Education Tech Research Dev (2019) 67:423–441
https://doi.org/10.1007/s11423-019-09653-8

DEVELOPMENT ARTICLE

From here to there! Elementary: a game-based approach to developing number sense and early algebraic understanding

Taylyn Hulse[1] · Maria Daigle[1] · Daniel Manzo[1] · Lindsay Braith[1] · Avery Harrison[1] · Erin Ottmar[1]

Published online: 6 February 2019

Abstract
This paper examines whether using *From Here to There! (FH2T:E),* a dynamic game-based mathematics learning technology relates to improved early algebraic understanding. We use student log files within FH2T to explore the possible benefits of student behaviors and gamification on learning gains. Using in app measures of student interactions (mouse clicks, resets, errors, problem solving steps, and completions), 19 variables were identified to summarize overall problem solving processes. An exploratory factor analysis identified five clear factors including engagement in problem solving, progress, strategic flexibility, strategic efficiency, and speed. Regression analyses reveal that after accounting for behavior within the app, playing the gamified version of the app contributed to higher learning gains than playing a nongamified version. Next, completing more problems within the game related to higher achievement on the post-test. Third, two significant interactions were found between progress and prior knowledge and engagement in problem solving and prior knowledge, where low performing students gained more when they completed more problems and engaged more with those problems.

Keywords Early algebra · Game-based learning · Math achievement

Algebra is frequently called the gateway to college (Welder 2012) due to its correlation with high school and college graduation rates, as well as employment earnings (National Mathematics Advisory Panel (NMAP) 2008). However, by middle school, many students are falling behind in mathematics and struggle with learning new concepts, particularly algebraic ideas. What makes algebra so difficult to learn? For many children, an introduction to algebra may be the first time that math notation is presented as abstract. In the elementary grades, math is often taught concretely with tangible objects to aid in forming connections between math and the real world (Bruner et al. 1966). However, the abstract nature of variables and algebra allows fewer opportunities to make such tangible connections that were

✉ Erin Ottmar
erottmar@wpi.edu

1 Worcester Polytechnic Institute, 100 Institute Rd, Worcester, MA 01609, USA

possible with arithmetic (Booth et al. 2014). This jump from concrete to abstract in middle school is one reason why many students fall behind; many become disengaged and never master even the most basic algebraic concepts (Stein et al. 2011).

It is likely that the decline in algebraic performance and formal math understanding stems from both lack of exposure to algebraic concepts as well as misconceptions that develop early, in critical windows where students form the foundations of math understanding. Increasingly, empirical data advances the idea that early childhood math competencies are good predictors for later academic achievement (VanDerHeyden and Burns 2009). To better prepare students for future algebraic learning, some researchers suggest introducing algebraic concepts in early elementary school (National Council of Teachersof Mathematics (NCTM) 2000; Stephens et al. 2015). Children begin to develop the ability to reason algebraically even before they begin formal schooling (Doig and Ompok 2010). Developmentally, many students are certainly capable of being introduced to algebraic idea early, provided that it is scaled down to meet their skill level (Bay-Williams 2001; Carpenter et al. 2005; Carraher et al. 2006). By exposing children to algebraic ideas earlier, as students progress in their mathematical thinking, they may be better prepared to learn more difficult concepts down the road (Bransford and Schwartz 1999; Koedinger et al. 2008; NCTM 2000). As a result, national organizations have subsequently begun a push in elementary mathematics to provide early intervention programs and introduce early algebraic ideas as an initiative structured to improve student readiness in algebra, and decrease the number of students that are under-prepared (National Council of Teachers of Mathematics (NCTM) 2014). The motivation for the current study derives from this developmental perspective.

In this study, we examine the effectiveness of early introduction of number sense and algebraic principles using *From Here to There! Elementary* (*FH2T:E*), a dynamic game-based approach to mathematics instruction that has worked well for middle school children (Ottmar and Landy 2017; Ottmar et al. 2015). *FH2T:E* was developed to provide a technology-based, self-paced game that allows students to dynamically transform, manipulate, and decompose numbers and operations to grasp the most basic mathematical concepts necessary for success (Ottmar et al. 2012, 2015). Prior work has found positive gains in learning from *FH2T*:E compared to a business-as-usual control condition (Braith et al. 2017). In this paper, we dig deeper into these findings to explore possible predictors and moderators. Using the student log data created during mathematical problem solving, we reveal latent constructs of mathematical proficiency within the context of FH2T. This study addresses three research questions: (1) Are there differences in learning between the gamified and non-gamified versions of FH2T:E?; (2) Do in-app measures of student problem solving process predict learning gains?; and (3) Do certain student behaviors within *FH2T*:E differentially predict learning for high or low-knowledge students?

Early readiness: the importance of building algebraic understanding and number sense

The ability to use symbolic representation in mathematics represents an important developmental milestone for children as they advance from number sense to more abstract algebraic thinking (Carr et al. 2011). However, formal algebra is not typically introduced to young children until they enter middle school (Bay-Williams 2001). In early elementary school, math instruction is often centered around the recognition of patterns within numerical expressions and their ability to extend such rules/patterns to other math expressions (Lins and Kaput 2004). Students in elementary school often first begin learning math as numbers by creating concrete representations and often do not recognize the flexible potential or function of the numbers (Carr et al. 2011). By second grade, many children gain the capacity to cognitively represent such numbers and begin doing so through abstract reasoning of numbers and their relations to one another (Carr et al. 2011). More broadly, there is evidence from the strategy-instruction literature indicating that when students are provided with ample opportunities to practice new strategies and understand the effectiveness of these new strategies then they are able to acquire and use these strategies more independently (Bay-Williams 2001; Carr et al. 2011).

