

Cost-Effectiveness of Algebraic Technological Applications

Matthew Finster, Lauren Decker-Woodrow, Barbara Booker, Craig A. Mason, Shihfen Tu & Ji-Eun Lee

To cite this article: Matthew Finster, Lauren Decker-Woodrow, Barbara Booker, Craig A. Mason, Shihfen Tu & Ji-Eun Lee (23 Oct 2023): Cost-Effectiveness of Algebraic Technological Applications, Journal of Research on Educational Effectiveness, DOI: [10.1080/19345747.2023.2269918](https://doi.org/10.1080/19345747.2023.2269918)

To link to this article: <https://doi.org/10.1080/19345747.2023.2269918>



© 2023 Westat. Published with license by Taylor & Francis Group, LLC



Published online: 23 Oct 2023.



Submit your article to this journal [↗](#)



Article views: 243



View related articles [↗](#)



View Crossmark data [↗](#)

Cost-Effectiveness of Algebraic Technological Applications

Matthew Finster^a, Lauren Decker-Woodrow^a, Barbara Booker^b, Craig A. Mason^c,
Shihfen Tu^d and Ji-Eun Lee^e

^aPrincipal Research Associate, Westat, Rockville, Maryland, USA; ^bIndependent Consultant, Cumming, Georgia, USA; ^cProfessor of Education and Applied Quantitative Methods, University of Maine, Orono, Maine, USA; ^dDirector, School of Learning and Teaching, Professor of Education and Applied Quantitative Methods, University of Maine, Orono, Maine, USA; ^eResearch Scientist, Learning Sciences & Technologies, Worcester Polytechnic Institute, Worcester, Massachusetts, USA

ABSTRACT

COVID-19 contributed to the largest student performance decline in mathematics since 1990. The nation needs cost-effective mathematic interventions to address this drop and improve students' mathematics performance. This study presents a cost-effectiveness analysis (CEA) of three algebraic technological applications, across four conditions: *From Here to There (FH2T)*, *Dragon Box 12+* (DragonBox), *Immediate Feedback and Active Control*. This CEA study uses impact measures from a student-level randomized control trial comparing student learning from the three treatment conditions to the Active Control condition with an analytic sample of 1,850 middle school students across 9 schools, 34 teachers, and 127 classes. The results from the CEA indicate FH2T costs \$39 per student and produces an average effect of 0.135 on algebraic achievement resulting in a cost-effectiveness ratio of \$291. DragonBox costs \$55 per student and produces an average effect of 0.269 on algebraic achievement resulting in a cost-effectiveness ratio of \$206. Overall, the current CEA study demonstrates the efficiency of FH2T and DragonBox as low-cost interventions for improving students' algebraic performance and addressing the nation's decline in mathematics.

ARTICLE HISTORY

Received 18 April 2022
Revised 10 July 2023
Accepted 20 August 2023

KEYWORDS

Randomized controlled trial; cost-effectiveness; cost analysis; ingredients method; algebra; game-based; *From Here to There*; *DragonBox*; ASSISTments

Performance in mathematics dropped substantially in the United States during COVID-19. But even before COVID-19, U.S. students underperformed in math compared to students in other countries (e.g., Provasnik et al., 2016; Schleicher, 2018), and many middle and high school students struggled to understand basic algebraic concepts (Kena et al., 2015; Kieran, 2006). This underperformance is particularly concerning because algebra is a critical foundation for learning advanced mathematics (National Mathematics Advisory Panel [NMAP], 2008). To efficiently address this drop and underperformance in mathematics, decision-makers need information on the cost-effectiveness of algebraic interventions.

CONTACT Matthew Finster  matthewfinster@westat.com  Principal Research Associate, 1600 Research Boulevard, Rockville, MD, USA.

© 2023 Westat. Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Therefore, this current study reports on the cost-effectiveness of three educational technology interventions on algebraic understanding among seventh graders across four conditions: (a) From Here to There (FH2T), (b) DragonBox 12+ (DragonBox), (c) Immediate Feedback, and (d) Active Control. The FH2T and DragonBox conditions represent the use of game-based applications. Immediate Feedback entails problem sets using an online homework system, ASSISTments. The Active Control condition mimics traditional homework assignments while still using technology. With the promise of such technological interventions to improve algebraic understanding, it is important to identify which may be the most cost-effectiveness. As one of the few cost-effectiveness studies of algebraic technological applications using impact estimates from a rigorous large-scale randomized controlled trial study, the article contributes to the literature about the efficiency of incorporating game-based technologies into instruction and is crucial for decision-makers aiming to increase the productivity of algebraic interventions to address learning loss stemming from disruptions of the COVID-19 pandemic.

Background

COVID-19 contributed to the largest student performance decline in mathematics in the United States since 1990 (NCES – NAEP, 2022). According to the latest results from the NAEP, also known as The Nation’s Report Card, released by the U.S. Department of Education’s NCES, the national average declines in mathematics scores for fourth- and eighth-graders were the largest ever recorded in that subject. Specifically, “The average eighth-grade mathematics score decreased by 8 points compared to 2019 and was lower than all previous assessment years going back to 2003.” (NCES – NAEP, 2022). The release of 2022 NAEP scores was followed by calls to identify and implement effective interventions in mathematics (and English Language Arts) to accelerate learning and address gaps in student achievement exacerbated by COVID-19 (e.g. Schneider, 2022).

Even before the large drop in student performance associated with COVID-19, there was cause for concern with U.S. students’ performance in mathematics. In particular, many middle and high school students fail to understand basic algebraic principles, such as which transformations are legal and appropriate (Marquis, 1998; National Mathematics Advisory Panel [NMAP], 2008), and how to convert between formal symbolic expressions and other representations (Koedinger & Nathan, 2004). These notational struggles make it almost impossible to understand advanced concepts, which often assume an understanding of algebraic notation. Given the implications on students’ performance, college graduation rates, and employment earnings (National Mathematics Advisory Panel [NMAP], 2008), the National Mathematics Advisory Panel [NMAP] (2008) highlighted algebra as an area of special concern.

While technology-based drill-and-practice and tutorials have been a common approach to supplemental instruction in algebra (National Mathematics Advisory Panel [NMAP], 2008), game-based applications that support algebraic instruction by teaching math through discovery-based puzzles may be more effective (Chan et al., 2022; Ottmar et al., 2012, 2015; Ottmar & Landy, 2017). Based on this research and prior evidence of game-based applications, a large-scale randomized control trial (RCT) was conducted to

evaluate the effectiveness of three educational technology interventions on algebraic understanding among seventh-graders, across four conditions: (a) From Here to There (FH2T), (b) Dragon Box 12+ (hereafter, DragonBox), (c) Immediate Feedback, and (d) Active Control (Decker-Woodrow et al., 2023).

However, relying solely on research about interventions' impacts without accounting for costs essentially promotes interventions with the largest effects irrespective of resources (see Levin et al., 1987). Hence, education researchers have argued that one must evaluate both the costs and the effects when considering educational interventions (e.g., Harris, 2009; Levin, 2001; Levin & Belfield, 2015; Levin & McEwan, 2001). Information about the costs and effects of an education intervention can assist decision-makers in assessing the productivity of interventions. That is, whether implementing a new intervention or other approaches may yield better results given the costs. And, when interventions are assessed on similar outcomes, they can be compared on their rates of efficiency (e.g., Yeh, 2010a), which provides decision-makers with information about the cost-effectiveness of a range of alternatives in achieving similar aims. By selecting more cost-effective or efficient interventions, education decision-makers could improve the productivity of education.

The Need for CEA of Technology-Based Algebraic Interventions

Given limited resources, education decision-makers are always under pressure to accomplish more with the same or even fewer resources (Levin et al., 2018; Levin & Belfield, 2015). Hence, there is a constant need to identify and implement relatively more cost-effective interventions to improve student performance. Also, given the substantial amount of public funding in education in the United States—over one trillion dollars (U.S. Government, 2012)—and the recent investment of the Coronavirus Aid, Relief, and Economic Security Act that provided funding to local education agencies through the Elementary and Secondary School Emergency Relief Fund to address the impact of COVID-19 on elementary and secondary schools, one may expect considerable information and scrutiny on how these resources can be used more efficiently. According to Levin and Belfield (2015), Cost-effectiveness analysis (CEA)—an approach that identifies which strategies will maximize outcomes for any given cost or produce a given outcome for the lowest cost—is the most versatile tool for this task.

