Received: 12 May 2022 | Accepted: 17 January 2023

DOI: 10.1111/bjet.13304

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Keep DRAGging ON: Is solving more problems in DragonBox 12+ associated with higher mathematical performance during the COVID-19 pandemic?

Jenny Yun-Chen Chan[1,2] | Chloe Byrne[1] | Janette Jerusal[1] | Allison S. Liu[1] | Justin Roberts[1] | Erin Ottmar[1]

[1]Social Science and Policy Studies, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

[2]Department of Early Childhood Education, The Education University of Hong Kong, Taipo, Hong Kong

Correspondence

Jenny Yun-Chen Chan, Department of Early Childhood Education, The Education University of Hong Kong, 10 Lo Ping Road, Taipo, Hong Kong SAR.
Email: chanjyc@eduhk.hk

Funding information

Institute of Education Sciences, Grant/Award Number: R305A180401

Abstract

Prior research has shown that game-based learning tools, such as DragonBox 12+, support algebraic understanding and that students' in-game progress positively predicts their later performance. Using data from 253 seventh-graders (12–13 years old) who played DragonBox as a part of technology intervention, we examined (a) the relations between students' progress within DragonBox and their algebraic knowledge and general mathematics achievement, (b) the moderating effects of students' prior performance on these relations and (c) the potential factors associated with students' in-game progress. Among students with higher prior algebraic knowledge, higher in-game progress was related to higher algebraic knowledge after the intervention. Higher in-game progress was also associated with higher end-of-year mathematics achievement, and this association was stronger among students with lower prior mathematics achievement. Students' demographic characteristics, prior knowledge and prior achievement did not significantly predict in-game progress beyond the number of intervention sessions students completed. These findings advance research on *how*, *for whom* and *in what contexts* game-based interventions, such as

DragonBox, support mathematical learning and have implications for practice using game-based technologies to supplement instruction.

KEYWORDS
algebraic learning, DragonBox 12+, in-game progress, mathematics achievement

Practitioner notes

What is already known about this topic

- DragonBox 12+ may support students' understanding of algebra but the findings are mixed.
- Students who solve more problems within math games tend to show higher performance after gameplay.
- Students' engagement with mathematics is often related to their prior math performance.

What this paper adds

- For students with higher prior algebraic knowledge, solving more problems in DragonBox 12+ is related to higher algebraic performance after gameplay.
- Students who make more in-game progress also have higher mathematics achievement, especially for students with lower prior achievement.
- Students who spend more time playing DragonBox 12+ make more in-game progress; their demographic, prior knowledge and prior achievement are not related to in-game progress.

Implications for practice and/or policy

- DragonBox 12+ can be beneficial as a supplement to algebra instruction for students with some understanding of algebra.
- DragonBox 12+ can engage students with mathematics across achievement levels.
- Dedicating time and encouraging students to play DragonBox 12+ may help them make more in-game progress, and in turn, support math learning.

INTRODUCTION

Algebraic knowledge is defined as the understanding of principles, procedures and strategies of working with generalized mathematical symbols (eg, letters, variables; Kieran, 2006). It provides an important foundation for higher-level STEM courses (Chen, 2013), yet students' often struggle with learning algebra (Kieran, 2006) and this struggle is even more apparent during the COVID-19 pandemic. Emerging studies have shown that students are making less progress in mathematics and other domains compared to their peers in prior years (Engzell et al., 2021; Kuhfeld et al., 2022). Preliminary research further showed that students' participation and engagement in online learning during the pandemic were lower than typical (Rutherford et al., 2021). Some research has shown that game-based learning—a method of learning where students can interact with educational materials in a way that is fun and playful—may support students' engagement with a topic and in turn improve learning (Connolly et al., 2012). We examine how students' engagement with an educational technology game, as measured by their in-game progress, may support their algebraic learning and mathematics achievement during the pandemic.

Game-based learning and in-game progress

Theories of game-based learning posit that learning occurs through playing games that are designed to improve educational outcomes (Plass et al., 2015). In this theoretical approach, learning is integrated with gameplay so that students acquire the targeted content through play. Some previous studies have shown the positive impacts of game-based learning (eg, Rowe et al., 2017; Sun-Lin & Chiou, 2017); however, other studies have not (eg, Attali & Arieli-Attali, 2015; Wrzesien & Raya, 2010). The increasing use of educational technology games combined with the mixed findings prompts more research to understand *how*, *for whom* and *in what contexts* these games support learning.

One potential mechanism of game-based learning is engaging students in the content and learning processes through play (Ke, 2008; Mora et al., 2015; Samur & Evans, 2012). Educational technology games often provide an interactive environment that maintains students' active engagement, allowing students to build up their knowledge and skills as they progress through games (Gee, 2003). Prior research has found that making more progress, which is defined as solving more problems or completing more exercises, within mathematical games positively predicts students' posttest scores on numeracy (Chan & Santos, 2022) and arithmetic (Hulse et al., 2019) even after controlling for their prior performance (ie, pretest scores). We contribute to this literature by testing the effects of in-game progress, as measured by the total number of completed problems, within DragonBox 12+ on students' later algebraic knowledge and mathematics achievement.

DragonBox 12+

DragonBox 12+ (hereafter DragonBox, https://dragonbox.com/products/algebra-12; Kahoot!, 2019) is an educational technology game that supports algebraic learning. In this game, students follow algebraic rules to manipulate digital characters and isolate a box containing a dragon, analogous to solving an equation for x (Figure 1). DragonBox has won several awards, including the Best App for Teaching and Learning in 2014 from the American Association of School Librarians. Its design has also been applauded for its research-backed principles, including embedded gestures and immediate feedback (Cayton-Hodges et al., 2015; Torres et al., 2016). Extensive research has suggested that mathematical learning is inherently perceptual, and students' physical experiences impact their mathematical thinking (Abrahamson et al., 2020; Alibali & Nathan, 2012; Goldstone et al., 2017; Kellman et al., 2010). In DragonBox, the active manipulation of symbols combined with the immediate feedback on students' actions may help students ground their abstract algebraic knowledge in physical experiences, developing an embodied routine for algebraic reasoning. This notion is supported by prior research examining a similar educational technology game that leverages action and perception in algebraic learning (eg, Chan, Lee, et al., 2022).

Several studies have found mixed effects of DragonBox on algebraic understanding and engagement. For example, in a sample of K-12 US students (5–18 years old) from 70 schools across 15 districts, the DragonBox developers and the University of Washington Center for Game Science found that 93% of students could correctly answer three consecutive algebra problems (eg, $d \times x + m = 8$) after 1.5 hours of playing a combined version of DragonBox 5+ and 12+ (Liu et al., 2015). Compared with a control group receiving typical instruction,

FIGURE 1 A sample problem and its problem-solving process within DragonBox. (a) Multiply the DragonBox by 2; (b) multiply all other terms on both sides by 2; (c) Factor 4 into 2·2; (d) Combine 2s in the numerator and denominator; (e) subtract both sides by 1/2; (f) The DragonBox is isolated and the user received 3 stars for solving the problem. This image is used with permission from DragonBox (Kahoot!).

eighth-grade Malaysian students (14 years old) in the DragonBox condition scored higher on assessments of algebraic thinking and mathematics attitudes after 16 hours of teaching and playing (Siew et al., 2016). Similarly, Indonesian middle-schoolers (12–15 years old) who used DragonBox as a supplemental activity in a semester-long algebra course outperformed their peers who did not use DragonBox on conceptual understanding (ie, the knowledge of algebraic principles, such as both sides of the equal sign represent the same quantity) and algebraic equation solving (Supriadi et al., 2020). Another study with eighth-graders in Norway (13–14 years old) found that after receiving 4 hours of instruction and using Dragon-

Box in pairs for another 4 hours within 4 weeks, students significantly improved their algebraic performance as measured by items selected from textbooks and standardized assessments (Dolonen & Kluge, 2015). Although students in that study showed high levels of usage and engagement with DragonBox, their learning gains were smaller compared with those of students who practised textbook-like algebra problems. Similarly, Long and Aleven (2017) found that seventh- and eighth-grade US students (12–14 years old) who played DragonBox for 3.5 hours across five sessions showed no improvements in problem-solving performance or confidence, although they reported higher enjoyment compared with students using an intelligent tutor system.