Research posits that the deficit in mathematical and algebraic understanding begins to arise as students enter the transitional shift between concrete representation of numbers and abstract conceptualization. This may occur due to a lack of understanding of number sense and the ability to see the flexibility and fluidity of expressions through operations (Kalchman et al. 2011). One of the most important developments in children's mathematical thinking is number sense, or flexibility in thinking about numbers (NCTM 2000). This involves being able to understand how to represent numbers in different ways, understand the size of numbers, and understand how different operations will impact the transformation of numbers (Sowder 1992). There are specific misconceptions and difficulties that students struggle with, namely, the overall understanding of order of operations, the use of parentheses within an expression, and the concept of equivalence (use of the equal sign) (Knuth et al. 2006; Welder 2012; Ottmar et al. 2012). For example, children often do not understand that parentheses also function as a multiplicative indicator as well as an organizational tool. As an example, the value of 18 may be written as 3×6 or 3(6). Importantly, 18 can also be written using multiple combinations of operations and symbols, like this: $(3+17) - 2$. Children who do not have a solid understanding of the order of operations would likely struggle to determine the appropriate order in which they could solve the expression, making complex math expressions that require multiple operations nearly unsolvable. These difficulties continue throughout schooling, with order of operations being noted as a major area of confusion for students learning algebra (Welder 2012).

Students also struggle with decomposition, or the ability to recognize and that any number can be broken down many combinations of other numbers (Clements and Sarama 2007). Decomposition as a math tactic is defined as the understanding that numbers are made of many different components, and may be rearranged in a way that makes the most sense to the student (Clements and Sarama 2007). When considering decomposition, students begin with a single number and are asked to explore its properties, for example, "*what two numbers can make 10?*" Inclination and tactics of decomposition are taught as early as kindergarten, as teachers see the meaningful action behind children understanding grouping, relationships, and patterns. Acting as a springboard for children's math understanding at an early age, decomposition is imperative to understanding more formal mathematical learning such as algebra. According to the NCTM Principles and Standards (2000), "students should be able to compose and decompose two-and three-digit numbers" by the second grade.

When students have a solid base in number sense and decomposition, they are more likely to be successful with algebra (VanDerHeyden and Burns 2009). Decomposition allows students to see various ways for them to approach problems (ex. $4\times6=24$ replace 6 to $4+2$ to see $4 \times (4+2)=24$, maintaining the same value about the equal sign). However, many times decomposition tasks only involve one operation and often this skill is not explicitly taught in relation to algebra despite its position as a fundamental algebraic concept (Clements 2000).

Equivalence in mathematics is also noted as a rudimentary foundation of algebra, and relies on strong quantitative skills fostered in early elementary mathematics teachings (Knuth et al. 2006). For instance, children when presented with $3+3=4+2$ instead of $3+3=6$ and $4+2=6$, may begin to understand the flexibility of the equivalence notation rather than view it as a rigid obstacle (i.e. the number to the right of the equal sign does not need to be the expressions definitive answer). The notation of equality and its role, is fostered in students understanding of the symbolism of equivalence, rather than as a directional symbol or one that separates problem from answer (Welder 2012). This understanding becomes critical in algebraic understanding as students must be able to correctly interpret equal sign and view its relation and equivalence (Knuth et al. 2006; Welder 2012). If provided early, exposure to not only decomposition of expressions and numbers, but also flexibility about the equal sign, may help increase mathematical understanding. Through the introduction and exposure of critical algebraic reasoning and fundamental concepts at earlier ages, children are provided the necessary tools to succeed in algebraic and future math conceptualization. Students who are successful in learning algebra progress through a series of conceptual steps that can be more precisely defined as number sense, representation, fact families, and (most importantly) decomposition (VanDerHeyden and Burns 2009).

It is upon this foundation that the learning of algebraic ideas is built in middle and high school. By following the natural development of number sense and cardinality, interventions that begin with building a solid foundation of number sense and the concrete properties of numbers may result in improved mathematical understanding. Several programs, such as *Building Blocks* (Clements and Sarama 2007), highlight the current push for earlier introduction and precedence placed upon the initial techniques and tools introduced in early math learning. Project LEAP (Blanton et al. 2015) takes a similar approach by focusing on the early introduction of topics within the algebra domain; subsequently, the purpose of Project LEAP's research is to address children's developmental understanding of algebraic topics and the relative impact on understanding by using specific early algebra learning techniques compared to traditional instruction (Blanton et al. 2015). Each of these programs are tailored around the fundamental belief that the practice of early algebra education is critical to success in later mathematics (Carraher and Schliemann 2007; Lins and Kaput 2004).

From here to there! (FH2T): a game-based perceptual learning intervention

The present study examines the benefits of early exposure to algebraic ideas in a game-based context. *From Here to There!* (*FH2T*) is a mathematics game that uses perceptual based interventions to introduce foundational algebraic concepts (Ottmar et al. 2015). *FH2T* is an intuitive program that relies on self-paced interaction and slowly introduces students to mathematical content through discovery-based puzzles that engage perceptual-motor systems (Ottmar et al. 2015; Ottmar and Landy 2017). This innovative game displayed on a web-based and/or touch-screen interface allows both physical and dynamic manipulations of the expressions by students, providing a powerful source of perceptual-motor experiences, which in turn may lead to increased acquisition of appropriate skills and operations (Ottmar et al. 2015). With this tool, instead of simply applying memorized procedures, students are able to directly interact with numbers as physical/tangible objects.

From Here to There was designed based on much work in cognitive science and math education that has found that integrating perceptual learning and technology is a valuable approach to teaching mathematics (Ottmar and Landy 2017; Cortes et al. 2015; Goldstone et al. 2010, 2011; Seitz and Watanabe 2005; Kalyuga 2009; Kellman et al. 2010). In *From Here to There!*, symbols are made into virtual objects that permit student manipulation to learn flexibility in numbers/expressions within the constraints of the natural laws of mathematics. It embodies student-based discovery-learning techniques by displaying numbers and symbols as physical objects and uses touch screen interfaces that permits fluid visualization aligned with appropriate cognitive content. Immediate feedback is provided to students as they attempt something that is mathematically invalid, but students are not allowed to commit to those mistakes. This helps make the innate structure of math symbols and expressions more explicit and visually determined (Ottmar et al. 2012, 2015). Physical manipulation of mathematical symbols has been deemed intrinsically engaging and offers a new yet natural way for children to understand symbolic constraints.

