed	The distinct number of problems completed	Sum of the distinct number of problems completed during the intervention
Average number of resets	Hitting the reset button in the game to set the equation back to an initial state and restart the problem	Total number of resets/total number of problems attempted
Proportion of reattempts	Revisiting to solve the same problem again after completing the problem	Total number of reattempts across problems/distinct number of problems completed
Average pause time (seconds)	The amount of time students spent before taking the first action in their first attempt to reach the goal state	Total amount of time spent before taking the first action/total number of problems attempted
Outcome variables of interest		
Understanding of mathematical equivalence (for pretest and posttest)	Students' knowledge of mathematical equivalence	Sum of the correctness on 6 items (correct = 1, incorrect = 0), ranges between 0 and 6
Productivity of solution strategies	Whether or not student's first mathematical transformation moves them towards the goal state	Productive solution strategy = 1, non-productive solution strategy = 0

Lastly, in order to explore qualitative differences in students' problem-solving processes and solution strategies among the clusters (research question 3), we built Sankey diagrams, a data visualization technique that depicts the flow of a process (e.g., problem-solving process) or multiple paths between stages as well as the frequencies or quantities from one set of values to another using the width of lines (Riehmann et al., 2005). To generate Sankey diagrams, we used "plotly.js," a JavaScript data visualization library (Sievert, 2020). Sankey diagrams consist of two main components: nodes and links. In this study, each node indicates the steps (i.e., transformations of expressions) made by the students, and the thickness of a link (i.e., paths in the diagram) represents the number of students who made that mathematical transformation in the game. We also visualized productive and unproductive first steps using color.

Results

Descriptive statistics and correlation analysis

Means, standard deviations, minimum and maximum values, and correlation coefficients for all variables used in the study are presented in Table 3. As most of the variables (e.g., the proportion of reattempts, average number of resets, average pause time) were positively skewed, we conducted a Spearman correlation analysis to examine the associations between the variables. Regarding the association with the posttest scores, two variables showed statistically significant and strong positive correlations: pretest scores ($rs(225) = .63$, $p < .001$) and the number of problems completed in the game ($rs(225) = .60$, $p < .001$). Two variables had moderate negative correlations with the posttest scores: the average number of resets in the game ($rs(225) = -.25$, $p < .001$), and average pause time ($rs(225) = -.28$, $p < .001$). The proportion of reattempts did not have a statistically significant association with the posttest scores. Together, the preliminary analysis suggests that students' posttest scores may be associated with aspects of their behaviors in the game.

Table 3 Descriptive statistics and correlations for the overall sample ($N = 227$)

Variable	1	2	3	4	5	6
1. Posttest scores	–					
2. Pretest scores	.63***	–				
3. Number of problems completed	.60***	.53***	–			
4. Proportion of reattempts	.08	.07	−.05	–		
5. Average number of resets	−.25***	−.25***	−.21**	.17*	–	
6. Average pause time (seconds)	−.28***	−.25***	−.59**	−.11	−.10	–
M	4.30	3.86	104.66	.05	.72	16.37
SD	1.54	1.59	31.97	.10	.37	11.13
Min.	.00	.00	31.00	.00	.11	6.49
Max.	6.00	6.00	173.00	.68	2.56	106.12

***$p < .001$, **$p < .01$, *$p < .05$

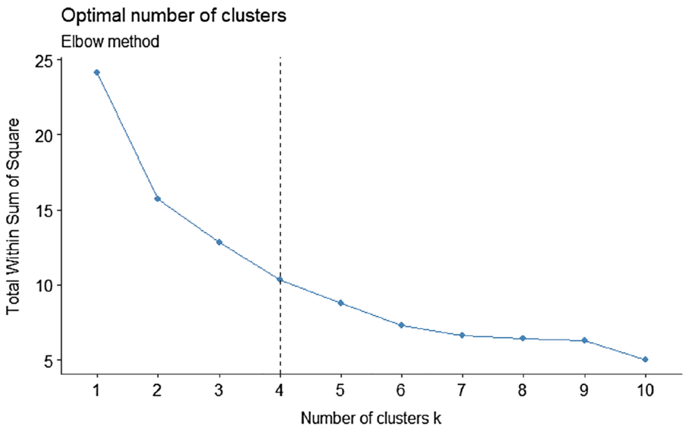

Fig. 2 The elbow method for determining the optimal number of clusters

Table 4 Descriptive statistics of each variable by cluster

Variables	Cluster 1: fast progressors ($n=60$)		Cluster 2: intermediate progressors ($n=116$)		Cluster 3: slow-steady progressors ($n=13$)		Cluster 4: slow progressors ($n=38$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number of problems completed	142.17	12.16	104.86	12.84	69.23	20.99	56.92	16.92
Average number of resets	0.55	0.21	0.75	0.31	0.82	0.54	0.87	0.55
Proportion of reattempts	0.03	0.04	0.04	0.06	0.38	0.13	0.03	0.05
Average of pause time (seconds)	11.23	3.08	14.36	5.42	14.03	6.02	31.40	18.67

Clustering based on the students' behavioral patterns

Before conducting *k*-means clustering analysis, we used the elbow method to determine the optimal number of clusters and validate the *k* value (see Fig. 2). As shown in Fig. 2, the "fviz_nbclust" R function suggested four cluster solutions (see the dotted line in Fig. 2), and the elbow point also indicated that $k=4$ would produce the optimal cluster results.