Prior studies typically involved testing the effects of playing DragonBox for 2–4 hours on algebraic learning and engagement. It remained unclear whether these improvements were associated with how much students played DragonBox, or whether improvements would have emerged regardless of the amount of in-game progress. Furthermore, to our knowledge, prior studies have only focused on DragonBox's impacts on algebraic performance. Given that algebraic knowledge supports general mathematics performance (eg, National Mathematics Advisory Panel, 2008), playing DragonBox may also have significant impacts on students' overall mathematics achievement. Finally, prior research did not explore the potential moderating effects of students' prior knowledge on the relation between playing DragonBox and learning outcomes. Addressing these questions is crucial for understanding the ways in which DragonBox supports mathematical learning.

Moderating effects of prior knowledge

Research has demonstrated that the effects of mathematics interventions vary depending on students' prior knowledge—the knowledge students already have and which interventions build upon. For example, among US first-graders (6–7 years old) with higher mathematics achievement, those who received an intervention involving peer-assisted supplemental activities showed greater gains in mathematics compared with those who did not. However, among students with lower mathematics achievement, those who did not receive the intervention showed greater gains compared with those who did (Wood et al., 2020). Other studies have reported mixed results, with some finding larger benefits of intervention among students with higher prior knowledge (eg, Swanson et al., 2008) and others finding the opposite (eg, Murphy et al., 2020). In the context of educational technology games, one study has examined the moderating effects of students' prior knowledge on the relation between in-game progress and later mathematical performance (Hulse et al., 2019). Solving more problems within the game was related to higher mathematical performance at posttest among students with lower, but not higher, prior knowledge, suggesting the select benefits of in-game progress among struggling students. To efficiently support all students in mathematical learning, it is important to understand when and for whom interventions are effective. We test the moderating effect of prior knowledge, as measured by students' algebraic knowledge and mathematics achievement prior to the intervention, on the relation between DragonBox progress and learning outcomes to contribute new findings regarding the role of prior knowledge in mathematical learning.

Factors associated with in-game progress

Finally, we explore whether students' characteristics, such as biological sex, race/ethnicity and prior knowledge and achievement are associated with their progress within DragonBox. Although prior studies have reported significant differences in sex and race/ethnicity in students' mathematics achievement, these differences tended to be related to situational contexts, such as educational experience and socioeconomic status (eg, Linn & Hyde, 1989; Sonnenschein & Galindo, 2015). For example, gender stereotypes have contributed to female students showing more negativity towards mathematics, less participation and worse achievement compared with their male counterparts (Nosek & Smyth, 2011). Similarly, Black female students reported lower levels of engagement in mathematics classes compared with other students (Martinez & Guzman, 2013). However, one study using the process data within an e-textbook found that female versus male students were more likely to spend more time and provide more detailed responses to mathematical questions, suggesting higher engagement (Reinhold et al., 2020). Furthermore, students with higher versus lower prior knowledge tend to report higher learning engagement (Dong et al., 2020) and show higher engagement in the learning process (Pecore et al., 2017). By examining whether students' in-game progress varies by their sex, race/ethnicity or prior knowledge or achievement, we aim to provide a starting point for future work that supports learning through games among all students from diverse backgrounds and at different levels.

The current study

The research questions (RQs) and their corresponding hypothesis are pre-registered (https://osf.io/eg2bu) and are as follows:

1. *Does students' in-game progress within DragonBox predict their later algebraic knowledge?* Based on prior research (Chan & Santos, 2022; Hulse et al., 2019), we hypothesize that students who solve more versus fewer problems in DragonBox may score higher on an algebraic knowledge posttest.
2. *Does the relation between in-game progress and later algebraic knowledge vary by students' prior knowledge?* Due to the mixed findings in prior research, we do not have a directional hypothesis. We explore whether the relation between progress and learning outcome varies by students' prior knowledge.
3. *Do the effects of in-game progress and its interaction with prior knowledge emerge in students' general mathematics achievement?* We explore these effects on students' end-of-grade mathematics achievement scores and do not have an a priori hypothesis.
4. *Do students' sex, race/ethnicity or prior knowledge or achievement predict their in-game progress?* We explore these associations to identify potential factors related to students' in-game progress. We do not have an a priori hypothesis.

We use the data collected from a randomized controlled trial (RCT) comparing the efficacy of four educational technology conditions to address our RQs. In the RCT, Decker-Woodrow et al. (in press) found that seventh-grade US students assigned to the DragonBox condition scored significantly higher on the algebraic knowledge post-test compared with students in the active control condition. Extending that work, we examine the impacts of DragonBox progress on students' mathematics performance and factors associated with in-game progress.

METHODS

Participants

The data were collected as a part of an RCT with seventh-graders during the 2020–2021 academic year amidst the COVID-19 pandemic. The students were recruited from a large, suburban school district that consisted of 10 physical schools and one virtual school in the Southeastern US. Due to the COVID-19 pandemic, the school district provided students and their families with two options for instruction: synchronous in-person instruction at school or asynchronous virtual instruction at home. Prior to the start of the academic year, the students and families selected their instructional format for the entire academic year, but they could change the format throughout the year. The research team randomly assigned students to one of four intervention conditions prior to the academic year, and the students used their assigned technology throughout the year (see details in Decker-Woodrow et al., in press).

The current study included 253 students who started the pretest, played DragonBox and started the post-test. Because our RQs focused on students' progress within DragonBox, the other three intervention conditions did not provide appropriate data for our RQs. The current sample was somewhat smaller than the original sample of 350 students who were assigned to the DragonBox condition and included in the prior analysis (Decker-Woodrow et al., in press). This discrepancy was due to the missing data in students' DragonBox progress. Because DragonBox is a privately-owned application that is not freely accessible through web browsers, the research team provided tablets installed with DragonBox for in-person students. Many virtual students struggled to follow written instructions to download the game for free on their own device and dropped out of the study.

We received students' demographic information and standardized achievement scores before and after the intervention directly from the school administrators. Any discrepancies in student information across school years were discussed and resolved with the administrators, ensuring the accuracy of the data. Among the 253 students, 52.2% were male, and 47.8% were female. In terms of race and ethnicity, 64.0% were White, 20.6% were Hispanic and 15.4% were other (8.3% Asian, 3.15% Black, 3.15% Multi-racial, 0.8% American Indian/ Alaska Native). Due to COVID-19, 92.5% of the sample received in-person instruction and 7.5% received virtual instruction; however, all students followed the same intervention procedure as detailed below.

Procedure

This research received approval from the Institutional Review Board at Worcester Polytechnic Institute. This research involved typical educational practices and did not require parental consent. Parents received an informational letter about the research and could opt their child out of the study.

In the RCT, students completed a pretest in September 2020, nine 30-minute intervention sessions (with a 2-week window to complete each), and a post-test from March to April 2021. During the pretest and post-test, students completed assessments on their algebraic knowledge and attitude towards mathematics. During each intervention session, students played DragonBox and reported their in-game progress. All students, regardless of their instructional format, completed the assessments and intervention sessions individually on a device at their own pace. For in-person students, teachers were asked to dedicate instructional periods for students to complete their study assignments. For virtual students, teachers assigned the assignments as a part of students' homework.

DragonBox

DragonBox is a commercially available, phone- or tablet-based application that allows students to interact with algebraic principles through on-screen actions. For example, students can combine the opposite numbers by dragging, overlapping and dropping one term on top of the other. DragonBox has 10 chapters, each with 20 problems. Students started from the first problem and progressed through simple (eg, arithmetic) and then complex topics (eg, factoring, distribution).

At the first intervention session, students followed the instructions provided in the study platform to open their player profile in the game, started a timer in the study platform and then played DragonBox. After 25 minutes, the timer rang; students returned to the study platform and reported their in-game progress. We asked students to report their progress because we did not have access to the log data within the game. At each subsequent session, students were prompted to report the chapter and problem at which they ended in the previous session to ensure that we obtained the data if students ran out of time in the previous session. Afterwards, they played DragonBox for 25 minutes continuing from where they left off, reported the chapter and problem at which they ended, and then reflected on the connections between DragonBox and their mathematics instruction.