Despite the growing investigation into the cost-effectiveness and cost-benefit of educational interventions in the last decade (e.g., Hollands et al., 2014; 2016; Levin et al., 2018; Yeh, 2010a, 2010b), there has been limited information about the cost-effectiveness of specific education interventions to inform decision-makers when they are considering multiple alternatives. Recent federal efforts and funding have increased emphasis on cost analysis, for example making it a requirement in NCER-funded research in 2018 (Schneider, 2020), and spurring additional cost analysis projects and resources, such as the Cost Analysis: A Starter Kit (Institute of Education Sciences, Institute of Education Sciences, 2020), the Cost Analysis in Practice (CAP) project, the Cost Analysis Standards Project Panel, and the corresponding report Standards for the Economic Evaluation of Educational and Social Programs (Cost Analysis Standards Project, 2021). State-level efforts have also worked to provide cost-benefit information

to policymakers. For example, the Washington State Institute of Public Policy (WSIPP) conducts and catalogs cost-benefit analysis (CBA) results of K-12 education programs to provide state policymakers with information that can lead to more efficient use of taxpayer dollars.

Despite these federal and state efforts, there is still limited information about the cost-effectiveness of specific educational interventions to inform decision-makers. And, in particular, there is very limited information on the costs, and consequently the cost-effectiveness, of technology-based approaches to supplement algebraic instruction. Early related work focused on the CEA of computer-assisted instruction (e.g., Fletcher et al., 1990; Keltner & Ross, 1996; Levin et al., 1987). While the specific cost findings are no longer relevant due to obsolete technology, the frameworks and general recommendations are still applicable today. In one of the first studies to examine the cost-effectiveness of computer-assisted instruction, Levin et al. (1987) provided estimates of the cost-effectiveness of computer-assisted instruction and three other educational interventions and concluded that the “most appropriate use of these results was to provide guidelines for the consideration of alternative interventions for increasing mathematics and reading achievement in elementary schools” (p.70). They concluded educators should question unqualified assertions that computer-assisted instruction is a more cost-effective intervention than other alternatives. This conclusion emphasizes the need for rigorous CEA of interventions because the results may contradict popular beliefs—among both researchers and policymakers—about which interventions should be implemented based only on effectiveness.

Despite early calls for research on the CEA of computer-assisted or technology-based interventions (Levin et al., 1987), there are few relatively recent studies that examine the costs and effectiveness of such applications, particularly for algebra. In one of the few cost analysis studies of algebraic (technological) curriculum, Daugherty et al. (2012) examined the costs of Carnegie Learning’s Cognitive Tutor[®] Algebra I relative to the other curricula in a randomized, controlled trial experiment in approximately 150 schools in seven states. Their analysis found the Cognitive Tutor[®] Algebra I curriculum cost was \$97 per student, compared with \$28 per student for the other algebra I curricula. Pane et al. (2014) reported on the effectiveness of Cognitive Tutor[®] Algebra I from a large-scale RCT, reporting that Cognitive Tutor[®] Algebra I improved the median student’s performance by approximately eight percentile points in the second year of implementation. While Daugherty et al. (2012) found that Cognitive Tutor[®] Algebra I was substantially more expensive than the comparative curriculum, Pane et al. (2014) also demonstrated that Cognitive Tutor[®] Algebra I was more effective. They concluded the cost must be weighed alongside the benefits, and educators should judge whether the positive effects are large enough to warrant the additional cost.

A CEA would allow decision-makers to compare the efficiency of interventions based on costs and effectiveness and determine which intervention yields a given level of effectiveness for the lowest cost. However, to our knowledge, the costs and impact measures of Cognitive Tutor[®] Algebra I and the comparison curriculum were never reported as cost-effectiveness ratios, which would have assisted decision-makers in determining which interventions were more efficient.

Besides the examples above, publishers' fees are the basis of many of the reported costs of technology-based approaches (e.g., What Works Clearinghouse [WWC] Intervention Report, The Expert Mathematician; WWC Intervention Report Cognitive Tutor[®]). Since the estimates do not include the full opportunity costs of the intervention, they likely underestimate the cost. A common critique of cost studies is that they do not include the costs of all the ingredients (Levin et al., 2018). Identifying and then valuing all of the resources or ingredients needed to implement the intervention using the ingredients method (Levin et al., 2018) provides a more complete version of the costs by accounting for the societal, district, and/or school costs to implement the intervention.

The ingredients method (Levin et al., 2018) is based on two primary principles—opportunity cost and cost accounting—and is the most widely recognized approach for estimating the full economic cost of a well-defined intervention (Shand & Bowden, 2022). The ingredients method requires a detailed account of all resources or ingredients required to implement an intervention to achieve a particular outcome in a specific setting. That is, district and/or school facilities, personnel, and equipment that are used to implement the intervention based on multiple data sources and/or methods, including program documentation, budgets, observations, interviews, and surveys of those who implemented an intervention (Levin et al., 2018). Fully accounting for the costs of the intervention provides a more accurate picture of the efficiency of the intervention and allows decision-makers to assess whether they have all the required ingredients to implement the intervention. We apply the ingredients method in this study to account for all of the resources required to implement the interventions. Moving from the need to assess the costs, in the next section, we discuss the effectiveness of algebraic technological applications.

Promising Approaches to Address Pandemic Learning Loss in Mathematics: Effectiveness of Algebraic Technological Applications

Research in math education and cognitive science has provided evidence of several factors that could improve the effectiveness of instructional mathematic software, including emphasizing conceptual understanding and algebraic structure (Knuth et al., 2005; McNeil et al., 2015; Rittle-Johnson et al., 2015; Schneider et al., 2011; Schoenfeld, 2007) and supporting symbolic reasoning as perceptual-motor learning (Catley & Novick, 2008; Goldstone et al., 2010; Jacob & Hochstein, 2008; Kellman et al., 2010; Kirshner & Awtry, 2004; Landy & Goldstone, 2007; Patsenko & Altmann, 2010). This literature indicates that students rely on the visual patterns available in notations to learn reasonable patterns of mathematical behaviors taken upon symbolic objects. Based on this finding, technological applications that utilize perceptual learning strategies and allow students to physically interact with objects on the screen through dynamic motion and play may provide a useful learning environment to explore mathematical ideas and algebraic structure and improve conceptual understanding.

Developers of algebraic technological applications have designed digital tools to support learning in different ways to improve mathematical understanding and performance among middle school students. One example is a game-based application

that embeds the pushing symbols framework and teaches math through discovery-based puzzles rather than procedural steps. The pushing symbols framework provides a concrete model of how to implement perceptual learning systems and core cognitive theory into math instruction using technology (Ottmar et al., 2012). Algebraic expressions are turned into interactive virtual objects that react according to their underlying mathematical properties. Users can dynamically manipulate and transform math expressions by directly dragging, tapping, slicing, sliding, and breaking apart parts of the equation on the screen. The goal is to make the user interface as natural and intuitive as possible. This approach also incorporates perceptual training, embodied cognition, and game design elements to address many of the factors that lead to low proficiency, including poor understanding of the equals sign (Knuth et al., 2005, 2006) and failure to connect procedural knowledge, conceptual understanding, and real-world applications (Clement et al., 1981; Rittle-Johnson et al., 2015; Schoenfeld, 2007).

Widely used technological applications, such as FH2T and DragonBox, are designed to engage students in interactive game-based learning, allowing students to interact with algebraic notations and solve puzzle-like problems in a playful environment. Other applications, for example, ASSISTments, are designed to provide timely support and feedback on homework, using problems that resemble those in traditional mathematics textbooks. Below, we provide an overview of the conditions assessed in this CEA, namely FH2T, DragonBox, Immediate Feedback, and Active Control, and evidence of their effectiveness in improving mathematical learning.

