From Here to There!: A Dynamic Algebraic Notation System Improves Understanding of Equivalence in Middle School Students

Jenny Yun-Chen Chan[a*], Ji-Eun Lee[a], Craig A. Mason[b], Katharine Sawrey[a], Erin Ottmar[a]

[a]Social Sciences and Policy Studies, Worcester Polytechnic Institute

[b]School of Learning and Teaching, University of Maine

* corresponding author: jchan2@wpi.edu

Author Contribution Statement:
Jenny Yun-Chen Chan: Conceptualization, Writing, Visualization, Project administration. **Ji-Eun Lee**: Data curation, Formal analysis, Writing, Visualization. **Craig Mason**: Formal analysis, Writing, Visualization. **Katharine Sawrey**: Conceptualization, Writing, Project administration. **Erin Ottmar**: Conceptualization, Writing, Supervision, Project administration, Funding acquisition

Acknowledgments
The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180401 to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We would like to thank the teachers and students for their participation; Dan Manzo, Avery Harrison, Hannah Smith, Taylyn Hulse, Kathryn Drzewiecki, Alisionna Iannacchione, and members of the Educational Psychology and Mathematics Learning Lab for their work; David Brokaw and members of Graspable Math Inc. for programming support; and Neil Heffernan and the ASSISTments Team for their support.

Disclosure Statement
Erin Ottmar is a founder and developer of Graspable Math, the technology supporting From Here to There!, and has declared a disclosure of interest.

PUBLISHED IN JANUARY 2021

Chan, J. Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2021). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000596

Abstract

Understanding equivalence is fundamental to STEM disciplines, yet misunderstandings and misconceptions inhibit students from fully appreciating or leveraging the concept. Using the game-based algebraic notation system, From Here to There! (FH2T), students explore ideas of equivalence by dynamically transforming expressions or equations among mathematically equivalent states. In the fall of 2019, 475 middle-school students participated in a randomized control trial where they worked in either FH2T or online problem sets with hints and feedback in ASSISTments over four 30-minute sessions during their math class. We found that (a) students in both conditions improved their understanding of mathematical equivalence from pretest to posttest, (b) students in the FH2T condition performed better on posttest compared to students in the problem set condition, and (c) the condition effect was comparable between students with high versus low prior knowledge. Together, the findings suggest that From Here to There! is an effective intervention for improving middle-school students' understanding of mathematical equivalence. The implications for research and practice on the usefulness of digital environments in mathematics education are discussed.

Keywords: Algebra and Algebraic Thinking, Technology, Instructional Activities, Middle School Mathematics

Educational Impact and Implications Statement

This study provides evidence that From Here to There!, a freely available game-based math technology, can improve mathematical understanding of equivalence in 6th and 7th grade students. From Here to There! integrates perceptual learning with mathematical puzzles, allows students to dynamically transform mathematical expressions and equations, and is designed to promote students' understanding of mathematical equivalence. Results reveal that FH2T improves learning for all students regardless of their prior knowledge, suggesting that it is a low-cost and effective math intervention for students.

From Here to There!: A Dynamic Algebraic Notation System Improves Understanding of Equivalence in Middle School Mathematics

Student misconceptions about equivalence and the equal sign have been noted as inhibiting success in upper-level mathematics and other STEM disciplines (Kieran, 2007; Knuth, Stephens, McNeil, & Alibali, 2006; Stephens et al., 2013; U.S. Department of Education, 2008). *Equivalence* is a foundational concept of mathematics involving the understanding that two mathematical objects, such as sets (e.g., 2 = 2), values (e.g., 2/3 = 4/6), or expressions (e.g., $(x - 1)(x + 1) = x^2 - 1$), represent the same value and are interchangeable. It is often formally represented by the equal sign, denoting that the two sides of an equation have the same value, so there is an equivalence relation between the expressions on each side of the equal symbol (e.g., 2 + 3 = 5; Kieran, 1981, 1992, 2007). While there has been an abundance of work that has explored interventions for improving students' understanding of equivalence in elementary grades (e.g., Alibali, Crooks, & McNeil, 2018; Blanton et al., 2015; McNeil et al., 2012), students continue to struggle with equivalence beyond elementary school years (McNeil et al., 2006). However, there is less work examining different approaches of training or improving students' conceptual understanding of equivalence at the middle school level.

In this study, we conduct a randomized controlled trial with 475 middle-school students, and test the effects of two different learning technologies, From Here to There! (FH2T) and an active control of online problem sets with hints and immediate feedback, on students' understanding of mathematical equivalence. In FH2T, we explore a novel approach of asking students to transform a perceptually different expression (e.g., $24 + y + 6 + 13$) to match a mathematically equivalent goal state (e.g., $13 + y + 30$). Instead of involving the equal sign, mathematical equivalence is implicitly embedded in the task, and students can experience the transformation of the starting expression changing into the goal state and the equivalent relation between the two expressions. Here, we compare student performance on mathematical equivalence items before and after the intervention to examine whether FH2T improves students' understanding of mathematical equivalence, and whether these effects vary based on students' initial levels of prior knowledge.