This approach is theoretically grounded in evidence supporting techniques of perceptual training to facilitate three fundamental algebraic perceptual processes (Ottmar et al. 2012, 2015). First, *symbols are treated as physical objects.* By envisioning the process as a fluid motion (including destruction, creation, and flexibility) the symbols become physical objects susceptible to manipulation. Second, *perceptual grouping affects mathematical performance* (Seitz and Watanabe 2005). When visualizing an expression, the placement of notation may lead the student to misunderstand the memorized standard order of operations. Third, *learning attentional tendencies is key to mastering mathematics.* This understanding comes from appropriate attendance to specific components of an expression (Carr et al. 2011; Seitz and Watanabe 2005; Welder 2012).Through this 'play-like' engagement with puzzle-based situations, the perceptual training provides students with the understanding of physical transformations and appropriate mathematical laws while creating a solid basis of understanding surrounding the decomposed properties and flexibility of numbers (Ottmar et al. 2012).

Originally intended for middle school students, the program has also been scaled down to be developmentally appropriate for elementary school students (*From Here to There!: Elementary (FH2T:E)*). The process of decomposition and promotion of early algebraic readiness is key to the design of *FH2T:E.* Each module presents a series of puzzles. Rather than simply solving the program, students are asked to make the given expression look like an equivalent expression that was specified in the goal (Fig. 1). To achieve this goal, students perform a series of dynamic interactions, including decomposing numbers ($8=5+3$ or $11-3$), performing operations to combine terms, rearranging terms to apply the commutative and associative properties, and adding terms to both sides of an equation. This promotes the essential algebraic skills of number sense and decomposition, as students must understand how to break apart and recombine numbers in order to progress. The unique features of *FH2T:E* such as its goal-state 'solution', provide a suitable environment for students to engage in trial-and-error decomposition and problem solving while remaining within the confines of natural math law. Beyond the game's intuitive and engaging nature, it also aligns with developmental trajectories of its target audience and provides an interactive and comfortable environment in which students may explore math topics. The game's universe-like module progression aligns with the Common Core standards and allows students to 'play' the game by slowly increasing in complexity through levels (i.e. subtraction, addition, order of operation). Preliminary results have established the

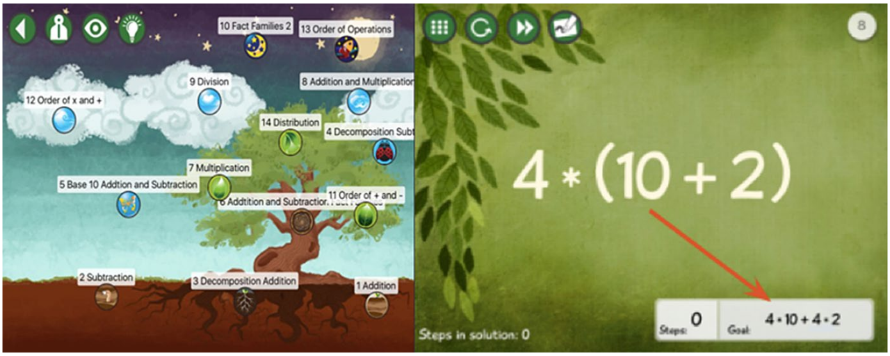

Fig. 1 *FH2T:E* World (left) and Problem View (right). The game includes 14 worlds, beginning with the most simple content (addition). Subsequent worlds become available as students progress through the game. Each problem has a designated puzzle-based "goal-state" where students rearrange, combine, and decompose the expressions to match the goal

feasibility of earlier introduction of pre-algebraic concepts and also demonstrated learning gains for second graders in comparison with children who did not interact with the game (Braith et al. 2017). However, it is unclear what components of the program relate to those learning gains (Fig. 2).

Examining student problem solving processes within games

One technologically driven instructional approach for education is game-based learning, where students can informally—and enjoyably—explore a topic. Games hold tremendous potential as mathematical instructional tools largely because of their ability to engage users while also having students practice mathematical concepts (Clark et al. 2016; Jere-Folotiya et al. 2014; Wouters et al. 2013) This higher level of engagement holds the potential to motivate students to practice problems, explore possibilities, and try new things in an engaging learning environment, which could prove particularly beneficial to those who typically struggle in traditional classroom settings (Kiili et al. 2015). Despite work suggesting that educational games may be useful for improving student engagement and math learning, when evaluating educational games, one must consider the possible learning mechanisms and behaviors that occur when students are interacting with the program. It remains to be determined whether the game-like features within a program (such as rewards and images) would be motivating or take away from learning and/or remain a superfluous detail neither adding nor taking away from the educational value of the game. It is also unclear whether the engaging aspects and learning benefits of FH2T result purely from the physical manipulation within the game, rather than the aesthetically pleasing features. Examining differences between a gamified version of *FH2T* and a non-gamified (plain, stripped down version of the game) allow us to address some of these questions.

In addition to providing potential engagement and learning gains, games can also provide researchers with important information, due to the rich data that can be collected while users are playing with the system (Gobert et al. 2013). Learning technologies not only have the potential to record the processing data needed for formative assessment, but can also provide students with immediate feedback, more