Measures

In-game progress

Informed by the prior literature (eg, Chan & Santos, 2022; Hulse et al., 2019), we used the total number of problems in DragonBox that students worked on throughout the nine intervention sessions as a measure of students' in-game progress. To derive this measure, we computed the total number of completed problems based on students' final chapter and problem. For example, a student who was on Chapter 4 Problem 11 at the end of the intervention would have worked on 71 problems (ie, (4–1) chapters × 20 problems per chapter + 11 problems). The total number of problems was a focal predictor for RQ1, and its interaction with students' pretest algebra scores was a focal predictor for RQ2. Both the main effect of in-game progress and its interaction with prior mathematics achievement were the focal predictors for RQ3. The total number of problems was a focal outcome for RQ4.

For the in-person students who used the researcher-provided tablets, we reviewed the tablets after study completion and recorded students' final problem. Even though 92.5% of the students were in person, we could only extract the progress data for 89.7% of students due to various reasons (eg, broken tablet screens, students started the assignments at home and continued using their personal devices). For the remaining students, we used their self-reported progress in the analyses. The correlation between researcher-recorded progress and student-reported progress was $r(186) = 0.807$, $p < 0.001$, suggesting that students' self-reported progress was a reliable measure of their in-game progress.

Sessions completed

We also recorded the number of intervention sessions students completed out of nine sessions as a proxy for the amount of time they engaged with the intervention. We controlled for this variable to test the unique effect of students' in-game progress, independent of their time spent with the intervention, on their algebraic performance and mathematics achievement.

Algebraic knowledge assessment

We selected 10 multiple-choice items from a previously validated measure (Star et al., 2015) for our algebraic knowledge assessments (Appendix A, Table A1). The items measured students' conceptual knowledge (eg, understanding equivalence; Items 3, 6, 7, 10), procedural knowledge (eg, solving for a variable; Items 1, 5, 8) and procedural flexibility (eg, evaluating different strategies; items 2, 4, 9) in algebraic equation solving and together tapped a range of students' knowledge targeted by the intervention. The pretest and post-test items were isomorphic, with different numbers of similar magnitudes in the questions and response options. Each item was scored as correct (1) or incorrect (0); the reliability of these items was acceptable (pretest: KR-20 = 0.74; post-test: KR20 = 0.79; Decker-Woodrow et al., in press). The total score out of 10 on the post-test was included as the outcome and the pretest as a covariate or moderator for RQs 1 and 2, respectively.

State standardized assessment on mathematics

Of the 253 students, we obtained 183 students' state standardized mathematics assessment scores at the end of fifth grade (May 2019) and seventh grade (May 2021) from the district. The sample size for the state standardized assessment was smaller than 253 because some students opted out of the assessment due to parental choice and concerns related to the COVID-19 pandemic. We used the fifth-grade state scores as the measure of prior achievement because the assessment was not administered at the end of sixth grade (May 2020) due to COVID-19-related disruptions.

The assessment measured students' mathematical understanding of a broad range of topics, including ratios, equations, geometry and statistics. The assessment included a variety of tasks (eg, computation, word problems) and answer formats (eg, multiple-choice, graphing). The scores on the assessment ranged between 265 and 740. We used students' seventh-grade state standardized scores (hereafter seventh-grade state scores) as the outcome variable and the interaction between fifth-grade state scores and in-game progress as a focal predictor for RQ3.

Analytic approach

We first conducted descriptive and correlational analyses to examine the distribution of, and correlation between, each variable, as well as to inform our primary analyses. To test the effect of in-game progress on students' later algebraic knowledge (RQ1), we conducted a linear regression with students' post-test algebra score as the dependent variable, the number of DragonBox problems completed as the focal predictor and the number of intervention sessions completed, instructional format, sex, race/ethnicity and pretest score as the covariates. Next, we added the interaction between students' pretest score and their in-game progress as the focal predictor of students' post-test score to test the moderating effect of prior knowledge on the relation between students' in-game progress and later algebraic knowledge (RQ2). To test the potential effects of in-game progress on students' general mathematics achievement (RQ3), we repeated the first two models but substituted pretest and post-test scores with fifth-grade and seventh-grade state scores, respectively. Finally, to identify potential factors associated with students' in-game progress (RQ4), we conducted a linear regression with student's in-game progress as the dependent variable, sex, race/ethnicity, pretest score and fifth-grade state score as the focal predictors and the number of intervention sessions completed and instructional format as the covariates.

We noted that the analyses for RQ3 only included 183 students, whereas the remaining analyses included 253 students. To ensure that the pattern of the results could not be attributed to the differences between samples, we repeated the remaining analyses using the sample of 183 students and reported the additional findings in Appendix B (see Tables B1–

B3 and Figure B1). The pattern of the results was identical for the two samples (N = 253 vs. N = 183). Therefore, we reported the findings for RQs 1, 2 and 4 based on 253 students in the Results section.

For all the linear regression models, we considered the nesting structure of students within teachers. The intra-class correlations were above 0.07 threshold (Niehaus et al., 2014) for post-test algebra scores (0.414), seventh-grade state scores (0.573) and in-game progress (0.363); therefore, we modelled the nesting structure. This modelling approach accounts for the variations between students of different teachers, whether known or unknown, improving the accuracy of the estimates. In the multilevel models, all continuous predictors were grand mean centered (ie, sessions completed, pretest scores, fifth-grade state scores, in-game progress), and all categorical predictors were dummy coded (ie, sex, race/ethnicity, instructional format). The analyses were conducted in RStudio (RStudio Team, 2020) using the *lme4* package (Bates et al., 2015). The final dataset and the R script are available on the Open Science Framework (https://osf.io/nfg4c).

RESULTS

Preliminary analyses

Descriptive statistics of, and the correlations between, the focal variables are reported in Table 1. As indicated by the average values of in-game progress and sessions completed, many students were actively solving problems in DragonBox and engaged with the intervention. The range of the values was large for both variables, suggesting large variability in students' in-game progress and overall intervention participation.

Next, students' algebra scores at pretest and post-test as well as their state standardized scores at fifth and seventh grades were widely distributed, indicating that the sample represented students with a wide range of mathematics performance levels. We noted that students' post-test scores and seventh-grade state scores were descriptively lower than students' pretest scores and fifth-grade state scores, respectively. The decrease was

TABLE 1 Descriptive statistics of and correlations between each variable.

	In-game progress	Sessions completed	Pretest score	Post-test score	Fifth-grade state score	Seventh-grade state score
In-game progress	–					
Sessions completed	0.57***	–				
Pretest score	0.24***	0.23***	–			
Post-test score	0.30***	0.20**	0.51***	–		
Fifth-grade state score	0.17*	0.09	0.61***	0.49***	–	
Seventh-grade state score	0.33***	0.22**	0.64***	0.60***	0.79***	–
Mean	103.02	5.16	4.21	4.10	560.97	549.25
SD	49.11	3.25	2.16	2.54	57.94	55.55
Min–Max	2–200	0–9	0–10	0–10	446–725	428–725
Skewness	0.06	−0.34	0.64	0.60	0.59	0.52
Kurtosis	−1.01	−1.33	−0.05	−0.46	0.36	−0.34

Abbreviations: Max, maximum; Min, minimum; SD, standard deviation.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.

not significant for the algebraic knowledge scores, $p = 0.455$, but significant for the state standardized scores, $p < 0.001$. We discuss these findings and their implications in the Discussion section.

Correlation analyses revealed the associations between intervention progress, algebraic knowledge scores and state scores. Specifically, students' in-game progress and sessions completed were moderately correlated with each other, suggesting that these two variables might have captured related yet distinct aspects of students' participation in the intervention; therefore, it was appropriate to control for the number of sessions completed to test the independent effects of in-game progress on students' later mathematics performance. Next, students' scores on the algebraic knowledge assessment and state standardized assessment were moderately correlated with each other, warranting the analyses on both types of scores.