Mathematical Equivalence and Students' Struggles

Although children as young as four years of age have some understanding of numerical equivalence—whether two sets have equal quantities of items (e.g., Mix, 1999; Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011), students continue to struggle with the concept of equivalence well into college as the notation becomes more complex and involves larger numbers, more operations, and generalized forms with variables (Crooks & Alibali, 2013). One common misconception many students have is holding an *operational view*, or viewing the equal sign as calling for computation. This view leads students to interpret "2 + 3 = 5" as two and three *makes* five, or judge "2 + 3 = 4 + 1" as an invalid equation because it does not follow the typical "operation = answer" format (Kieran, 1981; Knuth et al., 2006). Furthermore, when an equation involves operations on both sides (e.g., 2 + 3 = 4 + __), some elementary students may add up all the numbers to generate the answer, __ = 9, instead of balancing two sides of the equation and fill in __ = 1 (e.g., McNeil & Alibali, 2005; Perry, Breckinridge Church, & Goldin-Meadow, 1988). When asked to compare two arithmetic expressions (e.g., 10 – 2 + 4 and 10 + 4 – 2), middle-school students were inconsistent at parsing the expressions (i.e., knowing that "–" should go with "2" in the expressions above) and some students incorrectly judged the equivalence of the two expressions, especially when they involved large numbers (Chaiklin & Lesgold, 1984). These misconceptions of arithmetic equivalence have negative impacts on learning algebra in which letters represent unknown values (Kieran, 1992; Küchemann, 1980) and students need to understand the relations between variables (Usiskin, 1988) and the structure of algebraic expressions (Kieran, 1989).

Together, these misconceptions about equivalence are associated with difficulty in equation solving and have ramifications for algebra and higher-level mathematics.

Past work has shown that students, especially in elementary school, struggle with mathematical equivalence (e.g., McNeil & Alibali, 2000), thus most of the intervention studies on mathematical equivalence are with elementary school students and focus on students' understanding of the equal sign (Alibali et al., 2018; Knuth et al., 2006; Rittle-Johnson et al., 2011). In one study with second- and third-graders, students who received intervention on practicing arithmetic problems organized by equivalent sums (e.g., 2 + 3 = 5, 1 + 4 = 5, etc.) showed more improvement on solving mathematical equivalence problems (e.g., 2 + 3 = 4 + __) compared to students who practiced arithmetic problems organized by the first addend (e.g., 2 + 3 = 5, 2 + 4 = 6, etc.) or students who did not practice solving arithmetic problems at all (McNeil et al., 2012). Furthermore, a sustained, comprehensive early algebra intervention involving relational understanding of the equal sign and principles of identity (e.g., 2 + 0 = 2), inverse (e.g., 2 – 2 = 0), and commutative (e.g., 2 + 0 = 0 + 2) properties in third grade leads to improvement in students' understanding of mathematical equivalence and equations (Blanton et al., 2015). The improvements in these studies reflect aspects of the *relational view* of the equal sign, where two sides of the equal sign calculate to the same value, and a particular expression is just one of myriad ways to represent that quantity (Stephens et al., 2013).

The importance of equivalence extends well into middle school and these concepts are critically important for success in pre-algebra and algebra (Fyfe, Matthews, Amsel, McEldoon, & McNeil, 2018; Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Kieran, 1989; McNeil et al., 2006). However, studies have found that almost half of middle-school students in the U.S. held an operational view of the equal sign (Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007; Booth & Davenport, 2013; Knuth et al., 2006), indicating the need for equivalence intervention in middle-school. Whereas most elementary students struggle with math equivalence and students across aptitude levels benefit from the equivalence interventions, the understanding of equivalence varies between middle-school students. Therefore, the differences in middle-school students' level of understanding may moderate intervention effects, resulting in aptitude-by-treatment interactions.

Previous studies have revealed two competing hypotheses on how students' prior knowledge may moderate mathematical learning. First, students with high prior knowledge may be more equipped to benefit from the instruction or intervention, and improve more on mathematical performance compared to students with low prior knowledge (e.g., Swanson, Jerman, & Zheng, 2008; Wood, Mazzocco, Calhoon, Crowe, & Connor, 2020). Second, students with low prior knowledge may have more room for improvement, and learn more from an intervention compared to students with high prior knowledge (e.g., Murphy et al., 2020; Ramani & Siegler, 2011). Understanding whether students with high or low prior knowledge gain more from an intervention is important to provide effective and efficient instruction that supports all students.

Given that developing a formal understanding of math equivalence is crucial to algebra learning, identifying effective interventions that promote these equivalence skills in students is critical. Previous interventions have focused on the meaning of the equal sign (that both sides of equations equal the same amount) and the principles of arithmetic operations around the equal sign. However, students also need to be able to encode the location of symbols and operators, notice the arithmetic relations between numbers within equations, recognize that quantities can be represented in many ways, and transform equations following mathematical principles. In the current study, we test the efficacy of a *perceptual learning intervention*—From Here to There!—on middle-school students' understanding of equivalence, and explore the potential interaction between prior knowledge and intervention on students' learning of mathematical equivalence.