From the cluster analysis of the four behavioral variables within the game, the entire sample ($N=227$) was divided into four groups of students based on their behavioral patterns: Cluster 1 ($n=60$, 26.4%), Cluster 2 ($n=116$, 51.1%), Cluster 3 ($n=13$, 5.7%), and Cluster 4 ($n=38$, 16.8%). Table 4 shows the descriptive statistics of each variable by clusters. Figure 3 represents box plots of the four behavioral variables within the game by the cluster.

For interpretations of clustering results, we labeled four clusters based on the most salient characteristics that appeared in each cluster:

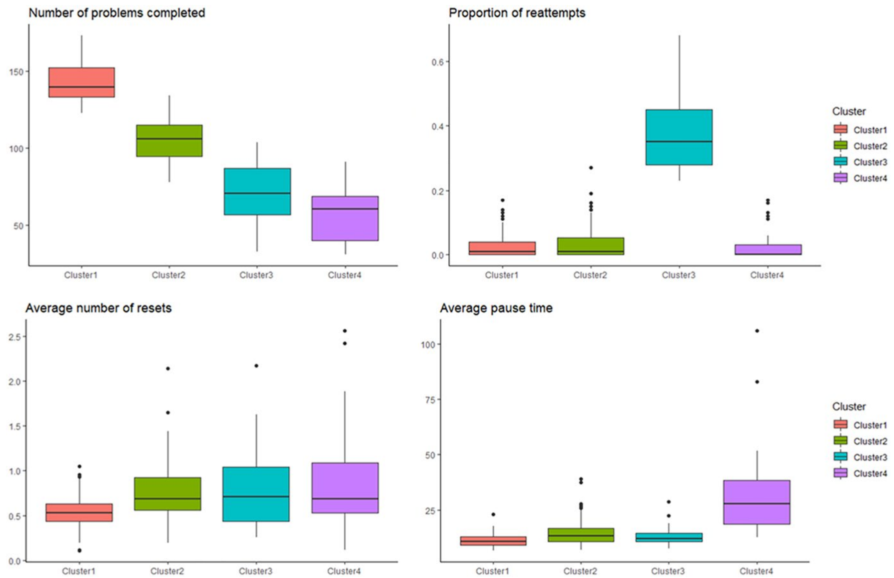

Fig. 3 Cluster results depicting the distribution of the behavioral variables for the four clusters. *Note* Cluster 1—*fast progressors*. Cluster 2—*intermediate progressors*. Cluster 3—*slow-steady progressors*. Cluster 4—*slow progressors*

1. *Fast progressors* (Cluster 1, $n=60$) included the students who had the highest number of problems completed but low values for the average number of resets, the proportion of reattempts, and the average pause time. The students in this cluster raced through the game and completed as many problems as possible, rather than retrying the problems again. The group also showed the lowest values of average pause time among the four clusters.
2. *Intermediate progressors* (Cluster 2, $n=116$) were the students with the medium values for all four variables (the number of problems completed, the average number of resets, the proportion of reattempts, average pause time). Approximately half of the students in the sample (51.1%) were included in this cluster.
3. *Slow-steady progressors* (Cluster 3, $n=13$) comprised of the students who showed the highest values for the proportion of reattempts, high values for the average number of resets, medium values for the average pause time, and low values for the number of problems completed. Only 5.7% of students in the sample belonged to this group. The students in this cluster explored the game slowly but steadily by retrying the problems they had completed before.
4. *Slow progressors* (Cluster 4, $n=38$) were the students with the lowest values for the number of problems completed and the proportion of reattempts, and the highest values for the average number of resets and the average pause time. The students in this cluster spent a longer amount of time before they made the first actions and reset the problems to the initial states more often than students in other clusters.

Changes in understanding of mathematical equivalence

Next, we examined whether there were significant improvements in students' understanding of mathematical equivalence from pretest to posttest within each cluster as identified by the cluster analysis of students' behavioral patterns. Table 5 presents means, standard deviations, medians of pretest and posttest scores, Z values, absolute learning gains, and normalized learning gains by the cluster. Although we did not include the students' pretest scores as input variables in the cluster analysis, there were descriptive differences in the pretest scores among the four clusters: fast progressors showed the highest pretest scores ($M=5.03$, $SD=1.03$, $MED=5.00$), while the slow progressors had the lowest pretest scores ($M=2.71$, $SD=1.59$, $MED=2.50$).

The Wilcoxon signed-ranks tests revealed that posttest ranks were statistically significantly higher than pretest ranks for three clusters: fast progressors ($Z=-2.333$, $p=.020$), intermediate progressors ($Z=-3.756$, $p<.001$), and slow-steady progressors ($Z=-1.983$, $p=.047$). The "slow progressors" cluster (Cluster 4) did not show a statistically significant difference between pretest and posttest scores.

Regarding the absolute learning gains, the "slow-steady progressors" showed the largest absolute learning gain scores ($g_{abs}=.77$) among the four clusters, followed by the "intermediate progressors" ($g_{abs}=.50$), "fast progressors" ($g_{abs}=.35$), and "slow progressors" ($g_{abs}=.26$) (see Fig. 4). Normalized learning gains for each cluster were further computed, and the values were .295 for fast progressors, .256 for slow-steady progressors, .157 for intermediate progressors, and −.032 for slow progressors, respectively.