From Here to There! (FH2T)

FH2T (<https://graspablemath.com/projects/fh2t>) is a game-based application that embodies the approach above by incorporating a pushing symbols framework and teaching math through discovery-based puzzles rather than procedural steps. Previous studies suggest that the game-based FH2T system may be effective in decreasing structural errors and improving math understanding (Chan et al., 2022; Ottmar et al., 2012, 2015; Ottmar & Landy, 2017). In addition, past evidence suggests that game-based dynamic systems like FH2T may help increase engagement, math efficacy, and interest in learning algebra, and may serve as a buffer against the detrimental effects of math anxiety on performance (Ottmar et al., 2012).

DragonBox

DragonBox (<https://dragonbox.com/products/algebra-12>) is another educational application that aims to teach algebraic concepts to students in an “intuitive, interactive, and efficient way.” The application incorporates a discovery puzzle-based approach, embedded gestures, multiple representation integration, varying levels of challenge, immediate feedback, and adaptability (Cayton-Hodges et al., 2015; Torres et al., 2016). Despite its promising design features and popularity (for example, it won a Gold Medal from the 2012 International Serious Play Awards, the Best Educational Game Award at the 2012 Fun and Serious Game Festival.), the research findings on its efficacy are mixed, with some research demonstrating significant gains in algebra problem

performance (Dolonen & Kluge, 2015; Liu et al., 2015; Shapiro, 2013) and students' attitudes toward math (Siew et al., 2016), and other research finding no improvements in problem-solving performance or student confidence (Long & Alevan, 2014, 2017), or lower learning gains compared to students using problems from standard algebra textbooks (Dolonen & Kluge, 2015). The mixed findings underscore Long and Alevan (2017) conclusion that more rigorous studies are needed to test out-of-game transfer of learning.

Immediate Feedback

Immediate Feedback entails problem sets using an online homework system, ASSISTments (<https://new.assistments.org/>; Heffernan & Heffernan, 2014). ASSISTments is a free, online tutoring system that offers immediate feedback to as they solve traditional textbook problems, and is currently being used by 50,000 students worldwide. In contrast to the two applications above, it does not include elements of game-based learning. The design of ASSISTments is based on the research that indicates when instruction is adjusted based on formative assessment and students are provided timely feedback, they show significant performance improvement (Bergan et al., 1991; Butler & Woodward, 2018; Shute, 2008; Speece et al., 2003). And, additional research underscores the importance of timely feedback being immediate (Azevedo & Bernard, 1995; Corbett & Anderson, 2001; Dihoff et al., 2003). Based on this research, ASSISTments is designed to provide students with immediate feedback and on-demand hints as scaffolds during problem-solving. Murphy et al. (2020) conducted an RCT and found that students in schools assigned to use ASSISTments learned more and the impact was greater for lower performing students.

Active Control

The Active Control condition mimicked traditional homework assignments while still using technology. In this condition, students received post-assignment feedback, including a report with feedback at the end of the problem set and the opportunity to review their responses, revisit problems, and request hints. It also used ASSISTments but removed the immediate feedback feature so that it mirrored traditional math problem practice while still using a device. This was done so that all students in the RCT classrooms were working on a device.

While there is a promise for algebraic technological applications to improve student performance and achievement, there are few studies, if any, that simultaneously examine the costs and assess the cost-effectiveness of such algebraic technological interventions. In addition to the research above, two recent meta-analyses examined the effects of game-based learning on students' mathematics performance and self-efficacy (Byun & Joung, 2018; Tokac et al., 2019). While these meta-analyses emphasized the need for more rigorous research, they also demonstrate the extent of research that is available about effectiveness, but not CEA. Without simultaneously considering the costs, decision-makers do not know the efficiency of such interventions. CEA is needed to assess the efficiency of these algebraic technological applications.

Purpose of the Study

The purpose of this study was to conduct a CEA of the four conditions—FH2T, DragonBox, Immediate Feedback, Active Control—using the ingredients method (Levin et al., 2018) and answering the following question: What is the cost-effectiveness of algebraic technological applications—specifically FH2T, DragonBox, and Immediate Feedback—on students’ algebraic performance compared to the Active Control? Or, in other words, which technological application—FH2T, DragonBox, or Immediate Feedback—yields a given level of effectiveness for the lowest cost compared to the Active Control? This CEA study uses the impact measures from a larger IES efficacy study that used a student-level RCT to compare student learning from FH2T, DragonBox, and Immediate Feedback to an Active Control condition. (Decker-Woodrow et al., 2023). We use the ingredients method to estimate the costs of the four conditions and combine those with the impact measures from the RCT efficacy study to estimate their cost-effectiveness. The results provide cost-effectiveness ratios for FH2T, DragonBox, and Immediate Feedback compared to the Active Control.

Methods

Background on Efficacy Study and Research Context

The goal of the larger efficacy study was to independently examine the efficacy of three widely used treatment conditions (i.e., FH2T, DragonBox, and Immediate Feedback using ASSISTments) on students’ algebraic understanding compared to the Active Control. In total, 37 teachers, 156 classes, and 3,612 students were randomly assigned into the four conditions, within classrooms, across 10 schools (9 in-person schools and 1 virtual school). The RCT study randomly assigned students to study conditions, within classrooms, ensuring study conditions were equivalent with respect to teacher characteristics and classroom curricula. This approach was possible because teachers were able to use all four technology-based interventions within the classroom. After the pretest assessment, one school dropped out, resulting in a final pool of 9 schools (8 in-person and 1 virtual), 34 teachers, 143 classes, and 3,271 students. From this pool of 3,271 students, 1,850 students across 127 classes and 34 teachers had both pretest and posttest assessments and served as the analytic sample, resulting in an overall attrition rate of 48.8% (Decker-Woodrow et al., 2023). In comparing attrition rates across conditions, differential attrition was not statistically significant, and differential and overall attrition rates were within tolerable threats of bias under optimistic assumptions (WWC, 2020). The analytic sample was evenly split between male (50.4%) and female (49.6%) students. The majority of students were White (52.4%) with 24.8% identified as Asian, 14.5% as Hispanic, 4.3% as African American, and 4% as another race/ethnicity (Decker-Woodrow et al., 2023).¹

As the larger study was conducted between September 2020 and April 2021, during the peak of the COVID-19 pandemic in the United States, the school district offered the students and their families a choice of instructional modality (in-person or virtual academy) for the 2020-2021 school year before the start of the fall semester. Among the

¹For additional information on the sample, including the random assignment process, research procedure, attrition rate, and demographic information by condition, see Decker-Woodrow et al. (in press).

analytic sample of 1,850 students, 1,245 students were in-person (67%) and 605 students were virtual (33%). The proportion of in-person and virtual students in each condition was similar to the overall proportions with 33% of students participating in the Immediate Feedback, FH2T, and Active Control virtually (and 67% in-person) and 31% of students participating in the DragonBox condition virtually (and 69% in-person). Regardless of students' instructional modality, all study sessions were administered online during their regular math classes (for in-person students) or as part of learning activities (for virtual students) and students worked individually at their own pace using their devices. Students received nine 30-minute intervention sessions across the school year, with a 2-week window to complete each session.

To assess impact, students received a pre and posttest assessment on algebraic knowledge. The algebraic knowledge assessment consisted of ten multiple-choice items from a previously validated measure of algebra across the conceptual understanding of algebraic equation-solving (e.g., the meaning of an equal sign), procedural skills of equation-solving (e.g., solving for a variable), and flexibility of equation-solving strategies. The measure of internal consistency Cronbach's α is .89 (see Star et al., 2015b for additional information)². The *pretest* assessment was administered in September 2020, approximately 1 week prior to the intervention sessions. The *posttest* assessment was administered between the end of March and the beginning of April 2021, approximately 2 weeks following the completion of the intervention. For students receiving instruction in person, teachers dedicated instructional periods for the study assignments in mathematics classrooms. For students receiving virtual instruction, teachers included the study assignments as a part of students' learning activities. To ensure that students spent a similar amount of time (i.e., 30 minutes per session) regardless of their condition assignments, a countdown timer was embedded in all technologies. See Table 1 for intervention conditions, assignment, and key ingredients.