Perceptual Learning and Conceptual Understanding of Equivalence

Substantial empirical work has demonstrated that perceptual processes are involved in learning (e.g., Catley & Novick, 2008; Goldstone, Landy, & Son, 2010; Patsenko & Altmann, 2010) and mathematics specifically (e.g., Kellman, Massey, & Son, 2010; Kirshner, 1989; Kirshner & Awtry, 2004; Landy & Goldstone, 2007). Perceptual features of symbols influence mathematics performance (e.g., McNeil, Uttal, Jarvin, & Sternberg, 2009), and humans adapt their perceptual-motor systems to organize these perceptual features to fit the needs in mathematical tasks. For instance, proximity is often used as a cue to group symbols in a way that aligns with mathematical principles. Adults tend to spontaneously group terms in ways that align with the order of operations: They write numbers around the multiplication sign closer together, and numbers around the addition sign further apart (e.g., $3\ +\ 4{\times}5$; Landy & Goldstone, 2007). When the symbols are spaced incongruent to the mathematical principles (e.g., $3{+}4\ \times\ 5$), even adults solve expressions incorrectly (i.e., add before multiplying in this example; Landy & Goldstone, 2010). These kinds of perceptual biases are thought to emerge with experience, and we rely on these perceptual processes to effectively and efficiently perform sophisticated cognitive tasks (Goldstone et al., 2010).

Since "what students notice mathematically has consequences for their subsequent reasoning" (Lobato, Hohensee, & Rhodehamel, 2013, p. 809), training one's perceptual and sensorimotor systems in symbolic notation may result in effective reasoning about the relationships represented by the symbols. Principles of grounded and embodied cognition suggest that successful perceptual training of algebraic structures engages cognitive systems that correctly embody mathematical rules and turn actions into meaning (Dourish, 2004). Grounding one's mathematical knowledge and reasoning in action and perception has also been shown to support the transfer of knowledge to new situations (Goldstone et al., 2010;

Goldstone, Landy, & Son, 2008; Landy & Goldstone, 2007b).

There is increasing evidence that utilizing perceptual features or engaging perceptual-motor systems in mathematical contexts can have a positive influence on learning (Kellman et al., 2010; Landy & Goldstone, 2007b; Ottmar et al., 2012; Ottmar et al., 2015). For instance, providing perceptual support by highlighting the equal sign in red leads to using new and correct strategies for solving mathematical equivalence problems in fourth-grade students (e.g., $2 + 3 + 4 = 2 + \_\_$; Alibali et al., 2018). Furthermore, students show improvement in simplifying algebraic expressions after a two-hour intervention that engages their perceptual-motor systems (Ottmar et al., 2012). In this intervention, eighth-grade students practiced simplifying algebraic expressions using gesture-actions (e.g., moving, combining, and substituting symbols) that apply dynamic transformations to expressions on screens. The system records all student actions, enacts valid transformations on the screen (e.g., turning $2 + 7x$ into $7x + 2$ when "2" is dragged to the right), and provides clear visual feedback on invalid transformations. As an example, in $2 + 7x$, when students attempt to add 2 and $7x$ by tapping "+", the expression shakes; when students attempt to transform $2 + 7x$ into $2x + 7$ by dragging "x" next to "2", "x" snaps back to its original location. Because x is an unknown in $2 + 7x$, adding 2 and $7x$ or moving x next to 2 are invalid, and the only valid actions the system enacts are commuting 7 and x around the multiplication sign (i.e., $2 + x \cdot 7$) or 2 and $7x$ around the addition sign ($7x + 2$). Following this intervention, students showed significant improvement on simplifying complicated expressions (e.g., $–5 + 4x + 2 – 6x – 8y + 2y$). Together, the findings suggest that learners leverage perceptual-motor features and feedback in mathematical learning and problem-solving, and that turning algebraic notations into tangible objects that enforce their own rules through physical movements may help improve mathematics learning.

Theoretical and Empirical Support for "From Here to There!"

FH2T (freely available online at https://graspablemath.com/projects/fh2t) is a dynamic research-based game application that implements perceptual learning theories to address cognitive and affective factors that lead to low proficiency in mathematics. While algebra instruction in school often focuses on memorizing and retrieval of abstract and arbitrary rules (Henry & Brown, 2008; Kirshner & Awtry, 2004), FH2T leverages and builds on students' knowledge of arithmetic for algebra learning. FH2T aims to help students identify how algebraic expressions are structured and think more flexibly about mathematical operations and properties, which in turn may improve students' proficiency and fluency in algebra (Ottmar et al., 2012). In particular, FH2T engages the perceptual-motor system to externalize the hierarchical structure of algebraic formalisms. In FH2T, the implicit structure of mathematics is made into explicit and interactive virtual objects so that students can touch and move symbols according to math principles in a virtual environment. By reifying math symbols as movable physical objects, students can realize that mathematical transformations are more dynamic, rather than procedural steps or a static re-copying of lines.