We then conducted linear regression analyses to examine how much of the variance in students' posttest scores was explained by cluster identifiers after controlling for their prior knowledge (i.e., pretest scores). In order to conduct the analysis, we created three dummy variables (being in the fast progressors group, being in the slow-steady progressors group, being in the slow progressors group) by selecting the "intermediate progressors" as a reference group (being in each group was coded as 1 = Yes and 0 = No). Table 6 shows the results of the linear regression analyses predicting posttest scores.

As shown in Model 1 in Table 6, the three cluster identifiers explained a statistically significant amount of variance in posttest scores ($F(3, 223)=27.435$, $p<.001$, $R^2=.270$, $R^2_{Adjusted}=.260$). The results also indicated that students in the "fast progressors" group performed significantly better on the posttest compared to students in the "intermediate progressors" group ($\beta=.329$, $t(226)=5.434$, $p<.001$). The students in the "slow progressors" group performed significantly worse on the posttest compared to students in the "intermediate progressors" group ($\beta=-.309$, $t(226)=-5.132$, $p<.001$). The posttest performance in the "slow-steady progressors" group did not significantly differ from the "intermediate progressors" group.

Next, students' prior knowledge (pretest scores) was added to the model as a control variable. The model explained 48% of the variance in posttest scores ($F(4, 222)=51.181$, $p<.001$, $R^2=.480$, $R^2_{Adjusted}=.470$). In terms of the cluster predictor variables, the students in the fast progressors group scored significantly higher on the posttest ($\beta=.138$, $t(226)=2.508$, $p=.013$) compared to the students in the intermediate progressors group, although the magnitude of the association became lower after adding prior knowledge as a control variable. Specifically, the regression coefficient decreased by 58.1%, indicating that a part of the association between being in the ''fast progressors'' group and the posttest scores was explained by students' prior

Table 5 Descriptive statistics, Wilcoxon signed-rank results, and learning gains for four clusters

Cluster	Pretest			Posttest			Wilcoxon signed-rank		Learning gain	
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*Z*	*p*	g_{abs}	*g*
Fast progressors ($n = 60$)	5.03	1.03	5.00	5.38	1.01	6.00	−2.333*	.020	.35	.295
Intermediate progressors ($n = 116$)	3.74	1.44	4.00	4.24	1.32	4.00	−3.756***	.000	.50	.157
Slow-steady progressors ($n = 13$)	2.85	1.63	3.00	3.62	2.06	4.00	−1.983*	.047	.77	.256
Slow progressors ($n = 38$)	2.71	1.59	2.50	2.97	1.46	3.00	−.974	.330	.26	−.032

***$p < .001$, *$p < .05$

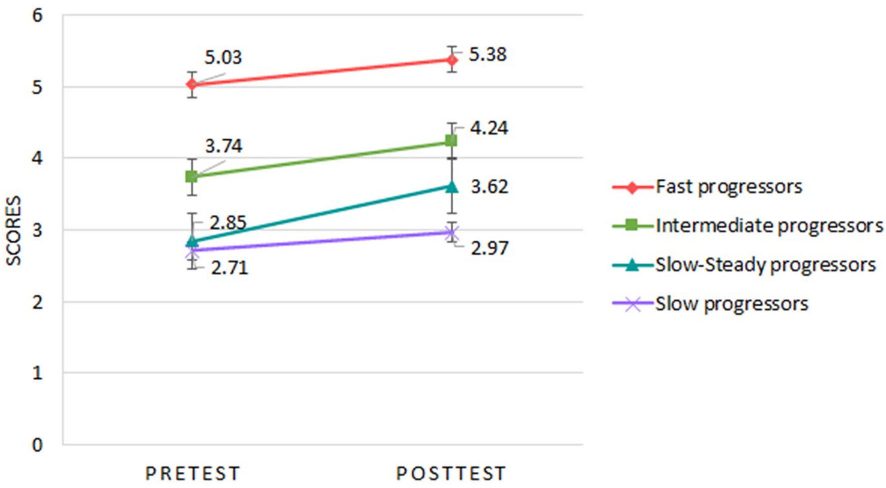

Fig. 4 Pretest and posttest scores by cluster

Table 6 Result of the multiple regression analysis predicting posttest scores

Predictors	*B*	*SE*	*β*	*t*	*p*	R^2	Adjusted R^2
Model 1						.270	.260
(Constant)	4.241	.123		34.556***	.000		
Fast progressors (1 = yes, 0 = no)	1.142	.210	.329	5.434***	.000		
Slow-steady progressors (1 = yes, 0 = no)	− .626	.387	− .095	− 1.620	.107		
Slow progressors (1 = yes, 0 = no)	− 1.268	.247	− .309	− 5.132***	.000		
Model 2						.480	.470
(Constant)	2.322	.228		10.195***	.000		
Prior knowledge (pretest scores)	.513	.054	.532	9.470***	.000		
Fast progressors (1 = yes, 0 = no)	.479	.191	.138	2.508*	.013		
Slow-steady progressors (1 = yes, 0 = no)	− .167	.331	− .025	− .504	.615		
Slow progressors (1 = yes, 0 = no)	− .739	.216	− .180	− 3.416**	.001		

***$p < .001$, **$p < .01$, *$p < .05$

knowledge. Second, the students in the slow progressors group scored significantly lower on posttest compared to those in the intermediate progressors group ($\beta = -.180$, $t(226) = -3.416$, $p = .001$). Lastly, the non-significant beta coefficient for the students in the slow-steady progressors group suggested that they performed similarly well on the posttest compared to students in the intermediate progressor group.