Approach to Estimating Costs

To conduct the concurrent CEA of the four conditions, we used the ingredients method (Levin et al., 2018; Levin & McEwan, 2001) and followed the steps in the Cost Analysis: A Toolkit (Institute of Education Sciences, 2020) and Cost Analysis Standards & Guidelines 1.1 (Hollands et al., 2021). The ingredients method of cost estimation involves three main steps to obtain accurate and consistent measures of cost: 1) identifying and specifying the ingredients required to obtain the evaluation results; 2) determining their costs; and 3) calculating total program costs and the average cost per participant (Levin et al., 2018). To identify resources needed to implement the algebraic technological applications, we collected data on all staff, materials, equipment, facilities, training, technical infrastructure, and other inputs required for implementing the algebraic technological applications during implementation. We reviewed programmatic materials, including the program's purpose, theory of change, and logic model, and conducted targeted interviews with technology coordinators and technical support personnel

²The 10 items are available at osf.io/bafdr

Table 1. Intervention conditions, assignment and key ingredients.

Condition	Description	Key ingredients				
		Assignment to condition	Personnel	Facilities	Equipment and materials	
Active control	Traditional homework assignments using technology	Randomly assigned within classroom	<ul style="list-style-type: none"> Middle school math teachers (virtual and in-person) 	<ul style="list-style-type: none"> Middle school classrooms, district/home office space 	<ul style="list-style-type: none"> Tablets Training materials and resources for the district-level coordinator 	<ul style="list-style-type: none"> Licensure fees (only DragonBox), transportation costs of a district-level coordinator translation services
Immediate Feedback	Problem sets using an online homework system, based on ASSISTments		<ul style="list-style-type: none"> District-level math content specialist, District-level coordinator 	<ul style="list-style-type: none"> district training space 		
DragonBox	Game-based application		<ul style="list-style-type: none"> Assistant coordinator District-level data (or information technology) staff 			
FH2T	Game-based application					

to identify resource demands by four broad categories: personnel, facilities, equipment and material, and other inputs.

To identify and estimate the price for each ingredient, we used several resources (e.g. CostOut[®] Database of Educational Resource Prices, district salary schedule, and national vendors). For the cost per unit of personnel, we used the district's salary schedule to ascribe prices to personnel based on hourly or daily rates. We also use the national price for the teachers' salaries in the sensitivity analysis. Based on the recommendation of Institute of Education Sciences Institute of Education Sciences (2020), we did not include students' time in the CEA. We obtained pricing for facilities (e.g., middle school classroom, and district offices) from CostOut. For the cost per unit of materials, equipment, and other inputs, we used their market value (for example, based on prices on Amazon.com and/or, when available, CostOut). Prices from national vendors are often the same for different locations, in which case, the national price equals the local price (Hollands et al., 2021). The costs of equipment were annualized, and, in some cases, prorated based on participants' usage (e.g., equipment and materials for district coordinator). In cases where we had to adjust prices, we adjusted to 2020-2021 dollars.

Based on the recommendations of Levin et al. (2018), we applied a societal perspective to estimate costs. That is, we include all program-related costs, regardless of who pays or contributes to the resources. For example, we included in-kind donations, such as the license fees for DragonBox, as part of the total costs. Also, we do not include application development costs, since they are considered sunk costs (Cost Analysis Standards Project, 2021) and are not relevant for districts and schools delivering the intervention. This perspective provides a "reference case" to allow for comparisons of the use of resources by different programs (Hollands et al., 2021).

To estimate the costs of the ingredients, for each item under the four broad categories, we established a unit of measure (e.g., hourly or daily rate of staff, sq. ft. of building), identified the quantity used by the intervention based on the approaches above, and multiplied it by the cost per unit. We inserted the unit of measure, price, and quantity of ingredients into CostOut, an existing IES-funded online tool (Hollands et al., 2015) developed for estimating the costs of educational interventions to estimate the costs of the ingredients for implementing each algebraic technological application. In several instances, we followed the recommendations from Shand and Bowden (2022) to identify the quantities used by the intervention, discussed further below.

As noted previously, the intervention took place during the peak of the COVID-19 pandemic and was implemented in a virtual and in-person setting. For this CEA, we estimated costs as in-person, which, in essence, uses district and school ingredient prices as shadow prices for remote ingredients. For example, the price of at-home office space for the district coordinator is estimated as the same as district office space. Using remote prices could potentially lower the overall costs proportionately for all conditions and would shift the costs to different groups. In this case, costs would shift from the district or school to an individual. But, given the RCT design that randomly assigned students within classrooms, this approach does not have implications for the comparison of cost-effectiveness ratios across the algebraic technological application, and it reflects how the interventions are intended to be implemented in the future.

After we obtained and identified the costs, we aggregated all the costs to estimate the total program cost across the program's lifespan, using a 30-year lifecycle for facilities and a 5-year lifecycle for equipment and materials. We also used a discount rate of 3%, as recommended by the Cost Analysis Standards Project, [2021]). The total program cost was divided by the number of participants who actually received and completed the treatment (i.e. the analytic sample) to estimate the upper-bound average cost per participant (Shand & Bowden, 2022). To determine the cost-effectiveness ratio, we divided the average cost per participant by the average effectiveness for each algebraic technological application.

Sample

This CEA study uses the sample from the IES efficacy study (Decker-Woodrow et al., 2023). Again, there was a final pool of 9 schools, 34 teachers, 143 classes, and 3,271 students participating at the start of the interventions. We used this sample to calculate the ingredient costs of the intervention with one exception—licenses for DragonBox. Given this is one of the few differences between the conditions and considering the unusually high attrition due to the COVID-19 pandemic, we use the analytic sample instead of the initial sample. Otherwise, this approach captures the costs of the intervention as implemented. From this pool of 3,271 students, 1,850 had both pretest and posttest assessments and constituted the analytic sample for this study. The final analytic sample for the IES RCT efficacy study was 1,850 students (FH2T $n = 753$, DragonBox $n = 350$, Immediate Feedback $n = 381$, and Active Control $n = 366$) in 127 classes, 34 teachers across 9 schools. Following the recommendation of Shand and Bowden (2022) regarding the sample over which to divide costs in the presence of attrition, we use the participants who received and completed the intervention to obtain an upper-bound cost estimate. The effectiveness measure that we adopt and apply in this study is derived from the analytic sample.

Effectiveness Estimates

The effectiveness estimates were obtained from an IES RCT efficacy study that examined the impact of FH2T, DragonBox, and Immediate Feedback compared to the Active Control condition (Decker-Woodrow et al., 2023). The authors assessed intervention effects through a series of 3-level hierarchical linear modeling (HLM) regression models, with 1,850 students nested within classrooms ($n = 127$) and nested within teachers ($n = 34$). Decker-Woodrow et al. (2023) conducted four HLM models predicting posttest scores on the algebraic knowledge assessment. The interventions' mean outcomes, that is the average post-test scores, were FH2T = 4.59 (SD = 2.96), DB = 4.81 (SD = 2.88), Immediate Feedback = 4.58 (SD = 2.90), and Active Control = 4.29 (SD = 2.79). The HLM model that controls for demographics such as gender, race/ethnicity, and gifted status; factors in pretest scores of algebraic knowledge; and adds number of completed assignments (dosage) and enrollment in physical or online classrooms as two post-randomization variables (i.e., Model 3) indicates the interventions are statistically significant ($\chi^2(3) = 20.863$, $p < 0.001$). The model coefficients indicate both FH2T ($\gamma = .361$) and

Table 2. Per student ingredient costs by intervention condition.

Ingredients	FH2T, Immediate Feedback, and Active Control		DragonBox	
	Costs per student	Percent	Costs per student	Percent
Personnel	\$32	80%	\$40	72%
Facilities	<\$1	2%	<\$1	1%
Equipment & materials	\$6	15%	\$6	11%
Other inputs (e.g., license fees)	<\$1	2%	\$9	16%
Total	\$39	100%	\$55	100%

Note. Differences due to rounding.