One of the important features of FH2T is that students learn algebra through discovery-based puzzles, rather than procedural steps. In each problem in FH2T, students are presented with two expressions: a starting expression, which is active and transformable, and a target goal state, which is mathematically equivalent to the starting expression (See Figure 1). Students' objective is to transform the starting expression ("$24 + y + 6 + 13$") into the target goal state ("$13 + y + 30$" in a white box) using algebraically permissible actions and learned gestures. It is important to note that there is no equal sign linking those two expressions; however, equivalence is an integral aspect of the game that the starting expression (and any subsequent state of the active expression) is mathematically equivalent to the goal state, and the transformation process demonstrates this equivalence. For instance, adding 24 to 6 transforms "$24 + y + 6 + 13$" into "$y + 30 + 13$", and commuting 13 to the left transforms "$y + 30 + 13$" into "$13 + y + 30$". Equivalence was preserved in *all* steps between the starting expression ($24 + y + 6 + 13$) and the goal state ($13 + y + 30$). In FH2T, students apply and build upon their arithmetic knowledge (e.g., $24 + 6 = 30$) to efficiently reach the goal, receive perceptual feedback on their actions, and uncover transformation paths between the two equivalent expressions.

Another important feature of FH2T is that students can dynamically manipulate and transform mathematical expressions by using various gesture-actions on the screen (e.g., dragging, tapping) to perform operations, break apart parts of the equation, and reach the goal state of the problem. Learning technologies, particularly dynamic systems that utilize motion, may offer a promising new approach to teaching mathematics that is not possible with traditional instruction (Arzarello, Bairral, & Dané, 2014; Byers & Hadley, 2012; Mcewen & Dubé, 2015; Sinclair & Heyd-Metzuyanim, 2014). Turning algebraic notations into tangible, movable objects that follow mathematical principles shows promise for transforming many of the traditional distinctions between abstract and concrete knowledge (Alibali & Nathan, 2012; Kaminski, Sloutsky, & Heckler, 2008; Uttal, Gentner, Liu, & Lewis, 2008). Indeed, preliminary evidence suggests that the use of dynamic symbols in a game-based environment can increase students' engagement and learning of algebraic concepts (Ottmar & Landy, 2017; Siew, Geofrey, & Lee, 2016). The program also responds to student actions, and the student receives immediate feedback on the validity of their actions. From a perceptual learning perspective, the experience of moving and transforming algebraic objects on the screen, reinforced by the visual feedback of changes to the expressions, may help students generalize notation mechanics and attend to relevant details.

Prior work using an earlier version of FH2T has shown positive results that the system may be effective for decreasing structural errors (e.g., incorrectly adding the 3 and 5 to make 8 in $3 + 5 \times 4$) and improving mathematical understanding for elementary and middle school students. For instance, one preliminary study with 85 sixth- and seventh-grade students showed that those in the condition similar to FH2T, referred to in the study as the fluid visualizations condition, experienced more gains on mathematics than students in both the manual calculations and control conditions (Ottmar et al,, 2015). Moreover, the students who completed more problems in the fluid visualizations condition scored higher on the posttest than students who completed fewer problems in the app (Cohen's d = .48), and this effect was significant above and beyond students' prior knowledge.

Similarly, another study with 185 second-grade students (Hulse et al., 2019) also found that, controlling for pretest performance, students who completed more problems within the game scored higher on the posttest compared to students who completed fewer problems. A further investigation revealed a significant interaction between in-game progress and students' prior knowledge on the posttest performance. Among students with low prior knowledge, those who solved more problems were more likely to have larger learning gains compared to those who solved fewer problems; among students with high prior knowledge, in-game progress was not related to posttest scores. The findings suggest that solving more problems in the game was more beneficial for students with low, as opposed to high, prior knowledge. In sum, the prior evidence suggests that FH2T may prove effective at improving mathematics performance in elementary and middle-school students, and perhaps be more beneficial for students with low prior knowledge, warranting a larger-scale randomized controlled trial to evaluate its effectiveness in middle school.

The Current Study

We conducted a pretest-intervention-posttest randomized controlled trial in the fall of 2019, where students were assigned to one of two intervention conditions: FH2T and an online problem set control with hints and immediate feedback in ASSISTments (Heffernan & Heffernan, 2014). ASSISTments is an online assignment platform where teachers can monitor student performance and progress, and students can request hints during problem-solving and receive immediate correctness feedback on their answers. A previous randomized controlled trial has revealed that ASSISTments was beneficial for student learning and increased students' mathematics performance compared to business-as-usual homework practice (Roschelle, Feng, Murphy, & Mason, 2016). As such, in this study, we used problem sets in ASSISTments as an active comparison group, rather than a true business-as-usual control, to examine the effects of playing FH2T on students' understanding of mathematical equivalence above and beyond an effective technology-based intervention (What Works Clearinghouse, 2020). Our specific research questions were:

(1) Do students make gains in understanding of mathematical equivalence after a two-hour intervention?
(2) Do students in the FH2T condition show a greater understanding of mathematical equivalence at posttest compared to students in the active problem set control condition?
(3) Does the intervention effect vary depending on students' prior knowledge?