Qualitative differences in the students' use of mathematical strategies

Lastly, we explored the qualitative differences in students' use of mathematical strategies based on their behavioral patterns identified from the cluster analysis in one sample problem (labeled as problem A). The students' task for this problem was to transform the start state (4*6*c*24*16) into the mathematically equivalent goal state (96*96*c) using a set of mathematical strategies. We first examined the productivity of initial solution strategies for "problem A" by the cluster (see Fig. 5). As shown in Fig. 5, the fast progressors (71.7%) showed the highest percentage of students who made a productive first step, followed by the intermediate progressors (57.8%), the slow-steady progressors (53.8%), and the slow progressors (47.4%).

Next, we explored variations in students' problem-solving processes through data visualizations. Figure 6 depicts Sankey diagrams of problem-solving processes by cluster, and colors in each diagram indicate the productivity of students' first steps (blue: productive, red: non-productive) (For work that introduces the use of Sankey diagrams for exploring problem-solving variations in more detail, see Lee et al., in press). Note that the minimum number of steps required to solve this problem is two steps (e.g., 4*6*c*24*16 → 96*6*c*16 [1 step] → 96*96*c [2 steps]).

As shown in Fig. 6, there was a huge variability in students' problem-solving processes across each of the clusters. Among the four clusters, the students in the fast progressors group (Fig. 6a) reached the goal state using the lowest average number of steps ($M = 4.42$, $SD = 3.58$), and 30% of students ($n = 18$) in this group solved the problem in the most efficient way (i.e., using two steps). Contrary to this, the slow-steady progressors group (Fig. 6c) took a higher average number of steps ($M = 9.00$, $SD = 7.96$) on their first attempt than the students in the other three clusters, and only one student in this group reached the goal state in the most efficient way (i.e., using two steps). The students in the intermediate progressors group (Fig. 6b) and the slow progressors group (Fig. 6d) showed similar values in terms of the average number of steps (intermediate progressors = 5.57 steps, slow

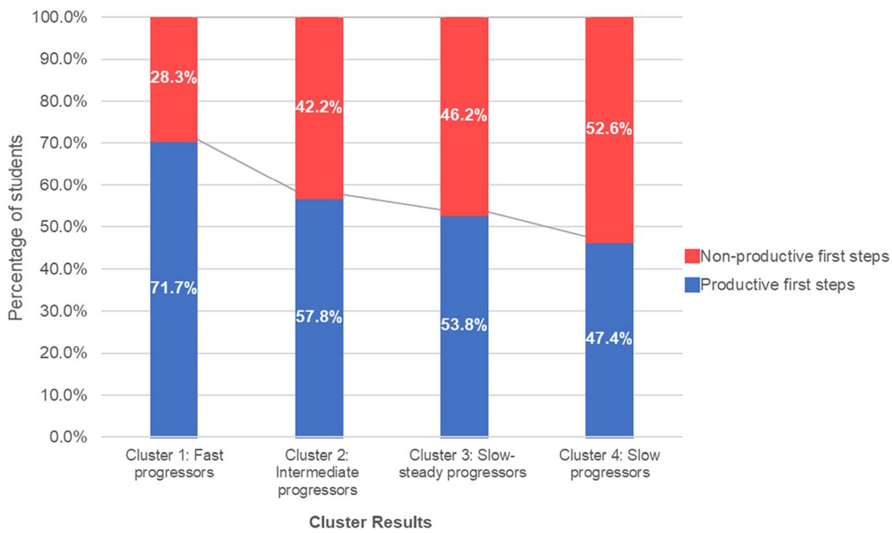

Fig. 5 Frequencies of productive vs. non-productive first steps by cluster (Color figure online)