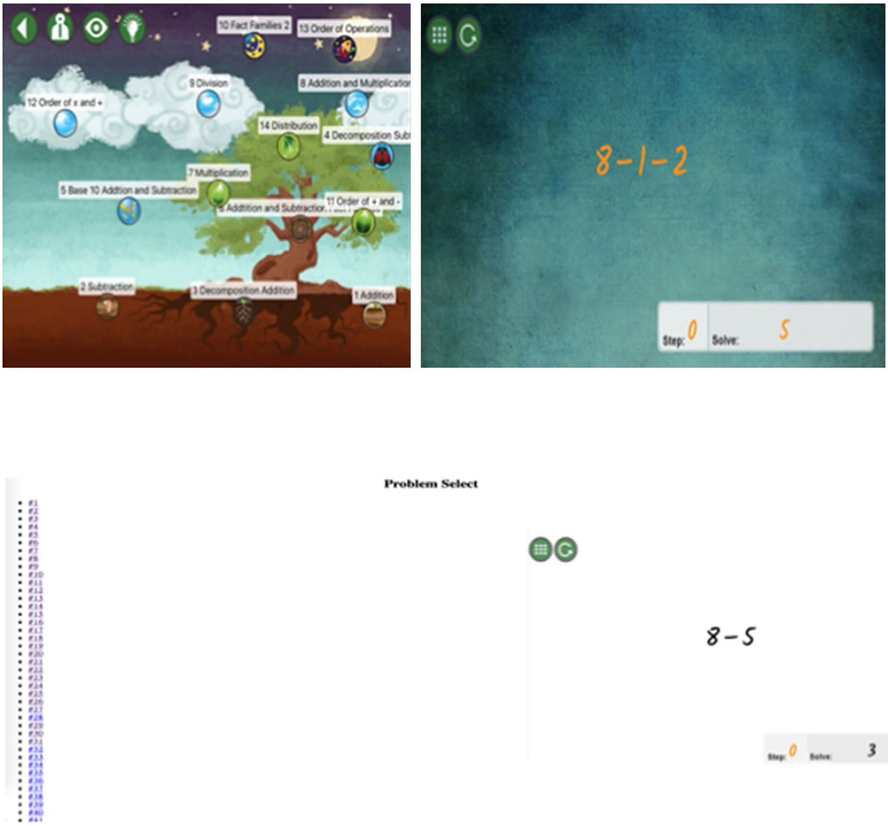

Fig. 2 *FH2T:E* gamified (top) and non-gamified (bottom) versions

individualized and self-paced learning, as well as more engagement through interactive content that rarely exists in more traditional forms of summative assessment (Kiili et al. 2015; Cayton-Hodges et al. 2015). Despite enormous advances in our understanding of teaching and learning mathematics over the past several decades, correctness is still the most commonly used measure to assess learning, most likely due to constraints such as time, cost, and feasibility in the classroom (Schoenfeld 2007). Logs from games provide us with unique innovative assessment features that measure the entire learning process of problem solving in real time (Chudowsky, and Pellegrino 2003; NRC 2001b). Through progress in using student log files to measure learning processes, far more information can be garnered in a shorter amount of time compared with traditional pencil and paper methods (Gee 2003; Gobert et al. 2013; Drasgow 2015; Shute 2011). For instance, when students solve math problems in a game format, information can be instantly gathered about their strategies, errors, and engagement. This information can then be used to predict gains in learning.

This work explores possible predictors and moderators that explain why students in a *FH2T:E* condition showed higher positive gains compared to a business-as-usual control condition (Braith et al. 2017). We examine how the behaviors and interactions within the *FH2T-E* game predict learning gains. We also examine whether interactions with the gain differentially benefit students with varying levels of prior achievement.

Method

Participants, experimental conditions, and procedures

One hundred eighty-five second grade students from ten classrooms in three different elementary schools in Massachusetts (116 female, 78 male) participated in this study. Students were randomzied into one of two experimental conditions: gamified versus non-gamified (explained below in materials). All students interacted with the app in their mathematics period during the school day for 20 min on four separate days, for a combined total of 80 min of play. This amount of time is both reasonable and practical for elementary students to practice and learn these concepts using typical classroom instruction. As part of the gamified condition, students played through the version of the game that possessed game-like features. Gamification. (i.e., any feature making the game more aesthetically pleasing rather than just the presence of math expressions) included the presence of levels, color, prizes, bonuses, stars, etc. The non-gamified version of *FH2T-E* was stripped down to display only the 18 math problems within each level. As students played through this plain version, there was no recognition of level completion or rewarded points for accuracy and efficiency. This lack of aesthetic features and reward-based prizes was intended to assess the degree to which the learning gains stemmed from the gamification features or the goal-state dynamic approach that the FH2T:E game provided. The math content and problems in each version was exactly the same. The only differences between conditions were the presence or absence of gamified visual material. Therefore, if differences in learning between conditions are statistically significant, results may highlight possible mechanisms by which *FH2T:E* leads to gains.

Measures

Data collected for this project included a combination of student scores on pre- and post-study worksheets and in-app data logs of the students interacting with the game.

Pre and post assessments

Prior to the introduction to the game, students completed a 15-item pre-study worksheet to assess prior math knowledge. These questions mirrored second grade math standards set forth by the Common Core (Common Core State Standards of Mathematics (CCSSM) 2010) and tested baseline understanding of decomposition, operational strategies, and basic notation. Completion of the pre-assessment was done 1 week before interaction with the game. A week after the four sessions were completed, students completed the post-study worksheet. The problems and expressions on the posttest were similar to those found on the pretest. To ensure baseline equivalence, an independent-samples *t* test was conducted to compare pretest scores for gamified and nongamified conditions. There was not a significant difference in pretest scores for the gamified ($M = 9.85$, $SD = 3.89$) and nongamified ($M = 9.95$, $SD = 3.60$) conditions; $t(183) = 0.17$, $p = 0.865$.

In app process data

As mentioned, *FH2T:E* has an extensive data logging system that records all student actions, mouse clicks and trajectories, errors, and moment-by-moment problem solving steps while interacting with the system. The reorded data was then compiled and aggregated across problems, levels, sessions and overall to create a series of variables that described problem solving processes. This paper uses the 34 overall variables to represent composite measures of student action and problem solving process in *FH2T:E* over the duration of the study.

An exploratory factor analysis, using Principal Axis Factoring, was then conducted to identify the number and structure of the factors underlying the overall data variables that were recorded within *FH2T:E* as students solved problems. Before conducting the factor analysis, all 34 of the initial variables were examined in a correlation matrix to test a few assumptions. It is recommended that all variables should be significantly correlated with at least one other variable (Tabachnick and Fidell 2007). It is also recommended that factors should not be correlated above 0.9, as that would violate assumptions of multicollinearity (Field 2009). There was only one factor that did not correlate with any others, *Star Score*. However, three groups of variables with correlations above 0.9: (1) *Extra Problems Completed* and (2) *Percentage Extra Completed, Average Time Per Problem, Problems Per Minute,* and *Average of Best Time*, as well as (3) *Distinct Problems Completed, Distinct Problems Unlocked, Percentage Problems Completed, Extra Problems Completed, Completed Stars* and *User Stars* (each highlighted in light gray). We chose to remove *Star Score* from analyses as it was an engineered measure from multiple other measures and did not correlate with any others. As for the groups of multicollinear variables, we decided to choose one variable from each group to represent the rest. We chose *Extra Problems Completed* to represent group 1 (about extra completed problems), *Average Time Per Problem* to represent group 2 (about time), and *Distinct Problems Completed* to represent group 3 (about overall completed problems). This left us with a total of 19 variables in the final analyses.