To address our research questions, we first conducted a preliminary series of paired sample t-tests to examine gains from pretest to posttest (RQ1); we then used hierarchical linear modeling (HLM) to estimate the effects of condition on posttest scores, controlling for pretest and other student characteristics, as well as classroom-level nesting (RQ2); and we explored the interaction between condition and pretest scores, using HLM (RQ3). We hypothesized that students in both conditions might experience gains from pretest to posttest, but students in the FH2T condition may show a greater understanding of mathematical equivalence at posttest compared to students in the problem set condition. We explored the potential moderating effect of students' prior knowledge on the effects of intervention in order to examine whether students with high or low prior knowledge benefit more from the intervention, but we did not have an a priori hypothesis regarding the direction of the interaction.

Methods

Participants

Ten teachers from six middle schools were recruited from a large, urban district in the Southeastern United States. Together, they taught 29 mathematics classes with a total of 689 students. Most students were in sixth grade (609 students, 88.4%), and the remaining 80 students (11.6%) were in seventh grade. All students in our study were placed in one of three levels of mathematics classrooms by the district: advanced (525 students, 76.2%), on-level (111 students, 16.1%), or support (53 students, 7.7%). The majority of students in our sample were advanced sixth grade students. Random assignment of the intervention condition occurred at the student-level, with the 689 students from the 29 classes randomly assigned to FH2T (n = 348) or problem set (n = 341) conditions.

Due to scheduling constraints, 19 students from one classroom did not participate in the study and 183 students did not complete at least 50% of the items on the pretest or posttest, thus, we excluded these students in the following analyses. The 50% cutoff on pretest and posttest was determined during preliminary analysis prior to testing the intervention effects. The students who completed less than 50% of items spent an average of 9 minutes on the pretest (Range: 0–12 minutes out of 45 minutes), and 2 minutes on the posttest (Range: 0–13 minutes out of 40 minutes), suggesting that they did not spend appropriate amount of time and dropped out of the assessments. Using the 50% cutoff on pretest and posttest allowed us to only include students with whom we had more accurate estimates of their equivalence understanding and learning. Additionally, we obtained access to data on student characteristics from the school district, and excluded 12 students who had missing demographic and past achievement information (i.e., gender, race, overall academic achievement status). The total number of excluded participants was 214 (FH2T: 121; problem set: 93). The final sample included the remaining 475 students: 227 (47.8%) were in the FH2T condition, and 248 (52.2%) were in the active control problem set condition.

The student demographics and pretest scores of the final sample ($N = 475$) were comparable between conditions (Table 1). The study was approved by and conducted in accordance with the human subjects guidelines of the Institutional Review Board.

Procedure

This study consisted of a 45-minute pretest, four 30-minute intervention sessions, and a 40-minute posttest in a span of six weeks. Teachers were asked to have their students complete one or two sessions a week, and the session assignments for that week were made available to the teachers and students each Sunday. Individual teachers decided the days that their students would work on the study assignments in class.

In Week 1, all students received a brief assessment on their equivalence understanding, an experimenter-designed task that measures students' sensitivity to perceptual differences in algebraic expressions, and questionnaires on their mathematics anxiety and mathematics self-efficacy. Students were then assigned two intervention sessions in Week 2 and another two intervention sessions in Week 3. The mathematical content was aligned between the two conditions, and all students solved problems involving four operations, negative numbers, fractions, and order of operations during their intervention sessions using their assigned technology. The posttest was given on Week 4, and it consisted of the mathematical equivalence assessment, the experimenter-designed task, and the questionnaire on mathematics anxiety. All study assignments remained available to students and teachers until the end of Week 6, allowing students additional time to complete any outstanding assignments.

All study assignments were administered online in mathematics classrooms during instructional periods, and students worked individually at their own pace using their own school-issued Chromebooks in the classroom. Although the pretest and posttest were designed to take approximately 40 minutes, students could take as long as they needed to complete the assessment and the questionnaires. A countdown timer was embedded in the FH2T and problem set conditions to ensure that students in the two conditions were allotted a comparable amount of time with each learning technology.

The assessments and all study sessions in both conditions were delivered to students using ASSISTments (Heffernan & Heffernan, 2014), an online platform that allows teachers to assign problem sets to students. To be clear, we used the ASSISTments platform in two different ways in this study: a platform to implement the randomized controlled trial, and as the technology used in the active problem set control condition. ASSISTments was designed to be used not only as a homework tutoring and feedback system for teachers and students, but also as a tool to help researchers efficiently conduct randomized controlled trials, collect finely grained data about student interactions, and analyze and share results within an existing data infrastructure (Heffernan & Heffernan, 2014). Thus, we used ASSISTments as the platform for administering pretest and posttest, providing instructions for the intervention sessions, maintaining intervention condition within students across sessions, and recording timing and fidelity data for *both* conditions.