DragonBox ($\gamma=.719$) had significantly larger effects on algebraic knowledge than the Active Control condition. However, Immediate Feedback had an insignificant effect of $\gamma = .254$. The standardized measure of intervention effects, that is the Hedges' g effect size estimates after controlling for all of the covariates, were FH2T = 0.135, DB = 0.269, and Immediate Feedback = 0.095.³ We use the Hedges' g effect size estimates as the effectiveness measures in the cost-effectiveness estimates. (See Table A1 in Appendix A for a summary of Model 3 coefficient estimates).

Results

Ingredient Costs of Algebraic Technological Applications

The ingredients' costs total \$39 per student for FH2T, Immediate Feedback, and Active Control, and \$55 for DragonBox. Table 2 presents the ingredient costs per student for the algebraic technological applications FH2T, DragonBox, Immediate Feedback, and Active Control. We discuss the conditions' ingredients costs by each category below.

Personnel

For personnel, the key ingredients identified were middle school math teachers (virtual and in-person), a district-level math content specialist, a district-level coordinator, an assistant coordinator, and district-level data (or information technology) staff. The implementation of the applications occurred in the classroom during the school day and consisted of 1 hour of teachers' time per assignment (including time for device management, login management, and student participation based on intervention sessions, device tracking, and teacher reports) amounting to 25 hours per teacher (including training) across the school year. We multiplied the number of teachers ($n = 34$) by the 25 hours per teacher to estimate the quantity of middle school math teacher hours at 850. We then multiplied the quantity of middle school math teacher hours (850) by an hourly rate of \$42.22 based on the district average annual teacher salary of \$60,795 and using the wage converter in CostOut to estimate the hourly rate using the academic year⁴. resulting in an estimate of \$35,890 for middle-grade math teachers. It is

³For continuous outcomes, the WWC recommends using the most commonly used effect size index, the standardized mean difference known as Hedges' g, with an adjustment for small samples. It is defined as the difference between the mean outcome for the intervention group and the mean outcome for the comparison group, divided by the pooled within-group standard deviation of the outcome measure.

⁴In-person and virtual teacher salaries did not vary. In CostOut, the academic year is 1,400 working hours [36 weeks, 5 days a week, 8 hours a day]

important to note that for DragonBox, teachers spent additional time with setup, login, and device management in August and September, estimated at 2 hours per teacher based on teacher reports. This additional time resulted in an additional cost of \$2,870 of middle school math teacher time being attributed to DragonBox. The next largest cost in personnel ingredients was the math coordinator that spent an estimated 315 hours, at an hourly rate of \$41.19 (based on academic year), supporting the implementation of the applications from July to May, resulting in an estimated \$12,970. The secondary math content specialist that supported startup and initial training and served as a liaison for district program implementation support spent an estimated 13 days, at a daily rate of \$329.55 (based on academic year), supporting the program implementation, resulting in a cost estimate of \$4,280. The remaining cost is attributable to district data staff that pulled outcome and demographic data on students and provided implementation reports. The time for this support includes 1 day per month of technical support for August through May, at a daily rate of \$329.55 (based on academic year), and 4.5 hours for each of the 13 sessions at an hourly rate of \$32.49 (based on academic year), resulting in an estimate of approximately \$5,200. These estimates totaled approximately \$58,340 in personnel ingredients costs across the four conditions, with differences due to rounding, with an additional \$2,870 for DragonBox.

The personnel costs per student for FH2T, Immediate Feedback, and Active Control are \$32. This cost is mostly attributable to middle-grade math teachers' time to implement the interventions (62%) and a district coordinator to support implementation (22%). Other personnel support staff (e.g., district-level information technology staff, content specialists, and assistant coordinators) account for the remaining 16% of personnel costs. The approximately \$8 increase in the personnel costs of DragonBox compared to the other conditions (\$32 compared to \$40) is attributable to increased middle-grade math teachers' time for setup, device, and login management only for that condition.

Facilities

For facilities, the key ingredients included middle school classrooms, district/home office space, and district training space. Based on the usage rates above, we estimated the quantity for each facility type and multiplied the quantities by prices available in the CostOut database of education resource prices for regular middle school classrooms (2020 prices) and mid-rise commercial offices (2020 prices). For example, we also estimated the cost of a middle school classroom based on students' aggregated usage during the intervention (accounting for approximately 11% of one classroom for a calendar year) and multiplied it by the annualized price of a classroom (using a 30-year lifecycle). The district also held a 4-hour in-person training for teachers at the beginning of the year in a space large enough for teachers to spread out due to COVID-19 (approximately 1020 square ft.). Based on the description of the building, we used the mid-rise commercial office (adjusted) price per sq. ft. in the CostOut database of education resource prices to estimate costs for a half day of use (based on a 30-year lifecycle). These facilities estimates totaled approximately \$1,540, resulting in facility ingredient cost of less than \$1 per student for each condition.

The facility cost per student for all conditions is less than \$1. Approximately 85% of the costs are attributed to classroom facilities. District office space and training space

account for the other approximately 15% of the facility costs. Again, the lack of variation in facility costs between the conditions is attributable to the random assignment of students to conditions within the classroom. In addition to using the same classroom facilities, the district provided training to all the teachers as a cohort, so there were also no differences in district training facilities between conditions. The facility costs for the district math coordinator are also the same across conditions because the coordinator oversaw the implementation of the four conditions equally.

Equipment and Materials

The primary equipment and materials were student tablets; other minor equipment and materials include training materials and resources for the district-level coordinator (laptop, phone, printer, supplies). The RCT study team provided tablets that were purchased at market rate for \$45 in 2019 and 2020 to the students. Students shared the tablets at approximately a 3 to 1 (student/tablet) ratio. We used the amortization calculator in CostOut to estimate the annual cost per tablet, using a 5-year lifecycle with a 3% discount rate (the recommended discount rate in Cost Analysis Standards Project, [2021]), resulting in approximately \$9.83. Using this approach, we estimated the ingredient cost of the tablets at \$10,700. The students also used district-owned Chromebooks to sign into accounts before using the tablets for DragonBox; however, this was for the research team needed to track student participation, so we excluded those costs and used only the cost of tablets to reflect future implementation of the conditions. The other equipment and materials (training supplies, coordinator equipment, and materials) were estimated at \$500. These estimates totaled approximately \$11,200, resulting in equipment and material ingredient costs of approximately \$6 per student for each condition. The tablets account for almost 96% of these costs.

Other inputs

Other inputs include licensure fees (relevant only to DragonBox at this time), transportation costs of a district-level coordinator, and translation services (for parent/guardian letters). The DragonBox license fee is \$7.99 per student, costing approximately \$2,800 for 350 students. (Again, given this is one of the few differences between the conditions and considering the unusually high attrition due to the COVID-19 pandemic, we use the analytic sample instead of the initial sample.). The other ingredients consisted of translation services (\$900), phone services for the coordinator (\$320), and coordinator transportation (\$400), amounting to a total of \$1,610 (differences due to rounding). For other inputs, the cost per student for FH2T, Immediate Feedback, and Active Control is less than \$1 compared to \$9 for DragonBox. The difference in other costs across interventions is attributable to the \$7.99 single-use license fee for DragonBox. While this reflects the license costs a district would have to pay to implement DragonBox in the future and is thus included in the primary analysis, given the licenses were donated in-kind, we also estimate the costs without the license fee in the sensitivity analysis. We also estimate the costs of the licenses for the initial sample of 3,271 students in the sensitivity analysis to demonstrate the difference between using the initial sample and the analytic sample for DragonBox.

Table 3. Cost-effectiveness results by intervention condition.

Components	FH2T	DragonBox	Immediate feedback	Active control
Cost per student	\$39	\$55	\$39	\$39
Effect size (Hedges' g) on mathematic achievement	0.135	0.269	0.095	0
CE ratio	\$291	\$206	\$414	—

Note. Differences due to rounding.