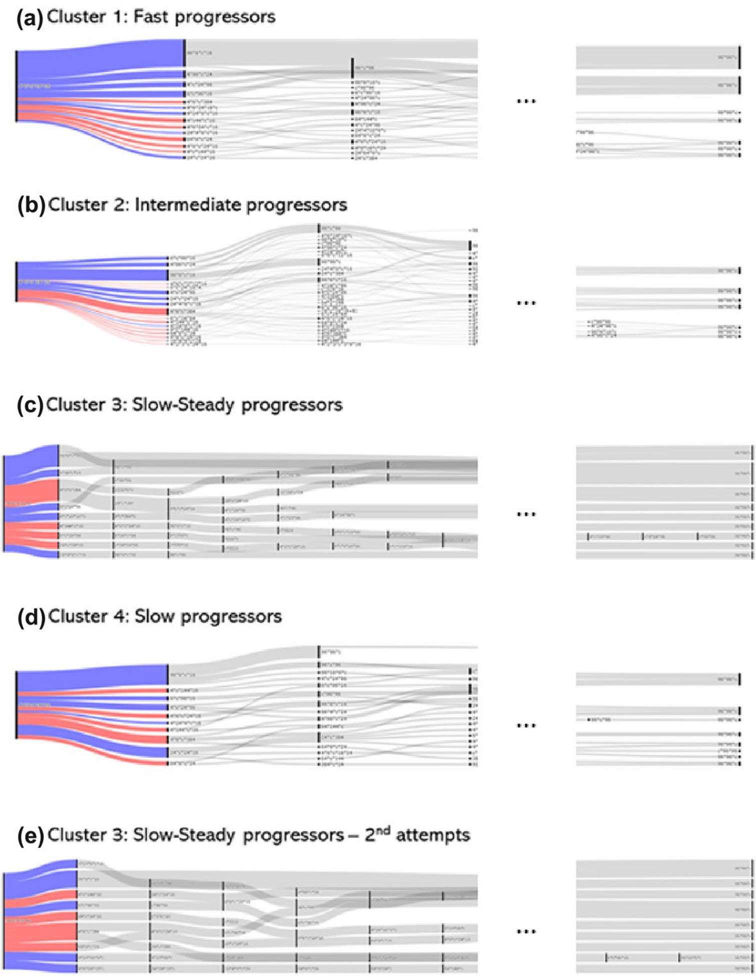

Fig. 6 The proportion of students' mathematical strategies used for problem A (start state: 4*6*c*24*16, goal state: 96*96*c) by cluster. *Note* For full image: https://tinyurl.com/2s489xyh

progressors = 5.05 steps) and the percentage of students who completed the problem in the most efficient way (intermediate progressors = 15.5%, slow progressors = 15.8%).

For the slow-steady progressors, we further investigated the students' second attempts for the same problem. The results indicated that 11 out of 13 students (84.6%) in slow-steady progressors reattempted this problem, while 10% or less of the students in the other three groups (fast progressors: 10.0%, intermediate progressors: 8.6%, slow progressors: 5.3%) reattempted the problem. Thus, we built another Sankey diagram to examine how the slow-steady progressors' problem-solving processes differed from their first attempts (Fig. 6e). As shown in Fig. 6e, slightly more students (54.5%) made productive first steps on their second attempt compared to their first attempt (53.8%), and the average number of steps ($M = 8.09$, $SD = 6.47$) became slightly lower in comparison with their first attempts. Interestingly, the students in the slow-steady progressors group did not necessarily solve the problem in a more efficient way in their second attempts but tried to explore different solutions to reach the goal state. For instance, one student in the slow-steady progressors group made a productive first step and completed the problem using two steps in the first attempt (e.g., 4*6*c*24*16→[step 1] 96*6*c*16→[step 2] 96*96*c). In the second attempt, this student made two more steps to reach the goal state than the first attempt but explored a different way to solve the problem (e.g., 4*6*c*24*16→[step 1] 6*c*96*16→[step 2] 96*c*96→[step 3] c*96*96→[step 4] 96*96*c). These findings suggest that slow-steady progressors may explore various solution strategies by reattempting the problems multiple times, and this behavioral pattern may lead to increased learning gains that reduce the gap between slow-steady progressors' and intermediate progressors' mathematical equivalence scores at posttest.

Discussion

While online educational games have been used to support students' mathematics learning, the effects of the games largely depend on various factors, in particular, students' behavioral patterns in the game (Martin et al., 2015; Shute et al., 2015). This study attempts to explore the subgroups of students that are characterized by different types of behavioral patterns in an online algebra learning game, FH2T, and unpack how students' behavioral patterns impact their understanding of mathematical equivalence.

For the first research question, we examined how students were grouped into different clusters based on their behavioral patterns (the number of problems completed, the average number of resets, the proportion of reattempts, the average pause time) in the game. The *k*-means clustering identified four groups of students: fast progressors (26.4%) who had the highest number of problems completed, intermediate progressors (51.1%) with the medium values of all behavioral variables, slow-steady progressors (5.7%) who showed the highest values of reattempts, and slow progressors (16.8%) with the lowest number of problems completed. Although most of the students (84.6%) in our sample were in advanced math classes, we still observed notable variability in their behavioral patterns in the game. Further, the behavioral patterns uniquely predicted student learning outcomes (i.e., posttest score) above and beyond their prior knowledge in mathematics, suggesting that the ways in which students worked with interactive educational games, like FH2T, were related to their learning. Therefore, it may be important to consider these individual differences while designing online educational games in order to provide an adaptive learning environment that supports learners with different behavioral patterns (Vandewaetere et al., 2011). These cluster results can form a foundation for designing an adaptive educational game environment based on learner profile data.

The second research question sought to examine how students' different behavioral patterns influenced their understanding of mathematical equivalence. The results indicated that the students in three clusters—fast progressors, intermediate progressors, slow-steady progressors—showed statistically significant increases in their understanding of equivalence after playing the game. Slow progressors, who completed the least number of problems in the game and had the lowest values of reattempts, did not show a significant increase in their understanding of mathematical equivalence.

More specifically, fast progressors, who had the highest pretest score on mathematical equivalence, solved more problems than students in the other clusters and rarely reset or reattempted problems; they also showed the highest posttest scores among the four clusters. This finding aligned with previous literature that the in-game progress was positively associated with student performance (Hulse et al., 2019; Martin et al., 2015; Shute et al., 2015). However, in terms of learning gains, slow-steady progressors, who did more resets and reattempts, showed the largest absolute learning gain scores among the four clusters. More importantly, although slow-steady progressors and slow progressors had similar starting points in terms of their prior knowledge before the intervention (i.e., pretest scores), only slow-steady progressors, who repeatedly reattempted the problems in the game, made significant increases in their understanding of mathematical equivalence after the intervention. Thus, these results corroborated the findings of other studies, which showed that students' retries or reattempts in the game positively correlated with their performance, and these behaviors might be indicators of students' persistence, productive failure, or conscientiousness (Chen et al., 2020; Shute & Ventura, 2013).