The KMO test values above 0.5 can be considered for EFA, with values above 0.9 considered as excellent (Hutcheson and Sofroniou 1999). Our KMO resulted in 0.751, which means our sample is adequate for producing reliable factors. The Bartlett test was significant $^{2}(171) = 5877.65$, $p < 0.001$, which means our correlations are significantly different from zero. With these considerations met, our sample was determined suitable for EFA (Table 1).

Using SPSS 22, Principal Axis Factoring was conducted using a Promax rotation. Promax was chosen as it is an oblique rotation that assumes the factors are correlated. Communalities describe the proportion of variance explained by the underlying factors and values above 0.5 are considered adequate for factor analysis. Communalities in our sample all resulted in values above 0.5. In fact, all variables except *Extra Problems* resulted in values above 0.9 (Table 2). Next, 5 factors were extracted using Kaiser's criterion (1958) criterion: that eigenvalues are greater than 1.00, that each factor be comprised of at least two factor loadings of > 0.40, and that the resulting components demonstrate good internal consistency.

Five factors with eigenvalues greater than 1.00 and sufficiently large loadings were extracted and they explained 29.74, 24.70, 21.29, 8.82, and 6.22% respectively, explaining a total of 90.78% of the variance (Table 3). The five factors are described in Table 2 and have been classified based on how variables loaded onto each factor. Factor 1, which included

Table 1 Structure coefficients from principal axis factor

Item	Engagement	Progress	Strategic flexibility	Strategic efficiency	Speed	Mean	SD
Total Go-Backs	0.935					25.41	49.42
Percentage of Attempts	0.908					2.00	0.91
Percent of Go-Backs	0.873					0.19	0.29
Number of Attempts	0.778					151.61	66.03
Overall Time Interaction	0.677					2601.70	868.83
Problems Completed		1.049				78.53	21.53
Completed Best Step		1.038				122.43	44.34
Extras Completed		0.774				18.06	7.24
Percentage of Resets			0.880			0.20	0.11
Ave Attempts Completed			−0.864			0.70	0.15
Total Resets			0.821			15.50	9.58
Average Resets			0.784			0.10	0.05
Ave Time-Step				0.834		7.84	2.65
User First Step				− 0.687		190.56	80.29
Percentage Stars				0.666		0.88	0.08
User Total Step				− 0.635		366.22	189.77
First Efficiency				0.554		2.67	1.81
Ave Time-Problem					0.999	27.48	9.96
Best Time					0.585	2077.05	1002.77
Eigenvalues	5.65	4.69	4.05	1.68	1.18		
Percent of Variance (%)	29.74	24.70	21.29	8.82	6.22	Total:	90.78

Table 2 Descriptive statistics and correlations

Factor correlations	1	2	3	4	5	6	7	8	9
1. Posttest score	–								
2. Pretest score	0.70**	–							
3. Gender	0.09	0.14	–						
4. Condition	−0.05	0.01	0.01	–					
5. Factor 1: engagement	0.21*	0.23	0.05	0.33**	–				
6. Factor 2: progress	0.27**	0.32**	0.05	0.26**	0.04	–			
7. Factor 3: strategic flexibility	0.12	0.06	0.07	−0.22**	0.16**	0.21**	–		
8. Factor 4: strategic efficiency	0.12	0.11	−0.07	−0.09	0.01	−0.23**	0.23**	–	
9. Factor 5: speed	−0.07	−0.08	−0.18*	0.04	−0.22**	0.51**	0.25**	−0.18*	–
Mean	74.23	65.91	0.53	0.61	0	0	0	0	0
Standard deviation	23.96	25.14	0.50	0.49	1	1	1	1	1

*p < 0.05; **p < 0.01

Table 3 Model results predicting post-test achievement

Parameter	Model 1	Model 2	Model 3	Model 4
(Constant)	27.95 (4.04)**	31.24 (4.34)**	33.91 (4.41)**	34.25 (4.36)**
Gender	−0.03 (2.61)	−0.37 (2.72)	0.52 (2.70)	0.54 (2.68)
PreTest % correct	0.68 (0.05)**	0.60 (0.06)**	0.58 (0.06)**	0.55 (0.06)**
Gamified	3.86 (2.61)	6.58 (3.16)*	5.51 (3.14)$^{+}$	8.39 (3.14)**
Engagement		1.90 (1.56)	15.17 (5.60)**	2.30 (1.53)
Progress		3.49 (1.81)$^{+}$	3.07 (1.80)$^{+}$	14.55 (4.09)**
Strategic flexibility		0.63 (1.58)	−0.74 (1.59)	0.22 (1.55)
Strategic efficiency		1.38 (1.48)	1.39 (1.46)	1.60 (1.45)
Speed		−1.48 (1.73)	−0.80 (1.72)	−1.28 (1.69)
Engagement × pretest			−0.17 (0.07)**	
Progress × pretest				−0.16 (0.05)**
F	57.54	22.45	21.25	21.93
R^2	0.50	0.52	0.54	0.55

Standard errors in parentheses

$^{+}p < 0.10$; *p < 0.05; **p < 0.01

Total Go-Backs, Percentage of Attempts, Percent of Go-Backs, Number of attempts, and *Overall Time Interaction* has been classified as *Engagement in Problem Solving*. This factor represents a measure of the number of problems solve: however, this measure does not represent greater progression through the app. students with higher scores on the go-backs factor were more likely to attempt and complete the same problems multiple times. Factor 2, which included *Distinct Problems Completed, Completed Best Step*, and *Extra Problems Completed* has been classified as *Progression*. For example, students with higher scores on this progression factor solved more distinct problems and progressed through the app more quickly. The distinction between factor 1 and factor 2 is important as it allows us to test whether it is simply practicing problems (attempting and completing the same problem more than once) or progression through the app (moving through the app and completing more unique problems) that is more beneficial for students. Factor 3, which included *Percentage of Resets, Average Attempts Completed, Total Resets,* and *Average Resets* has been classified as *Strategic Flexibility*. This represents a measure of how often students reset problems to try different approaches before successfully completing the puzzles. Factor 4, which included *Average Time Per Step, User First Step, Percentage Stars, User Total Step,* and *First Efficiency* has been classified as *Strategic Efficiency*. Higher scores for Factor 4 (strategic efficiency) represents using a minimal number of steps while solving problems. Finally, Factor 5, which included *Average Time Per Problem* and *Best Time* has been classified as *Speed*, a measure of student rate of solving problems. Correlations indicated that the 5 factors were also sufficiently independent of one another, indicating that they measure separate latent constructs.