CEA of Algebraic Technological Applications

The results from the cost analysis indicate FH2T costs \$39 per student. Dividing the cost of \$39 per student by the average effect size (Hedges' g) of 0.135 on math achievement results in a cost-effectiveness ratio of \$291 (differences in cost-effectiveness ratios due to rounding). DragonBox, on the other hand, costs \$55 per student and has an average effect size of 0.269 on math achievement, resulting in a cost-effectiveness ratio of \$206. And Immediate Feedback costs \$39 per student and had an average effect size of 0.095, statistically insignificant, resulting in a cost-effectiveness ratio of \$414. Given that a lower ratio means less cost per unit of gain, these results show that DragonBox is the most cost-effective of the three algebraic technological applications and that both DragonBox and FH2T are more cost-effective than Immediate Feedback. The results are presented in Table 3-CEA results by intervention condition.

Sensitivity Analysis

Estimating costs always involves uncertainty, and estimates may vary (Briggs et al., 2012), hence it is critical to perform sensitivity analyses that vary estimates to test the robustness of the results (Boardman et al., 2018). Approaches to sensitivity testing range in complexity from 1) identifying the largest, most influential parameters and varying them while holding all else equal (referred to as partial sensitivity analysis), 2) considering worst and best case scenarios based on multiple combinations of the least favorable and most conservative assumptions, 3) conducting Monte Carlo simulations that draw key parameter estimates from probability distributions (Boardman et al., 2018).⁵

We conducted a partial sensitivity analysis and varied the largest parameters in the model (e.g., staffing salaries, facility costs, and equipment costs). Since so many of the ingredients are similar across the algebraic technological applications, varying the parameters in the cost model does not change the implications of the results of the cost-effectiveness ratios. That is, we found efficiency gains moving to DragonBox and FH2T under the different scenarios. For example, using an average teacher salary rate of \$45.20 derived from the \$65,090 estimated average annual salary of teachers in public elementary and secondary schools in 2020-21 (NCES, 2021) instead of \$42.22 results in a cost per student of \$41 for FH2T, Immediate Feedback, and Active Control and \$57 for DragonBox, which changes the cost-effectiveness ratios to \$301 for FH2T, \$213 for DragonBox, and \$427 for Immediate Feedback. Additionally, we estimated the per student cost of DragonBox using the additional licensure fees for the initial 654 students participating at the start of the interventions, as opposed to the 350 in the analytic

⁵A Monte Carlo simulation is a model used to predict the probability of a variety of outcomes when the potential for random variables is present.

sample. Changing this parameter increase cost per student of DragonBox to \$62 and increases the cost-effectiveness ratio to \$232. However, as discussed above, this rate of student attrition is unlikely in future scenarios in the absence of the COVID-19 disruptions, though districts may expect to purchase more licenses for students that initially start the intervention than ones that complete it. While variations in cost parameters in the model do not change the implications of the CEA, they do demonstrate the possible range in per student costs and the cost-effectiveness ratios, which may be important for decision-makers considering multiple alternatives and feasibility (discussed below).

There are also a couple of ways DragonBox in particular may improve its efficiency. For example, given one of the few differences in costs for DragonBox stemmed from additional teacher time to set up devices, mitigating this issue in the future implementation of DragonBox could potentially reduce the cost per student to \$47 and improve the cost-effectiveness ratio to \$176. This may be a realistic scenario if the same teachers use DragonBox over multiple years and become more familiar with the application. And, while we include the license fees of DragonBox in the cost per student to capture all the costs regardless of who pays, if the licenses were donated in-kind (as was the case with the IES efficacy study), the cost per student decreases to \$48 and the cost-effectiveness ratio improves to \$177. Both of these scenarios reflect ways DragonBox could improve its productivity.

Discussion

Summary of Findings

While there is a large amount of research on the effectiveness of interventions and strategies for teaching algebra (Star et al., 2015a), there is a lack of comparable information about the cost-effectiveness of such algebraic interventions. This CEA study provides information about the costs and effectiveness of three algebraic technological applications and demonstrates the efficiency of two in particular—FH2T and DragonBox. Using DragonBox compared to FH2T increases efficiency by about one-third. Both are relatively low-cost, effective interventions for improving students' algebraic performance. The Immediate Feedback condition, on the other hand, had an insignificant effect, was the least efficient of the three conditions, and was approximately double the cost-effectiveness of DragonBox. Furthermore, the results of this CEA indicate that DragonBox had a large effect size ($ES = \geq .20$) and low cost ($< \$500$), as indicated in Kraft's Kraft (2020) schema for interpretation of cost-effectiveness ratios. FH2T falls within the medium effect size ($ES = .05$ to $< .20$) and is low cost.

Since CEA combines costs and impact measures, there are study limitations that pertain to both. Firstly, regarding the effectiveness measures, all the limitations of the impact study (Decker-Woodrow et al., 2023) pertain to this CEA study. The authors noted several limitations, including non-trivial attrition due to the pandemic, with almost half of the students dropping out of the study by the end of the intervention.⁶ This limitation is pertinent to the CEA because using the initial participating sample

⁶The authors also noted the dropout rate was not related to the intervention condition and there was no statistically significant differential attrition between conditions.

versus the analytic sample with unusually high student attrition results in a higher upper-bound estimate than would likely be typical. Given these unusual circumstances, we demonstrate different approaches to estimating these costs in the sensitivity analysis.

Another limitation of this CEA is the capacity to fully capture ingredients and corresponding costs of students that participated virtually. As previously discussed, since we estimated costs as in-person, district and school prices are used in the analysis instead of remote prices. Consequently, it does not account for differences in the costs of residential vs. school or district facilities and equipment, and it may overlook some unintended ingredients that are not part of the planned interventions (e.g., if parents/guardians elected to support student engagement in the study). The cost-effectiveness ratios are representative of implementing the interventions in person, which is how they are intended to be, and how the majority of students participated in the study. However, decision-makers considering using a combination of in-person and virtual instruction to deliver the interventions should be cautious making comparisons of the cost-effectiveness ratios to other algebraic interventions.

In addition to informing decision-makers' choices regarding resource allocation, further research could build on this work by comparing these cost-effectiveness estimates to other cost-effectiveness estimates of algebraic interventions. By routinely incorporating cost-effectiveness into evaluations of educational programs and applying a similar methodology (see, for example, guidelines and recommendations in Cost Analysis Standards Project, 2021; Levin et al., 2018; Institute of Education Sciences, 2020; Hollands et al., 2021; Shand & Bowden, 2022), it will allow for more precise cost-effectiveness comparisons and consequently be more informative for decision-makers. Additionally, researchers should aim to incorporate common measures into their CEA to help benchmark and compare results (see Schneider [2020] for a discussion and [EdInstruments](#) for specific examples.) Furthermore, advancements in CEA analysis have also led to PowerUp!-CEA, which is an excel workbook designed to aid in power analysis for multilevel randomized cost-effectiveness trials (Li et al., 2021) that researchers can apply in future studies. However, an important caveat of cost-effectiveness results is that they can only be compared when they use similar measures of effectiveness. We elaborate on how cost-effectiveness results can be used to inform decision-making below.

The Role of CEA in Decision-Making

The value of CEA lies in informing decisions to allocate limited resources to maximize effectiveness in achieving education aims. While the advantage of CEA analysis lies in being able to compare two or more alternatives, one of the major disadvantages of CEA is one can only compare cost-effectiveness ratios across interventions with similar goals, in this case, algebraic achievement. Decision-makers using the cost-effectiveness results from this study must only make comparisons across interventions that impact students' algebraic achievement. That is, the results can be used to make comparisons of one intervention to another in terms of relative cost-effectiveness on algebraic achievement. This is also an area for future research, that is, to examine if and how so decision-makers may use this information. Decision-makers need to assess whether the ingredients required to implement an intervention are available at comparable costs (Tsang, 1997).