Furthermore, while some studies have found that online educational games were more beneficial to low-performing students compared to high-performing students (Shin et al., 2012), our results showed that it was not beneficial to all low-performing students. Among the low-performing students, only those with a high level of behavioral engagement or persistence as measured by reattempts in the game showed significant improvement in their mathematical understanding. Further, as indicated by the regression analyses, the slow-steady progressors were catching up to the intermediate progressors at posttest, whereas the slow progressors were not. In addition, the behavioral patterns differed between fast, intermediate, and slow-steady progressors, yet all three groups improved on their understanding of mathematical equivalence, suggesting that there might be multiple pathways to effective learning from educational games. The findings on the slow vs. slow-steady progressors suggest that perhaps a future direction is to examine whether prompting students, especially low-performing students, to retry and reattempt problems can effectively improve their mathematical learning. If the findings align with the current results, instructional design of games can aim to foster persistence among students so they can learn from their previous mistakes and experiences (Strmečki et al., 2015) and explore multiple ways to solve the same problems. For example, a feature that enables students to review or reattempt the problem that they just solved can be added to a game so that they can explore a different strategy to solve the problem.

Lastly, we used data visualization techniques, specifically Sankey diagrams, to investigate the variability in students' solution strategies within and across clusters for one problem from the multiplication topic. This process revealed that there was a huge variability in students' solution strategies, productivity, and efficiency across the clusters. Specifically, most students in the fast progressors cluster made productive first steps and solved the problem in more efficient ways than their peers in other clusters. For the slow-steady progressors, almost half of the students made non-productive first steps, and many of them solved the problems in less efficient ways in comparison with the students in other clusters.

However, more than 80% of the students in the slow-steady progressors group attempted the problem again, so we further investigated the differences in their solution strategies between their first and second attempts to unpack what led to the largest learning gains among the four clusters. The results revealed that the students did not necessarily solve the problem in more efficient ways compared to their first attempts, but they tried to solve the same problem in a different way. Although these visualizations provide only a qualitative look into student strategies and behaviors by cluster on one problem, the findings suggest that resetting and reattempting may prompt thinking about various ways to solve the problems and may lead to increases in students' understanding of mathematical equivalence. This supports the findings of previous studies that understanding multiple solution strategies positively influences students' gains in conceptual knowledge, procedural knowledge, flexibility, as well as further learning in mathematics (Heinze et al., 2009; Rittle-Johnson & Star, 2007).

Limitations and future directions

A number of important limitations need to be considered. First, our study focused on only four behavioral variables in the game based on our review of previous literature, and other variables, such as errors or use of hints, were not considered. Further research might explore how other behavioral features influence students' in-game performance and post-test scores. Second, in order to investigate the qualitative differences among the clusters, the current study only looked at students' use of mathematical strategies. Future studies should explore other qualitative differences among the clusters, such as demographic characteristics, math anxiety, or math self-efficacy, which are known correlates of students' mathematical performance. Third, although we tried to qualitatively unpack what led to the largest learning gains for slow-steady progressors by comparing the strategies in their first attempts and second attempts, we only examined the differences in one problem in the game. Further, the current analyses only included a subset of students who completed at least 50% of the pretest and posttest in the larger randomized controlled trial. While the current sample might be more motivated or better performing in math compared to the initial randomized sample, the students' algebraic performance was not subject to the ceiling effect, and we still observed differences in their problem-solving strategies. A further study with a more representative sample should be conducted to investigate the differences in students' attempts for other problems or temporal changes in their gameplay. Despite the greatest gains in the slow-steady progressors cluster, only a small number of students were classified in this group. Future work should replicate the current findings with a larger sample. If the patterns of results are consistent across studies, another possible area of research would be to identify design features that encourage students to do more retries in the game. This work will advance our understanding of factors that lead to more reattempts and larger learning gains among students. Finally, it is important to note that our results (i.e., student classification, cluster labeling) are game-specific, so the results should be interpreted in regard to this context. Although learner profiling can provide useful information for designing an adaptive educational game environment, there are ethical concerns that students may be stereotyped or categorized in negative ways. As Tzimas and Demetriadis (2021) argued, learner profiling should not define students nor limit their learning experiences.

Conclusion

This study demonstrates the usefulness of cluster analyses for identifying key patterns of student behaviors from log data that are recorded as students play educational games. This study identifies four unique patterns of student behaviors related to students' resetting, reattempting, problems completed, and pause times. The findings form the foundation for (a) interpreting patterns of student problem-solving strategies and perseverance in complex log data, (b) using the behavioral patterns to predict learning gains, and (c) leveraging data visualization techniques to explore students' problem-solving approaches within educational games. They also have implications for fostering various behaviors within interactive educational games to help students steadily win the race of mathematics.

Acknowledgements We would like to thank the teachers and students for their participation, members of the Math, Abstraction, Play, Learning, and Embodiment (MAPLE) Lab for their work, members of the Learning and Development Lab for their assistance in coding strategies, Erik Weitnauer and members of Graspable Math Inc. for programming and data support, and Neil Heffernan and the ASSISTments Team for their support.