Approach to analysis

First, descriptives statistics and correlations were calculated for each factor and variables. Next, four multiple regressions were conducted to examine relations between predictors and outcomes. The first model examined whether there were differences in performance between students in the gamified and non-gamified condition. Next, in model 2, the 5 latent in-app process measures were added into the analysis to explore which game behaviors contributed to learning. Our next step was to examine whether certain behaviors within *FH2T:E* mattered more for high or low performing students. In this study, we hypothesized that the two indictors of problem solving practice within the app (progression and engagement with problem solving) may vary depending on students prior knowledge levels. In model 3, we examined the interaction between progression and prior knowledge, while in model 4, we examined the interaction between engagement with problem solving and prior knowledge.

Results

Means, standard deviations, and correlations among the pretest, posttest, and latent factors are presented in Table 2. Pretest scores were correlated with higher completion ($r = 0.27$), higher go-backs ($r = 0.24$), and higher post-test scores ($r = 0.70$). Solving problems more quickly (time) was related to greater completion ($r = 0.37$) and fewer go-backs ($r = -0.25$). Results from all models are presented in Table 3.

Research Question 1: Our first aim was to determine whether there were differences in math posttest performance between students who received the gamified and non-gamified conditions. Results suggest that there were no differences in post test performance between the gamified and non-gamified conditions ($p > 0.05$), when only condition, gender and pretest performance were used to predict posttest performance.

Research Question 2: After including in-app student interaction components, a significant effect of condition emerged ($p < 0.05$). Students in the gamified condition performed, on average, 6.58 points higher on the posttest than students in the non-gamified condition. Further, progress (factor 2) was approaching significance ($p = 0.056$), suggesting that students who progressed faster and completed more unique problems in the app

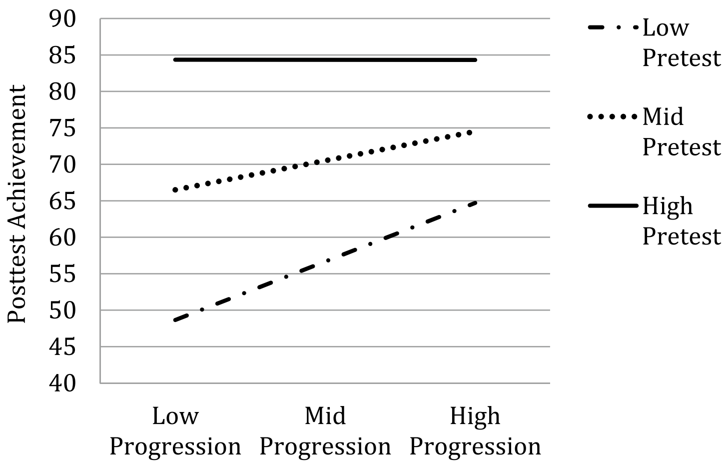

Fig. 3 Interaction of progression and prior knowledge on posttest achievement

may demonstrate higher posttest scores. More specifically, for every one standard deviation increase in completion, students performed approximately 3.07 points higher on the posttest. No other in-app measures predicted learning.

Research Questions 3 and 4: As displayed in Fig. 3, a significant interaction was present for Progress (factor 2) and prior knowledge. Students with lower initial pretest scores who completed more problems in the *FH2T:E* game demonstrated increased learning gains compared to students who completed less problems. However, posttest

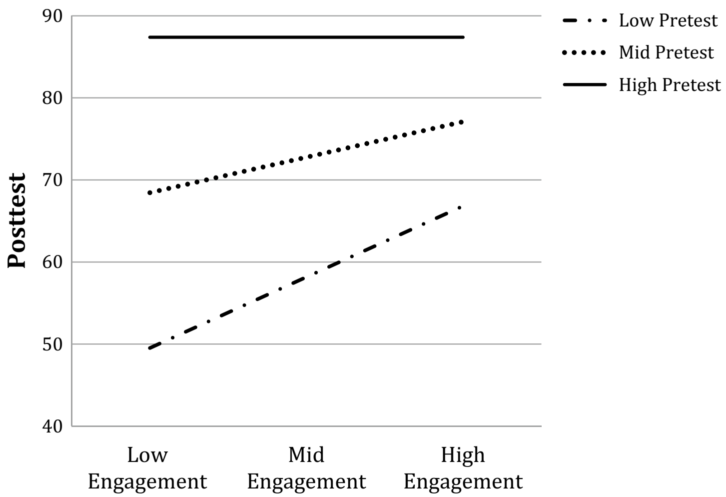

Fig. 4 Interaction of engagement with problem solving and prior knowledge on posttest achievement

achievement for initally high knowledge students was similar, regardless of the amount of problems students completed. A similar interaction and pattern emerged for Engagement with Problem Solving (factor 1, Fig. 4), with low knowledge students who engaged more with problems gained more than students who did not go-back and solve problems more than once. Engagement with Problem Solving did not seem to relate to achievement for high knowledge students.

Discussion

This study examined several factors related to student behavior and math learning within *From Here to There! Elementary*. Several main findings emerged from this study of second grade students. First, upon first examination, there did not appear to be significant differences in learning between gamified and non-gamified conditions. However, after accounting for in-app problem solving interactions, significant differences emerged, with students in the gamified condition being more likely to have larger gains on the posttest than students in the non-gamified condition. Next, solving more problems within the app could be related to higher achievement. Third, two significant interactions emerged, suggesting that solving problems within *FH2T:E* may be especially beneficial for low performing students: low performing students who solved more problems in the app and engaged in more behaviors in problem solving, including more attempts and going back to retry problems, were more likely to have larger learning gains than students who had initial higher levels of achievement.