In making any comparison between our estimates of the cost-effectiveness ratios for these conditions and estimates from other studies, decision-makers should pay close attention to what ingredients were included in any cost estimates. While the cost per student may be higher than reported costs of similar types of interventions (e.g., WWC Intervention Report *The Expert Mathematician*; WWC Intervention Report *Cognitive Tutor*[®]), it is because we identified and then valued all of the resources needed to implement the intervention using the ingredients method (Levin et al., 2018). This approach provides a more complete version of the costs by accounting for the societal, district, and/or school costs' to implement the intervention. Based on publisher fees alone, DragonBox would be approximately \$8 per student instead of our estimate of \$51. While our estimate includes the opportunity costs of implementing the intervention (e.g. costs of personnel, facilities, and equipment), districts and schools adopting the intervention may only directly incur this additional or incremental \$8 per student cost for implementing DragonBox.

Furthermore, it is important to consider the results of the CEA in regard to the scale of the intervention. This CEA used costs and effectiveness measures from a relatively large-scale RCT implemented across 10 middle schools. As these types of algebraic technological applications are brought to a larger scale, most of the variable costs will remain similar, however, some efficiencies may be gained. As discussed in the sensitivity analysis, in particular, as teachers implement the applications over multiple years, any additional setup time for DragonBox compared to the other applications may decrease, or DragonBox may find ways to streamline this process. Furthermore, any flat rate costs for implementing the program would decrease proportionately to the number of students using the program. These types of considerations may influence decision-makers' assessment of cost-feasibility.

Conclusion

This study is one of a few studies to provide decision-makers with information about the CEA of algebraic technological applications. Given the previous impact findings, both game-based applications were promising interventions that improve students' algebraic performance by training students' perceptual-motor routines in algebraic reasoning. This study contributes to that literature by further demonstrating that these game-based applications are also relatively low-cost interventions. Overall, the current CEA study demonstrates the efficiency of FH2T and DragonBox as interventions for improving algebraic performance. This information can be used by decision-makers considering the productivity of alternative algebraic educational interventions to improve the efficiency with which public resources are employed to address learning loss brought on by the pandemic.

Acknowledgments

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the U.S. Department of Education, Institute of Education Sciences (IES), award number R305A180401.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Open Research Statements

Study and Analysis Plan Registration

There is no study and analysis plan registration associated with this manuscript.

Data, Code, and Materials Transparency

The data, code, and materials underlying the results reported in this manuscript are not publicly available.

Design and Analysis Reporting Guidelines

Not applicable.

Transparency Declaration

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Replication Statement

This manuscript reports an original study.

References

- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111–127. <https://doi.org/10.2190/9LMD-3U28-3A0G-FTQT>
- Bergan, J. R., Sladeczek, I. E., Schwarz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal*, 28(3), 683–714. <https://doi.org/10.3102/00028312028003683>
- Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). *Cost-benefit analysis: Concepts and practice* (5th ed.). Cambridge University Press.
- Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J., Paltiel, A. D., & ISPOR-SMDM Modeling Good Research Practices Task Force (2012). Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, 32(5), 722–732. <https://doi.org/10.1177/0272989X12458348>
- Butler, A., & Woodward, N. (2018). Chapter One - Toward concision in the use of task-level feedback to promote learning. *Psychology of Learning and Motivation*, 69, 1–38. <https://doi.org/10.1016/bs.plm.2018.09.001>

- Byun, J., & Joungh, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science and Mathematics*, 118(3-4), 113–126. <https://doi.org/10.1111/ssm.12271>
- Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. *BioScience*, 58(10), 976–987. <https://doi.org/10.1641/B581011>
- 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., 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>
- Clement, J., Lochhead, J., & Monk, G. (1981). Translation difficulties in learning mathematics. *The American Mathematical Monthly*, 88(4), 286–290. <https://doi.org/10.2307/2320560>
- Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 245–252). Association of Computing Machinery.
- Cost Analysis Standards Project. (2021). *Standards for the economic evaluation of educational and social programs*. American Institutes for Research.
- Daugherty, L., Phillips, A., Pane, J. F., & Karam, R. (2012). *Analysis of costs in an algebra I curriculum effectiveness study*. (Report No. TR-1171-1-DEIES). RAND Corporation. https://www.rand.org/pubs/technical_reports/TR1171-1.html
- Decker-Woodrow, L. E., Mason, C. A., Lee, J.-E., Chan, J. Y.-C., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA Open*, 9, 23328584231165919. <https://doi.org/10.1177/23328584231165919>
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, 53(4), 533–548. <https://doi.org/10.1007/BF03395451>
- Dolonen, J. A., & Kluge, A. (2015). Algebra learning through digital gaming in school [Paper presentation]. Exploring the Material Conditions of Learning. The Computer Supported Collaborative Learning Conference 2015, In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), Volume 1. Gothenburg, Sweden. The International Society of the Learning Sciences. <https://doi.org/10.22318/cscl2015.232>
- Fletcher, J., Hawley, D. E., & Piele, P. K. (1990). Costs, Effects, and Utility of Microcomputer Assisted Instruction in the Classroom. *American Educational Research Journal*, 27(4), 783–806. <https://doi.org/10.3102/00028312027004783>
- Goldstone, R. L., Landy, D., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, 2(2), 265–284. <https://doi.org/10.1111/j.1756-8765.2009.01055.x>
- Harris, D. N. (2009). Toward policy relevant benchmarks for interpreting effect sizes combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3–29. <https://doi.org/10.3102/0162373708327524>
- Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4), 470–497. <https://doi.org/10.1007/s40593-014-0024-x>
- Hollands, F., Bowden, A. B., Belfield, C., Levin, H.M., Cheng, H., Shand, R., Yilin Pan, Y., & Hanisch-Cerda, B. (2014). Costeffectiveness analysis in practice: Interventions to improve high school completion. *Educational Evaluation and Policy Analysis*, 36(3), 307–326. <https://doi.org/10.3102/0162373713511850>
- Hollands, F. M., Hanisch-Cerda, B., Levin, H. M., Belfield, C. R., Menon, A., Shand, R., Pan, Y., Bakir, I., & Cheng, H. (2015). *CostOut - the CBCSE Cost Tool Kit*. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. www.cbsecosttoolkit.org
- Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness*, 9(1), 30–53. <https://doi.org/10.1080/19345747.2015.1055639>

- Hollands, F. M., Pratt-Williams, J., & Shand, R. (2021). *Cost analysis standards & guidelines, Module 1.1*. Cost Analysis in Practice (CAP) Project. <https://capproject.org/resources>
- Institute of Education Sciences. (2020). *Cost Analysis: A Toolkit* (IES 2020-001). U.S. Department of Education. https://ies.ed.gov/seer/cost_analysis.asp
- Jacob, M., & Hochstein, S. (2008). Set recognition as a window to perceptual and cognitive processes. *Perception & Psychophysics*, 70(7), 1165–1184. <https://doi.org/10.3758/PP.70.7.1165>
- Kellman, P. J., Massey, C. M., & Son, J. (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>
- Keltner, B., & Ross, R. (1996). *The cost of school based educational technology programs*. RAND.
- Kena G., Musu-Gillette L., Robinson J., Wang X., Rathbun A., Zhang J., Wilkinson-Flicker S., Barmer A., & Dunlop Velez E. (2015). *The condition of education 2015 (NCES 2015-144)*. U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>
- 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 Publisher.
- Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, 35(4), 224–257. <https://psycnet.apa.org/doi/10.2307/30034809> <https://doi.org/10.2307/30034809>
- 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*, 37(1), 68–76. https://doi.org/10.1007/978-3-642-17735-4_15
- Knuth, E., Stephens, A., McNeil, N., & Alibali, M. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, 37(4), 297–312.
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, 13(2), 129–164. https://doi.org/10.1207/s15327809jls1302_1
- Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 720–733. <https://doi.org/10.1037/0278-7393.33.4.720>
- Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In R. Light (Ed.), *New directions for evaluation* (pp. 55–68). Jossey-Bass. <https://doi.org/10.1002/ev.12>
- Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, 8(3), 400–418. <https://doi.org/10.1080/19345747.2014.915604h>
- Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). Sage Publications.
- Levin, H. M., Glass, G. V., & Meister, G. R. (1987). Cost-effectiveness of computer-assisted instruction. *Evaluation Review*, 11(1), 50–72. <https://doi.org/10.1177/0193841X8701100103>
- Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). Sage Publications.
- Li, W., Dong, N., Maynard, R. A., Spybrook, J., & Kelcey, B. (2021). *PowerUp!-CEA: A tool for calculating statistical power in multilevel randomized cost-effectiveness trials*. (Version 1.1) [Software]. <http://www.causalevaluation.org>
- 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. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In *Intelligent tutoring systems* (pp. 378–387). Springer International Publishing.