RQ1: In-game progress did not predict students' post-test algebra scores

A multilevel model revealed that students' in-game progress did not significantly predict their post-test scores on the algebraic knowledge assessment above and beyond the covariates, $p = 0.065$ (Table 2, Model 1.1). As expected, students' pretest scores significantly predicted their post-test scores, $p < 0.001$. For students scoring one point higher than the average on the pretest, their post-test scores were estimated to be 0.41 point higher than the post-test average. The remaining covariates did not significantly predict students' post-test scores. Together, the predictors accounted for 18% of the variances in students' post-test scores.

RQ2: The effect of in-game progress varies by students' pretest algebra scores

Building on Model 1.1, we added the Pretest × In-game Progress interaction. The interaction was significant, $p = 0.036$ and accounted for an additional 3% of the variance in students'

TABLE 2 Multilevel linear regression models predicting students' post-test algebra scores ($N = 253$).

Predictors	Model 1.1 *B* (*SE*) [95%CI]	Model 1.2 *B* (*SE*) [95%CI]
Intercept	4.21 (0.29) [3.66, 4.78]***	4.17 (0.28) [3.63, 4.71]***
Sessions completed	0.003 (0.06) [−0.12, 0.12]	−0.01 (0.06) [−0.12, 0.11]
Instructional format: virtual	0.15 (0.58) [−0.96, 1.34]	0.07 (0.58) [−1.04, 1.23]
Sex: male	−0.11 (0.27) [−0.63, 0.42]	−0.10 (0.27) [−0.62, 0.42]
Race: Hispanic	−0.04 (0.36) [−0.74, 0.65]	−0.06 (0.35) [−0.76, 0.61]
Race: Other	0.13 (0.43) [−0.69, 0.98]	0.04 (0.43) [−0.77, 0.88]
Pretest algebra scores	0.41 (0.07) [0.26, 0.58]***	0.41 (0.07) [0.27, 0.58]***
In-game progress	0.01 (0.004) [−0.0003, 0.014]	0.01 (0.004) [−0.0008, 0.013]
Pretest × in-game progress	–	0.003 (0.001) [0.0002, 0.005]*
Marginal R^2	0.18	0.21
Conditional R^2	0.32	0.32

Note: Instructional format: virtual = 1, in-person = 0; sex: male = 1, female = 0; race: Hispanic and Other are binary variables with White as the reference group.

*$p < 0.05$.

***$p < 0.001$.

post-test algebra scores. For students with higher pretest scores, more progress in DragonBox was associated with higher post-test scores. This association was not significant among students with lower pretest scores (Table 2, Model 1.2, Figure 2).

RQ3: The effects of in-game progress and prior mathematics achievement

Next, we repeated the models for RQ1 and RQ2 using students' seventh-grade state score as the outcome variable and fifth-grade score as a covariate. We found that students' in-game progress predicted their seventh-grade state scores above and beyond the covariates, $p = 0.012$. For students with average in-game progress, increasing their progress by one problem was associated with 0.17 point increase in their seventh-grade state scores (Table 3, Model 2.1). Adding the Fifth-grade Score × In-game Progress interaction revealed that the effect of in-game progress varied by students' fifth-grade state score, $p = 0.035$. The positive association between in-game progress and seventh-grade state scores was stronger among students with lower versus higher fifth-grade state scores (Table 3, Model 2.2, Figure 3).

RQ4: Factors associated with students' in-game progress

After controlling for the number of intervention sessions and instructional format, none of the predictors were significantly associated with students' in-game progress, *p*s >0.192 (Table 4). For students with an average number of sessions completed, completing one additional session was associated with 8.12 problems increase in their in-game progress, $p < 0.001$.

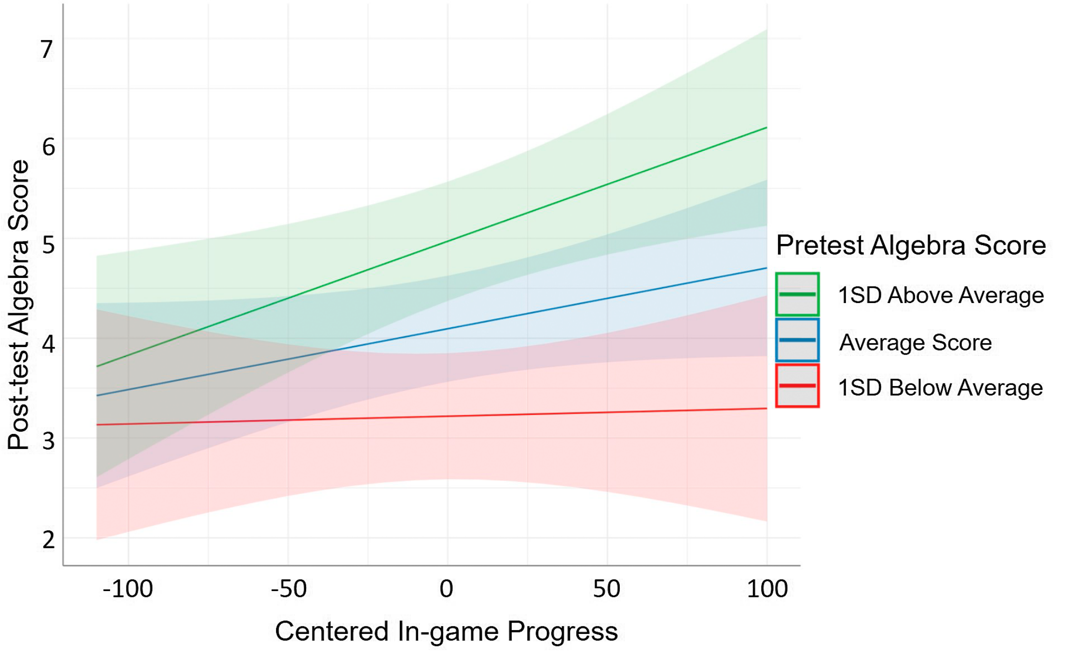

FIGURE 2 Pretest algebra score by in-game progress interaction on post-test algebra score.

TABLE 3 Multilevel linear regression models predicting students' seventh-grade state score (*N* = 183).

Predictors	Model 2.1 *B* (*SE*) [95%CI]	Model 2.2 *B* (*SE*) [95%CI]
Intercept	550.21 (5.70) [539.26, 561.60]***	550.22 (5.77) [539.17, 561.72]***
Session completed	1.64 (1.20) [−0.71, 3.96]	1.85 (1.20) [−0.51, 4.16]
Instructional format: virtual	−11.78 (10.48) [−32.09, 10.49]	−11.71 (10.36) [−31.85, 10.41]
Sex: male	2.81 (4.75) [−6.31, 12.16]	2.96 (4.69) [−6.02, 12.15]
Race: Hispanic	6.99 (5.98) [−5.18, 18.44]	8.59 (5.94) [−3.39, 19.95]
Race: Other	4.92 (7.50) [−9.47, 20.07]	6.94 (7.50) [−7.40, 21.88]
Fifth-grade scores	0.58 (0.05) [0.47, 0.71]***	0.57 (0.05) [0.47, 0.70]***
In-game progress	0.17 (0.07) [0.04, 0.30]*	0.18 (0.07) [0.05, 0.30]**
Fifth-grade score × in-game progress	–	−0.002 (0.001) [−0.003, −0.0001]*
Marginal R^2	0.52	0.52
Conditional R^2	0.66	0.67

Note: Instructional format: virtual = 1, in-person = 0; sex: male = 1, female = 0; race: Hispanic and Other are binary variables with White as the reference group.

*$p < 0.05$.