Funding The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180401 to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Declarations

Conflict of interest Erin Ottmar was a designer and co-developer of From Here to There! and Graspable Math. Three other authors declare that they have no conflict of interest.

Ethical approval All procedures performed in the study were approved by Institutional Review Board (IRB). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent is not required for this study.

References

Alibali, M. W., Crooks, N. M., & McNeil, N. M. (2018). Perceptual support promotes strategy generation: Evidence from equation solving. *British Journal of Developmental Psychology, 36*(2), 153–168. https://doi.org/10.1111/bjdp.12203

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning, 9*(3), 221–247. https://doi.org/10.1080/10986060701360902

Alonso-Fernandez, 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*, 103612. https://doi.org/10.1016/j.compedu.2019.103612

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60*(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J. S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education, 46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Boehmke, B., & Greenwell, B. M. (2019). *Hands-on machine learning with R*. CRC Press. https://doi.org/10.1201/9780367816377

Booth, J. L., & Davenport, J. L. (2013). The role of problem representation and feature knowledge in algebraic equation-solving. *The Journal of Mathematical Behavior, 32*(3), 415–423. https://doi.org/10.1016/j.jmathb.2013.04.003

Bush, S. B., & Karp, K. S. (2013). Prerequisite algebra skills and associated misconceptions of middle grade students: A review. *The Journal of Mathematical Behavior, 32*(3), 613–632. https://doi.org/10.1016/j.jmathb.2013.07.002

Cano, A. R., Fernández-Manjón, B., & García-Tejedor, Á. J. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology, 49*(4), 659–672. https://doi.org/10.1111/bjet.12632

Chan, J.Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2022a). From here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology, 114*(1), 56. https://doi.org/10.1037/edu0000596

Chan, J.Y.-C., Ottmar, E. R., & Lee, J.-E. (2022b). Slow down to speed up: Longer pause time before solving problems relates to higher strategy efficiency. *Learning and Individual Differences, 93*, 102109. https://doi.org/10.1016/j.lindif.2021.102109

Chang, M., Evans, M. A., Kim, S., Norton, A., Deater-Deckard, K., & Samur, Y. (2016). The effects of an educational video game on mathematical engagement. *Education and Information Technologies, 21*(5), 1283–1297. https://doi.org/10.1007/s10639-015-9382-8

Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International, 52*(1), 47–57. https://doi.org/10.1080/09523987.2015.1005427

Chen, F., Cui, Y., & Chu, M. W. (2020). Utilizing game analytics to inform and validate digital game-based assessment with evidence-centered game design: A case study. *International Journal of Artificial Intelligence in Education, 30*(3), 481–503. https://doi.org/10.1007/s40593-020-00202-6

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

Deater-Deckard, K., El Mallah, S., Chang, M., Evans, M. A., & Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11–14 year olds. *International Journal of Child-Computer Interaction, 2*(3), 101–108. https://doi.org/10.1016/j.ijcci.2014.08.001

Es-Sajjade, A., & Paas, F. (2020). Educational theories and computer game design: Lessons from an experiment in elementary mathematics education. *Educational Technology Research and Development, 68*(5), 2685–2703. https://doi.org/10.1007/s11423-020-09799-w

Fyfe, E. R., Matthews, P. G., Amsel, E., McEldoon, K. L., & McNeil, N. M. (2018). Assessing formal knowledge of math equivalence among algebra and pre-algebra students. *Journal of Educational Psychology, 110*(1), 87. https://doi.org/10.1037/edu0000208

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM, 41*, 535–540. https://doi.org/10.1007/s11858-009-0214-4

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development, 68*(4), 1875–1901. https://doi.org/10.1007/s11423-020-09807-z

Hulse, T., Daigle, M., Manzo, D., Braith, L., Harrison, A., & Ottmar, E. (2019). From here to there! Elementary: A game-based approach to developing number sense and early algebraic understanding. *Educational Technology Research and Development, 67*(2), 423–441. https://doi.org/10.1007/s11423-019-09653-8

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Bullock Mann, F., Barmer, A., & Dilig, R. (2020). *The condition of education 2020* (NCES 2020-144). U.S. Department of Education. National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020144

Karakoç, B., Eryılmaz, K., Turan Özpolat, E., & Yıldırım, İ. (2020). The effect of game-based learning on student achievement: A meta-analysis study. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-020-09471-5

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). Sthda.

Kassambara, A., & Mundt, F. (2016). *Package 'factoextra'*. https://cran.microsoft.com/snapshot/2016-11-30/web/packages/factoextra/factoextra.pdf

Kerr, D. (2015). Using data mining results to improve educational video game design. *Journal of Educational Data Mining, 7*(3), 1–17.

Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences, 19*(2), 262–268. https://doi.org/10.1016/j.lindif.2009.01.004

Knuth, E. J., Alibali, M. W., McNeil, N. M., Weinberg, A., & Stephens, A. C. (2005). Middle school students' understanding of core algebraic concepts: Equivalence & variable. *ZDM Mathematics Education, 37*(1), 68–76. https://doi.org/10.1007/bf02655899

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312.