After accounting for in-app behaviors, there is an advantage for gamification over non-gamification. Adding the support of gamified features may motivate students to engage with more difficult content that they have never learned before in a non-threatening environment. Furthermore, gamification may motivate these children to improve their problem solving strategies in order to receive rewards for the most efficient solution. However, it is important to note that efficiency and time were not significant predictors of mathematics learning. This is consistent with other work in mathematics education that values flexible problem solving process and thinking over speed and efficiency, even from the early years of mathematics instruction (Baroody 2003). Completing more unique problems and progressing further through the app was related to improved learning, providing additional evidence of learning benefits by engaging with and using the app. It may be that completing more problems provided more opportunities for learning by increasing exposure to different types of content and problems that young children may have never seen before, such as more complex opportunities for decomposition with multiple operations.

The significant interaction effects identify differences in the more subtle aspects of interaction with the program and addresses the question, *Who does FH2T:E help most?* Results suggest that playing with and completing more problems in *FH2T:E* appears to be more beneficial for low performing students compared to high performing students whose learning did not significantly change. This may be due to the fact that low performing students have more to gain in terms of learning and *FH2T:E* can give low performing students a valuable learning opportunity. One benefit of online math games is that students can progress through the app at their own pace, allowing lower performing students to continue to practice mathematical concepts and problems within a safe environment. Interestingly, it does not seem to matter if low performing students complete more unique problems that continue to progress them through the app or if students practice the same easier problems multiple times (repeated practice). Similar patterns of gains in achievement are observed for both types of problem solving practice for low and average performing students. These findings are consistent with other work examining the benefits of math apps that allow for differentiation of learners with varying achievement levels (Moyer-Packenham and Suh 2012), pointing to the importance of allowing students to re-do problems (attempts, go-backs) and solve math problems at their own pace. These findings are promising for using perceptually-guided puzzle-based problem solving as a means of decreasing the achievement gap between high knowledge and struggling students. Future studies should also address whether *FH2T:E* will benefit students with different demographic characteristics (racial, ethnic, linguistic, cultural, etc.) than those in the study population.

While we cannot definitively say why *FH2T:E* especially helped struggling students, it may be that the perceptual feedback, hints, and ability to reset and retry problems created new affordances for students that typically paper and pencil assessments does not provide. One plausible explanation may be that the puzzle-based design of the game was more motivating and engaging and less threatening for struggling students that the emphasis on correctness. Although we did not specifically measure math anxiety in this sample, these patterns are consistent with prior work in *FH2T* which suggested that students with higher levels of math anxiety and lower prior knowledge who engaged with *FH2T* solved more problems and did not experience detrimental effects of math anxiety on achievement compared to students who received more traditional instruction (Ottmar et al. 2015). Future studies should more closely examine the in-app data to compare the behaviors and relations between low and high performing students.

Implications for math teaching, research, and practice

These results suggest that it is feasible and productive to use games to support young students algebraic thinking through practicing early algebraic content, such as decomposition and order of operations. All students, regardless of their prior knowledge, were able to easily progress through the game. The flexible and accessible nature of the *FH2T:E* program supports the creation of new games in the future that can introduce physical interaction with content via a technological interface. From a preparation for future learning perspective, games might be especially effective because they can provide both motivation and learning gains while gradually exposing students to more difficult content and feedback within a supportive learning environment. The accessibility that web-based games provide may not only provide affordable opportunities for students to continue their math practice during the school year, but it may also serve as a promising intervention to bridge the gap over summer break when students often lose ground in content understanding.

Game-based learning technologies also have the potential to measure and assess student learning *during* the problem solving process. Though many instructional technologies have the ability to record all student interactions, there is little research on how these data can be used and mapped onto learning constructs of mathematical practice during instruction. This study is the first time that we have explored the predictability of new measures of in app interactions to assess mathematical learning within the *FH2T:E* game context. The additional information provided by this data revealed previously hidden effects of game-based components on learning. Following these findings, future research directions should include studies to expand and generalize the *FH2T-E* approach within this age range and to develop additional versions of gamed-based perceptual learning algebra interventions designed for even younger students (Clements and Sarama 2007; Lins and Kaput 2004).

Now that we have identified five factors that seem to reflect student interaction with the game, the next step is to validate these factors within a different data set. Once validated, we can more generally use these composite scores to predict learning, as well as create profiles of student behavior to better understand which students succeed and fail. This could begin to tease apart differences in age, prior knowledge, and engagement with the app and shed light on how students, despite differing starting points, could utilize *FH2T:E* to increase mathematical performance. Future studies could include outcome measures reflecting differences in student engagement, motivation or strategy obtained from the in-app data logged for each individual student's "game-session." Finally, within this in-app data, the *FH2T:E* program has the capability to analyze errant attempts made by students as they approach solving various items. Thus, it enables researchers to visualize both the effective strategies used by students and the errors and maladaptive approaches. This sort of data could be used to examine questions of mathematical flexibility and intervene earlier by providing immediate feedback and additional practice more effectively.

Conclusion

Overall, this study provides further evidence of efficacy for the *From Here to There!: Elementary* game on improving student mathematical understanding. By providing games that embed developmentally appropriate content and activities may make the introduction of early algebraic concepts into school classrooms more feasible and impactful.

Acknowledgments The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305A110060 to University of Richmond and Indiana University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors are also grateful to the many teachers and students who helped make this research possible.

References

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills* (pp. 1–34). Mahwah, NJ: Lawrence Erlbaum Associates.

Bay-Williams, J. M. (2001). What is algebra in elementary school? *Teaching Children Mathematics, 8*(4), 196–200. http://www.jstor.org.ezproxy.wpi.edu/stable/41197754.

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

Booth, J. L., Barbieri, C., Eyer, F., & Par-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *Journal of Problem Solving, 7*(1), 3.

Braith, L., Daigle, M., Manzo, D., & Ottmar, E. (2017). *Even elementary students can explore algebra! Testing the feasibility of From Here to There!, a game-based perceptual learning intervention.* Poster presented at the American Psychological Society Conference, Boston, Massachusetts.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61–100.

Bruner, J., Olver, R., & Greenfield, P. (1966). *Studies in cognitive growth: A collaboration at the center for cognitive studies*. New York: Wiley.

Carpenter, T., Levi, L., Franke, M., & Zeringue, J. (2005). Algebra in elementary school: Developing relational thinking. *Zentralblatt Für Didaktik Der Mathematik, 37*(1), 53–59. https://doi.org/10.1007/BF02655897.