While it may appear that using the system in this dual way may unfairly benefit students in the problem set control condition, we feel that any advantage the control students may have makes it an even stronger comparison condition for testing FH2T. That is, if the familiarity and experience of the technological environment contribute to student performance on posttest, students in the problem set control condition should outperform students in FH2T condition. However, if we detect positive effects for FH2T above and beyond the active control, the findings provide clear evidence that the improvement may be due to aspects of FH2T experience and cannot be attributed to students' familiarity with the learning technologies.

The Measure of Mathematical Equivalence Understanding.

Although the pretest and posttest also included several measures on perception of algebraic expressions and attitude toward mathematics, the current analyses focused on the mathematical equivalence assessment consisting of six items from previously validated measures of performance (Rittle-Johnson et al., 2011: Cronbach's α = .94 -.95; Star et al., 2014; Cronbach's α = .89). Questions are listed in Table 2. Within the six items, Items 1 and 2 focused on balancing two sides of the equation; Item 3 assessed students' definition of the equal sign; Item 4 involved arithmetic operation on both sides of the equal sign; Items 5 tapped into students' understanding of the relations between addition and multiplication in equivalent expressions; and Item 6 involved transforming equivalent expressions by applying order of operations and distribution. Although limited in measuring students' comprehensive understanding of algebra, these six items together assessed a range of students' conceptual knowledge in mathematical equivalence. Isomorphic posttest questions were created by substituting numbers of similar magnitudes in the pretest questions and the response options. Each item was scored as correct (1) or incorrect (0), and the total score out of 6 on the posttest was included as the outcome and the pretest score was included as a covariate in the analyses. The reliability of these six items was KR-20 = 0.63 at pretest and KR-20 = 0.64 at posttest. Since the reliability on these assessments were within acceptable range but lower than the preferred 0.80, likely owing to the limited number of items and binary scoring, we reported descriptive findings of the individual items to further explore how the intervention may impact students' understanding of equivalence.

The assessments were administered using the test-mode in ASSISTments, where hints and correctness feedback were not available during or after the assessments. The questions were presented one at a time, and students entered their answers via the keyboard or selected a response option using a mouse. After entering or selecting an answer, students clicked the "Submit Answer" button for the system to record their response, and then clicked the "Next Problem" button to move on to the next question (Figure 2).

Intervention Conditions

In both conditions, students accessed their assigned learning technology through the ASSISTments platform. Each day, students logged in to the platform using a username and password. The assignment of the day was presented using a clickable link that directed students to their assigned learning technology—further work in ASSISTments or FH2T. A countdown timer was embedded into each intervention session to ensure that both conditions were matched on time.

From Here to There! (FH2T)

As described earlier, FH2T is a research-based technology game, where students transform mathematical expressions from an initial state to a specified mathematically equivalent goal state. In FH2T, numbers and mathematical symbols become virtual objects that students can pick up and move. Students acquire new gesture-actions (e.g., using a mouse cursor to tap the addition sign to add) through brief video demonstrations, and use these gesture-actions to transform expressions from one state to another.

A sample FH2T problem with a series of gesture-actions to the goal state is illustrated in Figure 3. In this example, the objective is to transform 7 + 2 + 10 + 8 into 5 + 2 + 5 + 15 (Figure 3a). The student first dragged the 7 on top of 8 (Figure 3b) to produce the sum of 15 (addition; Figure 3c). Then, the student turned on the keypad (Figure 3d) and selected 10 (Figure 3e) to substitute 10 with 5 + 5 (decomposition; Figure 3f). Last, the student dragged a 5 to the left to commute it with 2 (commuting; Figure 3g). When the active expression matches the goal state (Figure 3h), a clover board appears showing the number of clovers awarded based on the number of steps taken to solve the problem (Figure 3i). Students received more clovers for solving the problem using fewer steps. It is important to note that the problems do not simply ask students to click and solve for the answer. Rather, students need to attend to the similarities and differences between the initial and the goal states in order to successfully and efficiently complete the problem.

Beyond combining terms and dragging to commute, gesture-actions in FH2T allow users to enact most forms of symbolic manipulation, including each of the four basic operations, number decomposition, distribution, factoring, and properties of equality (e.g., performing arithmetic operations to both sides). The system also provides immediate feedback on invalid mathematical transformations. When students make a mathematical error or attempt an invalid transformation (for example, trying to turn $2 + 7x$ into $9x$ or $5 + 3 \times 4$ into 8×4), the expression automatically shakes and snaps back to the starting expression, which signifies that it was an invalid mathematical transformation, without indicating the correct action.