**$p < 0.01$.

***$p < 0.001$.

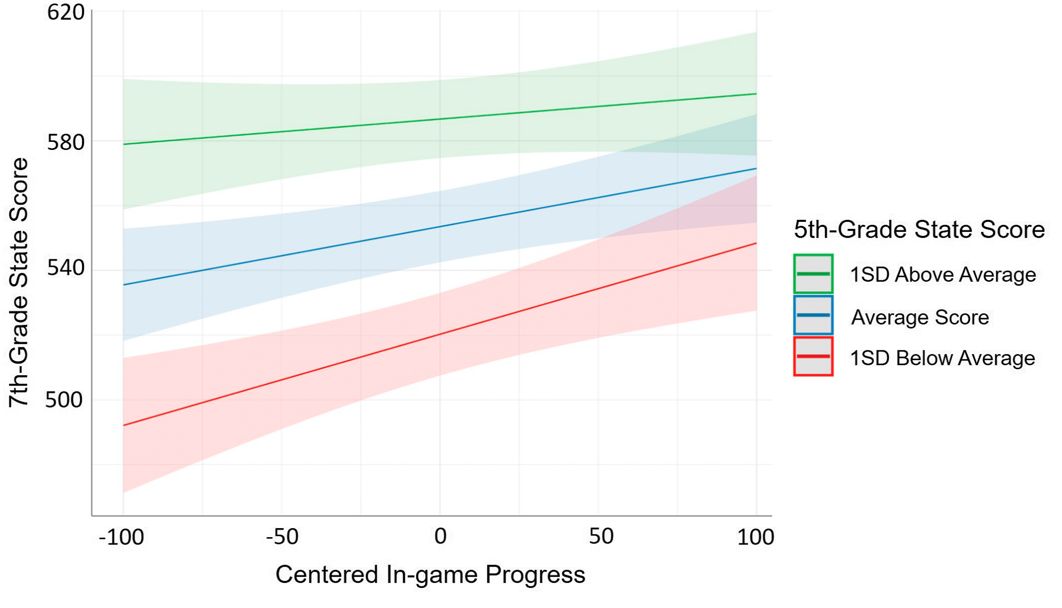

FIGURE 3 Fifth-grade state score by in-game progress interaction on seventh-grade state score.

DISCUSSION

Three main findings emerged from the current study. First, among students with higher prior knowledge, higher in-game progress was associated with higher algebraic performance after the intervention. Second, higher in-game progress was associated with higher mathematics achievement at the end of seventh grade, and this effect was stronger among students with lower versus higher prior achievement. Finally, students who completed more intervention sessions made more in-game progress; their biological sex, race/ethnicity, prior algebraic

TABLE 4 Multilevel linear regression model predicting students' in-game progress ($N = 253$).

Predictors	Model 3 *B* (*SE*) [95%CI]
Intercept	97.35 (5.68) [86.33, 108.27]***
Sessions completed	8.12 (1.07) [6.06, 10.19]***
Instructional format: virtual	12.39 (11.20) [−9.16, 34.19]
Sex: male	3.95 (5.17) [−6.14, 13.90]
Race: Hispanic	−4.24 (6.92) [−17.81, 9.05]
Race: Other	8.39 (8.03) [−7.32, 23.83]
Pretest algebra score	2.03 (1.55) [−0.97, 5.03]
Fifth-grade state scores	0.02 (0.06) [−0.10, 0.14]
Marginal R^2	0.34
Conditional R^2	0.46

Note: Instructional format: virtual = 1, in-person = 0; sex: male = 1, female = 0; race: Hispanic and Other are binary variables with White as the reference group.

***$p < 0.001$.

knowledge and prior mathematics achievement did not predict in-game progress beyond the intervention sessions.

Although the positive effects of in-game progress were promising, on average, students' performance on the algebraic assessment did not significantly improve from pretest to post-test, and their mathematics achievement scores slightly decreased from fifth grade to seventh grade. While surprising, this finding is consistent with the national US report on the decrease of mathematical performance during the COVID-19 pandemic (Engzell et al., 2021; Kuhfeld et al., 2022). The long-term educational impacts of the pandemic have yet to be seen, but the current study provided a valuable glimpse of its potential impact on algebraic learning. We interpret the findings with caution and in the context of the pandemic disruptions.

Algebraic performance

Contrary to our hypothesis and prior research (Chan & Santos, 2022; Hulse et al., 2019), we did not find an overall positive effect of in-game progress on students' algebraic performance. However, we observed an interaction such that the positive effect of in-game progress was only significant among students with higher pretest algebra scores. While beyond the scope of the current study on in-game progress, we found that, in the larger RCT, students who received either DragonBox or another interactive algebraic game notably improved on all four conceptual knowledge items from pretest ($M = 46\%$) to post-test ($M = 53\%$), whereas their improvements over time were inconsistent on other items (Chan et al., in press). Given the seemingly stagnant performance on the overall algebraic assessment from pretest to post-test, it is exciting to observe student improvement in conceptual knowledge, especially during the COVID-19 pandemic. Together, the findings from the RCT corroborate with prior research reporting the potential benefits of playing games like DragonBox on students' conceptual knowledge of algebra, more specifically understanding of mathematical equivalence (Chan, Lee, et al., 2022; Supriadi et al., 2020).

The current finding extends the literature on in-game progress (Chan & Santos, 2022; Hulse et al., 2019) and DragonBox (Liu et al., 2015; Siew et al., 2016; Supriadi et al., 2020) in two important ways. First, it provides additional evidence supporting the notion that in-game progress may be positively related to some students' learning outcomes. Specifically, the active experience of balancing two sides of the screen while isolating the dragon box may help students develop a perceptual-motor routine of algebraic equation solving, improving their understanding of algebra. Although this potential explanation seems counterintuitive, it is supported by extensive theoretical and empirical work on the perceptual and embodied nature of learning (eg, Abrahamson et al., 2020; Alibali & Nathan, 2012; Goldstone et al., 2017; Kellman et al., 2010; Supriadi et al., 2020). For example, performing actions by using one's body to form triangles, even without pedagogical language to connect actions with mathematical concepts, helps undergraduate students gain insights into geometric proofs (Nathan et al., 2014). Furthermore, the active experience of manipulating and transforming algebraic expressions prior to a formal lesson on mathematical principles, such as the commutative property, improves seventh graders' performance on simplifying algebraic expressions (Ottmar & Landy, 2017).

When asked about the connections between DragonBox and mathematics in the current study, students' anecdotal report after their gameplay suggested that they recognized some of these connections as they described their actions in mathematical terms (eg, fractions, equations, multiply, divide, etc.). For example, one student recognized that they could multiply a term in DragonBox by dragging a number to the term. They further explained that '… just like in math when you do equations you [*sic*] [,] whatever you do to one side you do to the other, which is the case with dragonBox [*sic*]. When I dragged one icon to one side [,] I would have to do the same to the other side. Another connection I could make was that the rules for cancelling out numbers are the same'. One student made a similar comment about balancing equations by saying that, '…when you add something to one side like a[n] animal, you have to also apply it to the other side as well'. Students also connected the action of combining the same pictures at the top and bottom of a fraction in DragonBox to combining like terms in mathematics. The seemingly frivolous procedure of moving pictures and boxes, therefore, may help students discover and practice algebraic concepts, such as equivalence, just as procedural and conceptual knowledge iteratively support each other throughout mathematical development (Rittle-Johnson, 2017). However, more qualitative research is needed to test this hypothesis and explore the potential mechanisms through which DragonBox supports algebraic learning.

Second, the finding suggests a nuanced dosage effect in that the students with higher prior knowledge may benefit more from solving more problems in the game. A potential explanation for this select benefit of in-game progress is that students with higher, as opposed to lower, foundational knowledge of algebra may be better equipped to make connections between the mathematical ideas hidden within DragonBox and algebraic concepts taught in classrooms, and in turn benefit more from DragonBox as they progress through the game. This explanation aligns with some prior work demonstrating that students with higher versus lower prior knowledge benefit more from mathematics interventions (Swanson et al., 2008; Wood et al., 2020) and that students with higher prior knowledge may be more likely to notice mathematical structures that support learning (Novick, 1988) and equation solving (Chan, Ottmar, et al., 2022; Lee et al., 2022). The moderating effect of students' prior algebraic knowledge, therefore, highlights its importance in the relation between in-game progress and later algebraic performance. To maximize the potential benefits of playing DragonBox, it may be helpful for students to first establish foundational algebraic knowledge through instruction and then engage with problem solving through game play.