- Long, Y., & Alevan, 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>
- Marquis, J. (1998). Common mistakes in algebra. In A. F. Coxford & P. Shulte (Eds.), *The Ideas of Algebra, K-12: National Council of Teachers of Mathematics yearbook* (pp. 204–205). National Council of Teachers of Mathematics.
- McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of mathematical equivalence. *Journal of Educational Psychology*, 107(2), 423–436. <https://doi.org/10.1037/a0037687>
- 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>
- National Mathematics Advisory Panel [NMAP]. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education.
- NCES – NAEP. (2022). *Mathematics and reading scores of fourth- and eighth-graders declined in most states during pandemic, Nation’s Report Card shows*. <https://nationsreportcard.gov>
- NCES. (2021). *Table 411.60. Estimated average annual salary of teachers in public and secondary schools, by state: Selected years, 1969-70 through 2020-21*. National Center for Education Statistics. https://nces.ed.gov/programs/digest/d21/tables/dt21_211.60.asp
- Ottmar, E. R., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on mathematics learning. *Journal of the Learning Sciences*, 26(1), 51–78. <https://doi.org/10.1080/10508406.2016.1250212>
- Ottmar, E. R., Landy, D., & Goldstone, R. L. (2012). Teaching the perceptual structure of algebraic expressions: Preliminary findings from the Pushing Symbols intervention. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 2156–2161). Cognitive Science Society.
- Ottmar, E. R., Landy, D., Goldstone, R., & Weitnauer, E. (2015). Getting from here to there: Testing the effectiveness of an interactive mathematics intervention embedding perceptual learning. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127–144. <https://doi.org/10.3102/0162373713507480>
- Patsenko, E. G., & Altmann, E. M. (2010). How playful is routine behavior? A selective attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology*, 139(1), 95–116. <https://psycnet.apa.org/doi/10.1037/a0018268> <https://doi.org/10.1037/a0018268>
- Provasnik, S., Malley, L., Stephens, M., Landeros, K., Perkins, R., & Tang, J. H. (2016). *Highlights from TIMSS and TIMSS Advanced 2015: Mathematics and science achievement of U.S. students in grades 4 and 8 and in advanced courses at the end of high school in an international context* (NCES 2017-002). U.S. Department of Education, National Center for Education Statistics. <http://nces.ed.gov/pubsearch>
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, 27(4), 587–597. <https://psycnet.apa.org/doi/10.1007/s10648-015-9302-x> <https://doi.org/10.1007/s10648-015-9302-x>
- Schleicher, A. (2018). PISA 2018: Insights and interpretations. Organization for Economic Co-operation and Development Publishing. <http://rb.gy/ti2xdk>
- Schneider, M. (2020). *Making common measures more common*. Institute of Education Sciences. <https://ies.ed.gov/director/remarks/5-05-2020.asp>
- Schneider, M. (2020). *The value of cost analysis*. Institute of Education Sciences. <https://ies.ed.gov/director/remarks/8-31-2020.asp>
- Schneider, M. (2022). *Addressing the COVID learning crisis*. Institute of Education Sciences. <https://ies.ed.gov/director/remarks/09-13-2022.asp>

- Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, 47(6), 1525–1538. <https://psycnet.apa.org/doi/10.1037/a0024997>
<https://doi.org/10.1037/a0024997>
- Schoenfeld, A. H. (Ed.). (2007). *Assessing mathematical proficiency* (vol. 53). Cambridge University Press.
- Shand, R., & Bowden, A. B. (2022). Empirical support for establishing common assumptions in cost research in education. *Journal of Research on Educational Effectiveness*, 15(1), 103–129. <https://doi.org/10.1080/19345747.2021.1938315>
- Shapiro, J. (2013). *It only takes about 42 minutes to learn algebra with video games*. Forbes. <http://www.forbes.com/sites/jordanshapiro/2013/07/01/it-only-takes-about-42-minutes-to-learn-algebra-with-video-games/>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Siew, N. M., Geoffrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using DragonBox 12+ app. *Research Journal of Mathematics and Technology*, 5(1), 66–79.
- Speece, D. L., Molloy, D. E., & Case, L. P. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice*, 18(3), 147–156. <https://doi.org/10.1111/1540-5826.00071>
- Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., Lyskawa, J., McCallum, W. G., Porath, J., & Zbiek, R. M. (2015a). Teaching strategies for improving algebra knowledge in middle and high school students (NCEE 2014-4333). In *National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences*. U.S. Department of Education. <http://whatworks.ed.gov>
- Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015b). Learning from comparison in algebra. *Contemporary Educational Psychology*, 40, 41–54. <https://doi.org/10.1016/j.cedpsych.2014.05.005>
- 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://scholar.harvard.edu/contrastingcases/publications/learning-comparison-algebra-0>
<https://doi.org/10.1111/jcal.12347>
- Torres, R., Toups, Z. O., Wiburg, K., Chamberlin, B., Gomez, C., & Ozer, M. A. (2016). Initial design implications for early algebra games. In *Proceedings of the 2016 annual symposium on computer-human interaction in play companion extended abstracts* (pp. 325–333). Association for Computing Machinery CHI PLAY '16.
- Tsang, M. C. (1997). Cost analysis for improved policy-making in education. *Educational Evaluation and Policy Analysis*, 19(4), 318–324. <https://doi.org/10.3102/01623737019004318>
- U.S. Government. (2012). *U.S. government spending*. http://www.usgovernmentspending.com/us_education_spending_20.html
- Yeh, S. S. (2010a). The cost effectiveness of 22 approaches for raising student achievement. *Journal of Education Finance*, 36(1), 38–75. <https://doi.org/10.1353/jef.0.0029>
- Yeh, S. S. (2010b). The cost-effectiveness of NBPTS teacher certification. *Evaluation Review*, 34(3), 220–241. <https://doi.org/10.1177/0193841X10369752>

Appendix A

Table A1. Summary of HLM Model 3 predicting posttest scores ($N = 1,850$).

	Model 3	
	Coef	95% CI
Intercept	3.462***	[3.178, 3.745]
FH2T	0.361**	[0.092, 0.630]
Dragon	0.719***	[0.408, 1.030]
Immediate Feedback	0.254	[-0.037, 0.545]
Pretest	0.279***	[0.235, 0.323]
Asian	0.824***	[0.430, 1.219]
Other race	0.130	[-0.109, 0.370]
Male	-0.255**	[-0.409, -0.100]
Gifted	0.612***	[0.325, 0.899]
Accelerated	1.668***	[1.178, 2.158]
EIP	-0.183*	[-0.360, -0.006]
IEP	0.208	[-0.063, 0.479]
Virtual	0.415*	[0.048, 0.782]
Assign %	0.120***	[0.088, 0.153]

Note. Model 3: Condition: $\chi^2(3, N = 1,850) = 20.863, p < 0.001$; Race Effect: $\chi^2(3, N = 1,850) = 19.962, p < 0.001$.

Key: FH2T: From Here to There; Dragon: DragonBox; Immediate: Immediate Feedback; Pretest: Pretest score (centered); Asian: Asian; Other race: Race/ethnicity other than Asian or White/non-Hispanic; Male: Male; Gifted: Identified as gifted; Accelerated: Student in an accelerated mathematics program; EIP: Enrolled in Early Intervention Program; IEP: Has an Individual Education Program; Virtual: Enrolled in a virtual class; Assign: Number of assignments completed.

Source: Decker-Woodrow et al., 2023.