FH2T consists of 14 worlds that focus on different mathematical concepts, and each world contains 18 problems (a total of 252 problems). Students start from simple topics and build up their knowledge and skills throughout the game. In this study, all students in the FH2T condition started from World 1: Addition, and worked their way through the game (World 2: Multiplication, World 3: Order of Operations + and ×, World 4: Subtraction and Negative Numbers, World 5: Mixed Practice of + and −, and World 6: Division, World 7: Order of Operations, World 8: Equation + and −, World 9: Inverse Operations + and −, World 10: Distribution, World 11: Factoring, World 12: Equation +, −, ×, and ÷, World 13, Inverse Operations, World 14: Final Review).

All students were given 30 minutes to play FH2T for each session. After 30 minutes, the system would log students out of the game and save their progress. When students returned by clicking the assignment link in each subsequent session, they were able to start where they left off. Because students worked through the problems on their own devices at their own pace, the progress within the game varied across individuals. On average, students solved 104.7 distinct problems (*SD* = 31.96, Min. = 31, Max. = 173) in FH2T across four intervention sessions. Among the 227 students in the FH2T condition, about half (48.9%) reached World 8 (Equation + and −) or higher; only 19 students (8.4%) reached World 11 (Factoring) or beyond. Mapping to the assessments, Worlds 1 (Addition), 2 (Multiplication), and 3 (Order of Operations + and ×) correspond to the concepts tested in Items 1, 2, 3, and 5; Worlds 4 (Subtraction and Negative Number) and 8 (Equation + and −) correspond to Item 4; Worlds 7 (Order of Operation) and 10 (Distribution) correspond to Item 6.

Problem Sets with Hints and Immediate Feedback in ASSISTments

Students in the problem set condition solved traditional mathematics problems in ASSISTments, a free online tutoring system for homework and problem-solving (Heffernan & Heffernan, 2014). Problem sets in ASSISTments were selected as an active control condition because it bookends different aspects of FH2T. Problem sets in ASSISTments cover mathematical content well-aligned with traditional instruction and offer hints and immediate feedback (e.g., the correctness of the answers) to students. However, unlike FH2T, the system does not include perceptual learning features so that the problems are presented with a static textbook format. Moreover, it does not have game design elements, such as rewards, level, and challenge, to motivate student learning.

The problems in ASSISTments were selected and adapted from three open-source middle-school mathematics curricula: Utah Math Project (2016), Illustrative Mathematics (2017), and Engage NY (2014), so that the problems aligned with traditional instruction and the topics covered in FH2T. The topics covered in the problem sets included: addition and multiplication, subtraction and negative numbers, division and fractions, and order of operations. Four problem sets each consisting of 24 to 39 questions were developed for the four mathematical topics in this study, and students started on the first question of a problem set at the beginning of each intervention session. On average, the students answered 115 questions (*SD* = 17.47, Min. = 25, Max. = 129) in the problem set condition across four intervention sessions. Among the 248 students in the problem set condition, 245 students (98.8%) completed all problems in Assignment 1: addition and multiplication, 237 (95.6%) completed Assignment 2: subtraction and negative number, 229 (92.3%) completed Assignment 3: division and fraction, and 236 (95.2%) completed Assignment 4: order of operations (Note that students did not have to complete the previous assignments to move forward). Mapping to the assessments, Assignment 1 (addition and multiplication) corresponds to the concepts tested in Items 1, 2, 3, and 5; Assignment 2 (subtraction and negative number) corresponds to Item 4; Assignment 4 corresponds to Item 6.

A variety of task and answer formats were used. The task formats included computation, word problems, and representation interpretations; the answer types included numbers, algebraic expressions, multiple choice, and open response. The questions were presented one at a time (Figure 4a), and the majority of the answer types were numbers, algebraic expressions, or multiple choice that were automatically graded. Students had the opportunity to request three hints and the final solution at any time (Figure 4b). Students received immediate correctness feedback after each answer submission (i.e., “Correct!” or “Sorry, try again. [Student answer] is not correct.”), and they had to enter the correct answer to move on to the next problem. If students answered the problem correctly on the first attempt without requesting any hints, they received a green check on the problem. If students attempted the problem multiple times or used hints, they received an X mark on the problem.

A green X was given if students answered a problem correctly after using one to three attempts or hints. A red X was given if students answered a problem correctly after using more than three attempts or hints. A red X with a yellow box was given if students requested all hints and the solution. Less than 15% of the questions were in the open-response format and they were included to gauge students' reasoning during problem solving. Students did not receive correctness feedback or hints on these questions. Instead, students saw the "Answer Recorded" message when they submitted their responses, and moved on to the next question. The open responses were recorded for later analyses but not automatically graded by the system.

Students solved traditional mathematics problems in ASSISTments for 25 minutes and then were directed to their assignment report to review their performance on each problem for the remaining 5 minutes (Figure 4c). Students were instructed to spend as much time as needed to review the problems and their answers. It is important to note that although ASSISTments provides additional functions for teachers to build their own problems, assign problem sets, monitor student progress, and review student performance, the current study only utilizes the student focused features, including hints, immediate correctness feedback, and performance reports in ASSISTments.