Mathematics achievement

Given the prior focus on students' algebraic performance (Dolonen & Kluge, 2015; Liu et al., 2015; Long & Aleven, 2017; Siew et al., 2016; Supriadi et al., 2020), the current study is the first to report DragonBox's impacts on general mathematics achievement. We found that on average, students' state standardized scores decreased over time, potentially owing to the negative impacts associated with educational disruptions during the COVID-19 pandemic (Engzell et al., 2021; Kuhfeld et al., 2022). Despite the learning loss, we found a positive effect of in-game progress, suggesting that solving more problems in DragonBox may have mitigated the decrease in students' mathematics achievement. Furthermore, this positive effect was stronger among students with lower mathematics achievement.

One potential explanation of this finding is that simply engaging students in mathematics through playing DragonBox may help foster a positive attitude towards mathematics and in turn support their later performance, especially among students with lower mathematics achievement. Prior work has shown that the game-based approach can increase students' engagement with learning (Mora et al., 2015) and that playing DragonBox fosters students' positive attitudes towards mathematics (Dolonen & Kluge, 2015; Long & Aleven, 2017; Siew et al., 2016). Focus groups with teachers in the RCT revealed that students were excited and happy when they noticed the connections between DragonBox and classroom instruction, suggesting that the game may foster students' positive attitudes towards learning mathematics (Miyaoka, et al., in press). Future work should replicate the current finding and delineate the mechanisms through which in-game progress and prior knowledge influence students' mathematics attitudes and performance. Doing so will help identify ways that game-based approaches support different aspects of mathematical learning and inform practices that maximize their benefits across students at varying levels of performance.

Factors associated with in-game progress

To identify potential factors related to students' in-game progress, we tested whether students' biological sex, race/ethnicity, pretest algebraic knowledge and fifth-grade mathematics achievement were associated with students' in-game progress. These factors were selected based on prior research reporting their associations with mathematical learning and engagement (Dong et al., 2020; Martinez & Guzman, 2013; Nosek & Smyth, 2011; Pecore et al., 2017; Reinhold et al., 2020). Contrary to prior work, we found that none of these factors were significantly associated with students' in-game progress beyond the number of intervention sessions students completed, implicating that the effects of these factors may be inconsistent or situational (Linn & Hyde, 1989; Sonnenschein & Galindo, 2015). We did find that students who completed more intervention sessions made more in-game progress. A potential implication is that, regardless of students' sex, race/ethnicity or prior knowledge or achievement, dedicating time and encouraging students to solve problems in DragonBox may help them make progress and in turn support mathematical learning.

Limitations and future directions

The current study had several limitations. First, our definitions and measures of the constructs by no means represented all perspectives in the field, and the study only focused on the effects of in-game progress within DragonBox, limiting the generalizability of the current findings. However, our definitions and measures were grounded in existing work (eg, Chan & Santos, 2022; Hulse et al., 2019; Star et al., 2015), allowing us to connect with the larger literature on educational technologies, algebraic knowledge and mathematics achievement. Furthermore, the current results contributed novel findings on the effects of in-game progress, advancing our understanding of how DragonBox supports mathematical learning.

Second, the current sample was a subset of students from an RCT conducted in one school district; therefore, the study did not capture a wide range of demographic or socioeconomic diversity, limiting the generalizability of the results. However, the large range of the scores on the pretest and fifth-grade mathematics achievement indicated that the sample consisted of students at varying performance levels. Furthermore, the current sample was large compared with some prior work (eg, Long & Aleven, 2017; Siew et al., 2016; Supriadi et al., 2020); the study was conducted in an authentic classroom context over an academic school year; and the pattern of results was robust across two subsamples (Appendix B). Future studies should replicate the current findings with participants from a wide range of backgrounds to ensure the generalizability of the results.

Third, the study was conducted during the pandemic and many known or unknown factors might have impacted our findings. Despite these challenges, we aimed to contribute to the cumulative understanding of learning by explicitly articulating our study contexts, grounding our work in the literature and triangulating our inferences using available data (Grosz et al., 2020). For example, while we were not able to collect additional qualitative data or account for all possible factors given the already stressful situation, we quantitatively accounted for students' instructional format and the nesting structure of the data. This approach succinctly accounted for a salient feature of the instruction during the pandemic and the potential systematic variations between students. Furthermore, we discussed some qualitative anecdotes from the larger project to supplement quantitative results and provide additional information on the study context. Although we were unable to capture a comprehensive picture of student learning during the pandemic, our findings nonetheless provide a piece of the puzzle on student learning during this difficult time.

CONCLUSION

We examined the associations between students' DragonBox progress and two mathematics outcomes. We found that these associations varied depending on students' prior performance. If students' later mathematics performance is related to their engagement with DragonBox and its interaction with students' prior performance, parents and educators may dedicate varying amounts of time for students to play the game depending on their prior performance. Doing so may maximize the efficacy of the game to efficiently support students with varying levels of mathematical knowledge and achievement.

ACKNOWLEDGEMENTS

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through an Efficacy and Replication Grant (R305A180401) to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the US Department of Education. We thank the participating teachers and students for their support with this study. We also thank Barbara Booker, Craig Mason, Shihfen Tu, Ji-Eun Lee and members of the Math Abstraction Play Learning and Embodiment Lab for their work on this project.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The final dataset for the current study is available on the Open Science Framework (https://osf.io/nfg4c).

ETHICS APPROVAL STATEMENT

This research received approval from the Institutional Review Board at Worcester Polytechnic Institute.

ORCID

Jenny Yun-Chen Chan https://orcid.org/0000-0003-1479-5935
Allison S. Liu https://orcid.org/0000-0003-1075-2575
Justin Roberts https://orcid.org/0000-0001-5906-6336
Erin Ottmar https://orcid.org/0000-0002-9487-7967

REFERENCES

Abrahamson, D., Nathan, M. J., Williams-Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education*, *5*, 147.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers and Education*, *83*, 57–63. https://doi.org/10.1016/j.compedu.2014.12.012

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Cayton-Hodges, G. A., Feng, G., & Pan, X. (2015). Tablet-based math assessment: What can we learn from math apps? *Educational Technology & Society*, *18*(2), 3–20.

Chan, J. Y.-C., Closser, A. H., Ngo, V., Smith, H., Liu, A., & Ottmar, E. (in press). Examining shifts in conceptual knowledge, procedural knowledge, and procedural flexibility in the context of two game-based technologies.

Chan, J. Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2022). From here to there! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology*, *114*(1), 56–71. https://doi.org/10.1037/edu0000596

Chan, J. Y.-C., Ottmar, E. R., Smith, H., & Closser, A. H. (2022). Variables versus numbers: Effects of symbols and algebraic knowledge on students' problem-solving strategies. *Contemporary Educational Psychology*, *71*, 102114. https://doi.org/10.1016/j.cedpsych.2022.102114

Chan, J. Y.-C., & Santos, F. (2022, June). How does Calcularis influence the development of numerical skills in children with Dyscalculia? Analyses of children's interaction within the game. In C. Gilmore & L. Cragg (Co-chairs), *Using large-scale educational app data to investigate mathematical learning processes* [Symposium]. The 2022 Mathematical Cognition and Learning Society Conference, Belgium.

Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields*. Statistical analysis report. NCES 2014-001.

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

Decker-Woodrow, L., Mason, C. A., Lee, J. E., Chan, J. Y.-C., Sales, A., Liu, A., & Tu, S. (in press). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA Open*.

Dolonen, J., & Kluge, A. (2015). Algebra learning through digital gaming in school. In *Proceedings of the 11th meeting of the international conference on computer supported collaborative learning* (Vol. *1*, pp. 252–259). International Society of the Learning Sciences.

Dong, A., Jong, M. S. Y., & King, R. B. (2020). How does prior knowledge influence learning engagement? The mediating roles of cognitive load and help-seeking. *Frontiers in Psychology*, *11*, 591203. https://doi.org/10.3389/fpsyg.2020.591203

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *118*(17), e2022376118. https://doi.org/10.1073/pnas.2022376118

Gee, J. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, *1*(1), 20. https://doi.org/10.1145/950566.950595

Goldstone, R. L., Marghetis, T., Weitnauer, E., Ottmar, E. R., & Landy, D. (2017). Adapting perception, action, and technology for mathematical reasoning. *Current Directions in Psychological Science*, *26*(5), 434–441. https://doi.org/10.1177/0963721417704888

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255.