Lee, J. E., Hornburg, C. B., Chan, J. Y. C., & Ottmar, E. (2022). Perceptual and number effects on students' initial solution strategies in an interactive online mathematics game. *Journal of Numerical Cognition, 8*(1), 166–182. https://doi.org/10.5964/jnc.8323

Lee, J. E., Stalin, A., Ngo. V., Drzewiecki, K., Trac, C., & Ottmar, E. (in press). Show the flow: Visualization of students' solution strategies with Sankey diagrams in an online mathematics game. *Journal of Interactive Learning Research*.

Lynch, K., & Star, J. R. (2014). Teachers' views about multiple strategies in middle and high school mathematics. *Mathematical Thinking and Learning, 16*(2), 85–108. https://doi.org/10.1080/10986065.2014.889501

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2014). *Package 'cluster'*. https://cran.microsoft.com/snapshot/2014-10-10/web/packages/cluster/cluster.pdf

Martin, T., Petrick Smith, C., Forsgren, N., Aghababyan, A., Janisiewicz, P., & Baker, S. (2015). Learning fractions by splitting: Using learning analytics to illuminate the development of mathematical understanding. *Journal of the Learning Sciences, 24*(4), 593–637. https://doi.org/10.1080/10508406.2015.1078244

McNeil, N. M. (2008). Limitations to teaching children 2+ 2= 4: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development, 79*(5), 1524–1537. https://doi.org/10.1111/j.1467-8624.2008.01203.x

McNeil, N. M., Chesney, D. L., Matthews, P. G., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Wheeler, M. C. (2012). It pays to be organized: Organizing arithmetic practice around equivalent values facilitates understanding of math equivalence. *Journal of Educational Psychology, 104*(4), 1109–1121. https://doi.org/10.1037/a0028997

Moon, J., & Ke, F. (2020). In-game actions to promote game-based math learning engagement. *Journal of Educational Computing Research, 58*(4), 863–885. https://doi.org/10.1177/0735633119878611

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education.

Newton, K. J., Lange, K., & Booth, J. L. (2020). Mathematical flexibility: Aspects of a continuum and the role of prior knowledge. *The Journal of Experimental Education, 88*(4), 503–515. https://doi.org/10.1080/00220973.2019.1586629

Ottmar, E. R., Landy, D., Goldstone, R. L., & Weitnauer, E. (2015). Getting from here to there: Testing the effectiveness of an interactive mathematics intervention embedding perceptual learning. In *Proceedings of the thirty-seventh annual conference of the Cognitive Science Society* (pp. 1793–1798). Cognitive Science Society.

Riehmann, P., Hanfler, M., & Froehlich, B. (2005, October). Interactive Sankey diagrams. In *IEEE symposium on information visualization* (pp. 233–240). IEEE.

Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldoon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology, 103*(1), 85–104. https://doi.org/10.1037/a0021334

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836–852. https://doi.org/10.1037/a0016026

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology, 43*(4), 540–560. https://doi.org/10.1111/j.1467-8535.2011.01197.x

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M., & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education, 86*, 224–235. https://doi.org/10.1016/j.compedu.2015.08.001

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. The MIT Press.

Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and shiny*. https://plotly-r.com.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, 18*(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Stephens, A. C., Knuth, E. J., Blanton, M. L., Isler, I., Gardiner, A. M., & Marum, T. (2013). Equation structure and the meaning of the equal sign: The impact of task selection in eliciting elementary students' understandings. *The Journal of Mathematical Behavior, 32*(2), 173–182. https://doi.org/10.1016/j.jmathb.2013.02.001

Strmečki, D., Bernik, A., & Radošević, D. (2015). Gamification in e-learning: Introducing gamified design elements into e-learning systems. *Journal of Computer Science, 11*(12), 1108–1117. https://doi.org/10.3844/jcssp.2015.1108.1117

Tabach, M., & Friedlander, A. (2008). Understanding equivalence of symbolic expressions in a spreadsheet-based environment. *International Journal of Computers for Mathematical Learning, 13*(1), 27–46. https://doi.org/10.1007/s10758-008-9125-7

Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning, 35*(3), 407–420. https://doi.org/10.1111/jcal.12347

Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development, 69*(2), 1101–1133. https://doi.org/10.1007/s11423-021-09977-4

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes. *Computers & Education, 143*, 103680. https://doi.org/10.1016/j.compedu.2019.103680

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior, 27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

Vanbecelaere, S., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2021). The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, noncognitive and efficiency outcomes. *British Journal of Educational Technology, 52*(1), 112–124. https://doi.org/10.1111/bjet.12957

Ji-Eun Lee is a Research Scientist in the Learning Sciences and Technologies program at Worcester Polytechnic Institute. Her interests are applying learning analytics and educational data mining techniques to improve instructional design and student learning in online learning environments.

Jenny Yun-Chen Chan is an Assistant Professor of Early Childhood Education at the Education University of Hong Kong. Her research focuses on how contexts and students' cognitive skills influence math learning, with the goal of informing practices that adapt to students' needs.

Anthony Botelho is an Assistant professor of Educational Technology in the School of Teaching and Learning at the University of Florida. He seeks to impact learning by studying aspects of student cognition, behavior, and affect through the application of quantitative methods grounded in learning theory and is passionate about using a human-in-the-loop design approach to build that research into practice.

Erin Ottmar is an Associate Professor of Learning Sciences, Technology, and Psychology at Worcester Polytechnic Institute. Her research focuses on the design, development, and testing of innovative technology interventions that embed cognitive, developmental, perceptual, and educational principles of learning into everyday mathematics practice.