Carr, M., Taasoobshirazi, G., Stroud, R., & Royer, J. M. (2011). Combined fluency and cognitive strategies instruction improves mathematics achievement in early elementary school. *Contemporary Educational Psychology, 36*(4), 323–333.

Carraher, D. W., & Schliemann, A. D. (2007). Early algebra and algebraic reasoning. *Second Handbook of Research on Mathematics Teaching and Learning, 2,* 669–705.

Carraher, D. W., Schliemann, A. D., Brizuela, B. M., & Earnest, D. (2006). Arithmetic and algebra in early mathematics education. *Journal for Research in Mathematics Education, 37,* 87–115.

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.

Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take? *Theory into Practice, 42*(1), 75–83.

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

Clements, D. H. (2000). 'Concrete' manipulatives, concrete ideas. *Contemporary Issues in Early Childhood, 1*(1), 45–60.

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education, 38,* 136–163.

Common Core State Standards Initiative (CCSSI) (2010). *Common Core State Standards for Mathematics.* Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Cortes, K. E., Goodman, J. S., & Nomi, T. (2015). Intensive math instruction and educational attainment long-run impacts of double-dose algebra. *Journal of Human Resources, 50*(1), 108–158.

Doig, B., & Ompok, C. (2010). Assessing young children's mathematical abilities through games. *Procedia - Social and Behavioral Sciences, 8,* 228–235. https://doi.org/10.1016/j.sbspro.2010.12.031.

Drasgow, F. (Ed.). (2015). *Technology and testing: Improving educational and psychological measurement*. Routledge.

Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: Sage.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 20.

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences, 22*(4), 521–563.

Goldstone, R., Landy, D., & Brunel, L. C. (2011). Improving perception to make distant connections closer. *Frontiers in Psychology, 2* 385.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science, 2*(2), 265–284.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Los Angeles: Sage.

Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J. C., Sampa, F., Yalukanda, C., Westerholm, J., et al. (2014). The effect of using a mobile literacy game to improve literacy levels of grade one students in zambian schools. *Educational Technology Research and Development, 62*(4), 417–436.

Kalchman, M. (2011). Using the math in everyday life to improve student learning: The Math in Everyday Life homework assignment builds student confidence and competence in mathematics. *Middle School Journal, 43*(1), 24–31.

Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. *Learning and Instruction, 19*(5), 402–410.

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.

Kiili, K., Devlin, K., Perttula, T., Tuomi, P., & Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games, 2*(4), 37–55.

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

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science, 32*(2), 366–397.

Lins, R., & Kaput, J. (2004). The early development of algebraic reasoning: The current state of the field. In K. Stacey, H. Chick & M. Kendal (Eds.), *The future of the teaching and learning of algebra: The 12th ICMI study* (pp. 47–70).

Moyer-Packenham, P. S., & Suh, J. M. (2012). Learning mathematics with technology: The influence of virtual manipulatives on different achievement groups. *Journal of Computers in Mathematics and Science Teaching, 31*(1), 39–59.
National Council of Teachers of Mathematics, Inc (NCTM). (2000). Standards for grades Pre-K-2. *Principles and standards for school mathematics* (Vol. 1, pp. 72–141).
National Council of Teachers of Mathematics. (2014). *Algebra as a strand of school mathematics for all students: A position of the national council of teachers of mathematics*. Retrieved from https://www.nctm.org/uploadedFiles/Standards_and_Positions/Position_Statements/Algebra_2014-04.pdf.
National Mathematics Advisory Panel (NMAP). (2008). https://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf.
National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. In J.W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.). Washington, DC: National Academies Press.
Ottmar, E., & Landy, D. (2017). Concreteness fading of algebraic instruction: Effects on learning. *Journal of the Learning Sciences, 26*(1), 51–78.
Ottmar, E., Landy, D., & Goldstone, R. L. (2012). *Teaching the perceptual structure of algebraic expressions: Preliminary findings from the pushing symbols intervention*. The Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society, 2156–2161.
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.
Schoenfeld, A. H. (2007). *Assessing mathematical proficiency* (Vol. 53). Cambridge: Cambridge University Press.
Seitz, A., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in cognitive sciences, 9*(7), 329–334.
Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction, 55*(2), 503–524.
Sowder, J. T. (1992). Making sense of numbers in school mathematics. In G. Leinhardt, R. Putnam, & R. A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* (pp. 1–46). East Sussex: Psychology Press.
Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research, 81*(4), 453–492.
Stephens, A., Blanton, M., Knuth, E., Isler, I., & Gardiner, A. M. (2015). Just say yes to early algebra! *Teaching Children Mathematics, 22*(2), 92–101.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.
VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education, 18*(1), 71–91.
Welder, R. M. (2012). Improving algebra preparation: Implications from research on student misconceptions and difficulties. *School Science and Mathematics, 112*(4), 255–264.
Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, Erik D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *American Psychological Association.*

Taylyn Hulse is a Ph.D. student in learning sciences and technologies at Worcester Polytechnic Institute. Her research is focused around the design and implementation of learning technologies in K-12 STEM classrooms.

Maria Daigle is an undergraduate pursuing a double major in psychology and biomedical engineering and minor in Spanish at WPI with research interests in educational and developmental psychology, neuroscience, and bioinstrumentation.

Daniel Manzo is a programmer, pedagogue, and researcher with interests in web applications, interactive media, serious games, and education. He is the founder of KnockOut Media and has authored numerous projects in these genres. He is currently pursuing a Ph.D. in Learning Sciences and Technologies at WPI.

Lindsay Braith received her undergraduate degree in psychological science from WPI in 2017. She is continuing her education towards a Masters Degree in Applied Behavioral Analysis and Assumption College with intention to sit for her BCBA exam in the summer of 2019.

Avery Harrison is a Ph.D. student in the Learning Sciences and Technology program at WPI. She is interested in research that explores the intersections of gesture, embodied cognition, and technology in K-12 mathematics education.

Erin Ottmar is an Assistant Professor of Psychology and Learning Sciences at WPI. Her research explores how cognitive and non-cognitive pathways, technology, and instruction combine to produce learning and growth for all children in K-12 mathematics classrooms.