Hulse, T., Daigle, M., Manzo, D., Braith, L., Harrison, A., & Ottmar, E. R. (2019). From here to there! Elementary: A game-based approach to developing number sense and early algebraic understanding. *Education Tech Research Development*, *67*, 423–441.

Kahoot! (2019). *Kahoot! and DragonBox join forces to create an awesome math learning experience for all*. https://kahoot.com/press/2019/05/09/kahoot-dragonbox-join-forces-create-awesome-math-learning-experience-for-all/

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay? *Computers & Education*, *51*(4), 1609–1620. https://doi.org/10.1016/j.compedu.2008.03.003

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, *2*(2), 285–305. https://doi.org/10.1111/j.1756-8765.2009.01053.x

Kieran, C. (2006). Research on the learning and teaching of algebra. In Á. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education* (pp. 11–49). Sense. https://doi.org/10.1163/9789087901127

Kuhfeld, M., Soland, J., & Lewis, K. (2022). *Test score patterns across three COVID-19-impacted school years* (EdWorkingPaper: 22-521). https://doi.org/10.26300/ga82-6v47

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

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher*, *18*(8), 17–27. https://doi.org/10.2307/1176462

Liu, Y.-E., Ballweber, C., O'rourke, E., Butler, E., Thummaphan, P., & Popović, Z. (2015). Large-scale educational campaigns. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1–24. https://doi.org/10.1145/2699760

Long, Y., & Aleven, V. (2017). Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *ACM Transactions on Computer-Human Interaction*, *24*(3), 1–27. https://doi.org/10.1145/3057889

Martinez, S., & Guzman, S. (2013). Gender and racial/ethnic differences in self-reported levels of engagement in high school math and science courses. *Hispanic Journal of Behavioral Sciences*, *35*(3), 407–427.

Miyaoka, A., Decker-Woodrow, L., Hartman, N., Booker, B., & Ottmar, E. (in press). Emergent coding and topic modeling: a comparison of two qualitative analysis methods on teacher focus group data.

Mora, A., Riera, D., Gonzalez, C., & Arnedo-Moreno, J. (2015). A literature review of gamification design frameworks. *2015 7th International Conference on Games and Virtual Worlds for Serious Applications. VS-Games*, 1–8. IEEE.

Murphy, R., Roschelle, J., Feng, M., & Mason, C. A. (2020). Investigating efficacy, moderators and mediators for an online mathematics homework intervention. *Journal of Research on Educational Effectiveness*, *13*(2), 235–270. https://doi.org/10.1080/19345747.2019.1710885

Nathan, M. J., Walkington, C., Boncoddo, R., Pier, E., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof. *Learning and Instruction*, *33*, 182–193.

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

Niehaus, E., Campbell, C. M., & Inkelas, K. K. (2014). HLM behind the curtain: Unveiling decisions behind the use and interpretation of HLM in higher education research. *Research in Higher Education*, *55*(1), 101–122. https://doi.org/10.1007/s11162-013-9306-7

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125–1156.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520.

Ottmar, E., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on learning. *Journal of the Learning Sciences*, *26*(1), 51–78. https://doi.org/10.1080/10508406.2016.1250212

Pecore, J. L., Kirchgessner, M. L., Demetrikopoulos, M. K., Carruth, L. L., & Frantz, K. J. (2017). Formal lessons improve informal educational experiences: The influence of prior knowledge on student engagement. *Visitor Studies*, *20*(1), 89–104.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Reinhold, F., Strohmaier, A., Hoch, S., Reiss, K., Böheim, R., & Seidel, T. (2020). Process data from electronic textbooks indicate students' classroom engagement. *Learning and Individual Differences*, *83*, 101934.

Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives*, *11*(3), 184–190. https://doi.org/10.1111/cdep.12229

Rowe, J. P., Lobene, E. V., Mott, B. W., & Lester, J. C. (2017). Play in the museum: Design and development of a game-based learning exhibit for informal science education. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, *9*(3), 96–113.

RStudio Team. (2020). *RStudio: Integrated development for R. RStudio*. PBC. http://www.rstudio.com/

Rutherford, T., Duck, K., Rosenberg, J. M., & Patt, R. (2021). Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic. *Journal of Research on Technology in Education*, *54*, S94–S131. https://doi.org/10.1080/15391523.2021.1920520

Samur, Y., & Evans, M. (2012, April 13). The effects of serious games on performance and engagement: A review of the literature (2001-2011). *American Educational Research Association Conference*.

Siew, N. M., Geofrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes toward algebra: The effects of game-based learning using Dragonbox 12+ app. *The Research Journal of Mathematics and Technology*, *5*(1), 66–79.

Sonnenschein, S., & Galindo, C. (2015). Race/ethnicity and early mathematics skills: Relations between home, classroom, and mathematics achievement. *Journal of Educational Research*, *108*(4), 261–277. https://doi.org/10.1080/00220671.2014.880394

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

Sun-Lin, H. Z., & Chiou, G. F. (2017). Effects of comparison and game-challenge on sixth graders' algebra variable learning achievement, learning attitude, and meta-cognitive awareness. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(6), 2627–2644.

Supriadi, N., Ramadhona, K., Pratiwi, D. D., & Widyawati, S. (2020). Concept understanding skills and mathematical problem-solving skills in algebraic materials: The effect of DMR learning model assisted by DragonBox puzzle game during the Covid-19 pandemic. *Jurnal Pendidikan Matematika*, *11*(1), 191–198.

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *100*(2), 343–379. https://doi.org/10.1037/0022-0663.100.2.343

Torres, R., Toups, Z. O., Wiburg, K., Chamberlin, B., Gomez, C., & Ozer, M. A. (2016). Initial design implications for early algebra games. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 325–333. https://doi.org/10.1145/2968120.2987748

Wood, T., Mazzocco, M. M. M., Calhoon, M. B., Crowe, E. C., & Connor, C. M. (2020). The effect of peer-assisted mathematics learning opportunities in first grade classrooms: What works for whom? *Journal of Research on Educational Effectiveness*, *13*(4), 601–624. https://doi.org/10.1080/19345747.2020.1772422

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.

How to cite this article: Chan, J. Y.-C., Byrne, C., Jerusal, J., Liu, A. S., Roberts, J., & Ottmar, E. (2023). Keep DRAGging ON: Is solving more problems in DragonBox 12+ associated with higher mathematical performance during the COVID-19 pandemic?. *British Journal of Educational Technology*, *00*, 1–24. https://doi.org/10.1111/bjet.13304

APPENDIX A

TABLE A1 The items in the algebraic knowledge assessment at post-test (Star et al., 2015).

Item	Type	Question	Answer
1	Procedural	Solve the equation for y. $3\ (y-2) = -1\ (y-2)+2$ a. 10 b. −2 c. 10/4 d. 4	c
2	Flexibility	Lynn solved the problem: $1/5\ (x+9) = 6$ Lynn's first step was $(5)\ 1/5\ (x+9) = 6\ (5)$ $x+9 = 30$ What step did Lynn use to get from the first line to the second line? a. combine like terms b. add or subtract the same quantity on both sides c. distribute across parentheses d. multiply or divide the same quantity on both sides	d
3	Conceptual	$6+2 = 8$ ↑ What does this symbol mean? a. two quantities on either side have the same value b. the total c. the problem has been solved d. what the answer is	a
4	Flexibility	Imagine you are taking a timed test. You want to use fast (and correct) ways to solve the problems so you can finish as many as possible. Which would be the best way to start the problem $6\ (y+3) = 19$? a. Alice's ‘multiply by 6 on both sides’ way b. Bob's ‘divide by 6 on both sides first’ way c. Channing's ‘subtract 19 from both sides first’ way d. David's ‘distribute first’ way	d
5	Procedural	Solve the equation below for x. $1 = (6-5x)/2$ a. 4/5 b. 0 c. 1 d. −4/5	a
6	Conceptual	If $12x+18 = 19$, which of the following must also be true? a. $2x+3 = 19$ b. $12x+18-18 = 19-18$ c. $12x-12+18-12 = 19$ d. $-12x-18 = 19$	b
7	Conceptual	Which of the following is equivalent to (the same as) $(a+2)+(a+2)+(a+2)+(a+2)+(a+2)$? a. $a+10$ b. $5a+2$ c. $5\ (a+2)$ d. a^5+10	c
8	Procedural	Solve the equation below for n. $10n+5 = 12n+17-6n$ a. 0 b. 3 c. 2 d. 3/2	b

(Continues)

TABLE A1 (Continued)

Item	Type	Question	Answer
9	Flexibility	Imagine you are taking a timed test. You want to use fast (and correct) ways to solve the problems so you can finish as many as possible. Which would be the best way to start the problem? (Choose the letter for the best way to start) 1/5 (3x + 2) = 10 a. Abby's 'divide by 5 on both sides first' way b. Blake's 'distribute first' way c. Chen's 'subtract 10 from both sides first' way d. Dean's 'multiply by 5 on both sides first' way	d
10	Conceptual	Which of the following is NOT equivalent to 18 (75 − 16)? a. 18 (59) b. 18 (−16 + 75) c. 18 (75) − 16 d. 18 (75) − 18 (16)	c

APPENDIX B

Because the analyses testing the effects of in-game progress and fifth-grade state scores on students' seventh-grade state scores only included 183 students (RQ3), we repeated the analyses for RQs 1, 2 and 4 with these same 183 students. By doing so, we (a) checked the robustness of the results across the two overlapping yet different samples (253 vs. 183 students) and (b) examined whether the pattern of results varied by the analytic samples. We found that the pattern of the results with 183 students was identical to that with 253 students reported in the manuscript. Below, we reported the descriptive statistics, correlations and model estimates for the RQs 1, 2 and 4 with 183 students who had both fifth-grade and seventh-grade state scores and, thus, were included in the analyses for RQ3 in the manuscript.

Descriptive and correlation analyses

TABLE B1 Descriptive statistics of and correlations between each variable (N = 183).

	In-game progress	Sessions completed	Pretest score	Post-test score	Fifth-grade state score	Seventh-grade state score
In-game progress	–					
Sessions completed	0.58***	–				
Pretest score	0.32***	0.29***	–			
Post-test score	0.37***	0.28**	0.52***	–		
Fifth-grade state score	0.19**	0.13	0.59***	0.47***	–	
Seventh-grade state score	0.32***	0.21**	0.65***	0.59***	0.79***	–
Mean	104.19	5.27	4.23	3.97	560.33	550.75
SD	48.39	3.32	2.14	2.53	58.25	55.38
Min–Max	13–200	0–9	0–10	0–10	446–725	428–725
Skewness	0.13	−0.37	0.62	0.67	0.58	0.54
Kurtosis	−1.03	−1.34	−0.07	−0.44	0.38	−0.27

Abbreviations: Max, maximum; Min, minimum; SD, standard deviation.

**p < 0.01.

***p < 0.001.

RQs 1 and 2: Effects of in-game progress and pretest algebra scores on post-test algebra scores

TABLE B2 Multilevel linear regression models predicting students' post-test algebra scores ($N = 183$).

Predictors	Model 1.1 *B* (*SE*) [95%CI]	Model 1.2 *B* (*SE*) [95%CI]
Intercept	3.93 (0.32) [3.32, 4.55]***	3.87 (0.29) [3.31, 4.42]***
Sessions completed	0.06 (0.07) [−0.08, 0.20]	0.05 (0.07) [−0.08, 0.17]
Instructional format: virtual	0.69 (0.67) [−0.59, 2.11]	0.61 (0.66) [−0.63, 1.94]
Sex: male	−0.22 (0.31) [−0.81, 0.37]	−0.24 (0.30) [−0.82, 0.34]
Race: Hispanic	0.10 (0.38) [−0.63, 0.82]	0.04 (0.37) [−0.68, 0.74]
Race: Other	0.37 (0.47) [−0.53, 1.30]	0.18 (0.47) [−0.71, 1.08]
Pretest scores	0.40 (0.08) [0.23, 0.58]***	0.39 (0.08) [0.23, 0.57]***
In-game progress	0.01 (0.004) [−0.001, 0.016]	0.007 (0.004) [−0.001, 0.015]
Pretest × in-game progress	–	0.004 (0.001) [0.002, 0.007]**
Marginal R^2	0.24	0.32
Conditional R^2	0.38	0.39

Note: Instructional format: virtual = 1, in-person = 0; sex: male = 1, female = 0; race: Hispanic and Other are binary variables with White as the reference group.

**$p < 0.01$.

***$p < 0.001$.

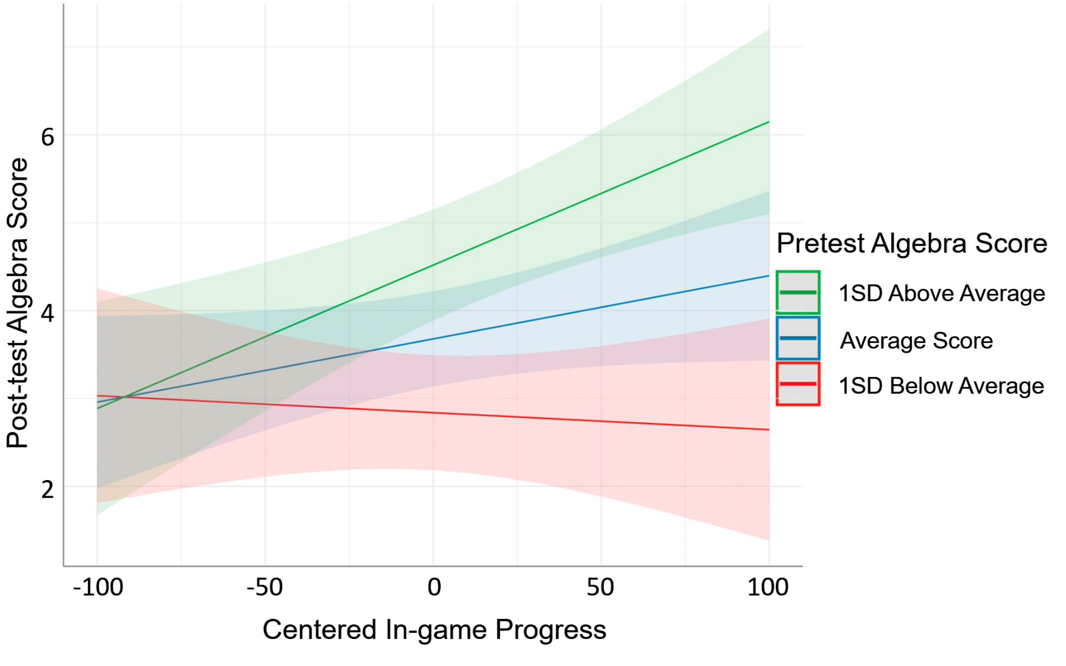

FIGURE B1 Pretest algebra score by in-game progress interaction on post-test algebra score ($N = 183$).

RQ4: Predictors of students' in-game progress

TABLE B3 Multilevel linear regression model predicting students' in-game progress (N = 183).

Predictors	Model 3 *B* (*SE*) [95%CI]
Intercept	100.37 (6.30) [88.21, 112.47]***
Sessions completed	8.35 (1.19) [6.07, 10.63]***
Instructional format: virtual	5.14 (11.84) [−17.59, 28.51]
Sex: male	4.36 (5.37) [−6.02, 14.76]
Race: Hispanic	−6.10 (6.75) [−19.27, 6.87]
Race: Other	7.60 (8.45) [−8.75, 23.84]
Pretest algebra score	2.49 (1.65) [−0.68, 5.69]
Fifth-grade state scores	0.003 (0.06) [−0.12, 0.12]
Marginal R^2	0.36
Conditional R^2	0.54

Note: Instructional format: virtual = 1, in-person = 0; sex: male = 1, female = 0; race: Hispanic and Other are binary variables with White as the reference group.

***$p < 0.001$.