Results

Descriptive statistics, correlations, t-tests, and standard multiple regressions were conducted using IBM SPSS Statistics version 25. Hierarchical linear models were conducted using HLM version 8.0 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2019). Means, standard deviations, minimum and maximum values, and correlation coefficients for all variables are presented in Table 3. Pretest scores ranged from 0 to 6, with a mean of 3.80 (SD = 1.61); posttest scores also ranged from 0 to 6, with a mean of 4.14 (SD = 1.56). Although the pretest and posttest scores were somewhat skewed (pretest = –0.31, posttest = –0.58), the entire range was reflected in the data. At pretest 23% of students scored two or below, 38% scored three or four, and 39% scored five or six; at posttest, 16% scored two or below, 36% scored three or four, and 47.4% scored five or six. Because the pretest and posttest scores were distributed throughout the entire range, they were treated as continuous variables in the analyses.

We conducted a Pearson correlation analysis for the pair of continuous variables (e.g., pretest - posttest scores) and computed the point-biserial correlation coefficients for the pairs of continuous and dichotomous variables (e.g., posttest scores - condition, grade, gender). For the pairs of dichotomous variables (e.g., gender - condition), phi coefficients were computed. As shown in Table 3, two variables showed statically significant, strong positive associations with the posttest scores: pretest scores ($r(473) = .65$, $p < .001$) and being in the advanced-level classes ($r_{pb}(473) = .56$, $p < .001$). Being identified as above grade level ($r_{pb}(473) = .45$, $p < .001$) or Asian ($r_{pb}(473) = .35$, $p < .001$) was also associated with higher posttest scores. The correlation analysis indicated potential associations among classroom level, student achievement level, ethnicity, and student performance, therefore we included these variables as covariates when estimating intervention effects in the following models.

RQ 1: Do Students Make Gains in Understanding of Equivalence after a Two-hour Intervention?

First, we examined whether students in the two conditions performed comparably on the pretest. An independent sample t-test on the pretest revealed that the baseline scores did not significantly differ between students in FH2T ($M = 3.86$, $SD = 1.60$) and problem set ($M = 3.75$, $SD = 1.62$) conditions, $p = .477$; nor were they significant in a 2-level HLM analysis based on students nested within classroom ($\gamma = 0.05$, $t(448) = 0.39$, $p = .699$).

Next, to examine the overall effects of intervention on students' performance, regardless of their intervention condition, we conducted a paired-sample t-test comparing the scores at pretest and posttest. Ignoring condition, students improved their performance from pretest (*M* = 3.80, *SD* = 1.61) to posttest (*M* = 4.14, *SD* = 1.56), $t(474) = 5.56$, $p < .001$. When each condition was examined separately, both were found to experience significant gains (FH2T: Posttest = 4.30 Gain = 0.44, $t(226) = 5.06$, $p < .001$; problem set: Posttest = 4.00, Gain = 0.25, $t(247) = 2.92$, $p = .004$).

RQ 2: Do Students in FH2T Condition Show Greater Understanding of Equivalence at Posttest Compared to Students in the Problem Set Condition?

We used HLM to examine differences in posttest scores, controlling for pretest performance and other student characteristics. A 2-level model (students nested within classrooms) was used on all analyses moving forward. We considered a 3-level model (classrooms nested within schools) but chose not to use it due to the small amount of variance at the school-level. While not reported here, the 3-level results were consistent with the 2-level model. In a preliminary, null model with no predictors, 52.2% of the variance in posttest scores was at Level 1, while 47.8% was at Level 2.

An initial model then examined the classroom (i.e., grade, instruction level) and student (i.e., gender, race, student achievement level, pretest) covariates, without considering the condition (FH2T vs. control) indicator. Robust standard errors were used to evaluate all effects, and multiparameter hypothesis tests were used to examine student race and instruction-level of the classroom as both of these constructs were reflected in two dummy variables. Not surprisingly, as seen in the first set of columns in Table 4, student pretest scores were highly related to posttest performance ($\gamma = 0.39$, $t(444) = 10.77$, $p < .001$), as was identification of “above grade level” on academic achievement ($\gamma = 0.31$, $t(444) = 3.15$, $p = .002$). In contrast, neither student gender ($\gamma = -0.14$, $t(444) = -1.10$, $p = .270$), nor race ($\chi^2(2, N=475) = 3.21$, $p = .199$) were found to have a statistically significant effect on posttest scores. At Level 2, both the grade-level of the class ($\gamma = 0.53$, $t(22) = 2.48$, $p = .021$) and instruction-level (advanced/supported, $\chi^2(2, N=475) = 95.23$, $p < .001$) were related to posttest scores in expected directions, with relatively higher performance among students in advanced classes ($\gamma = 1.13$, $t(22) = 5.03$, $p < .001$) and relatively lower performance in supported classes ($\gamma = -0.57$, $t(22) = -3.23$, $p = .004$), compared to those in on-level classes. It should be noted that the Level 2 variance no longer met a strict "$p < .05$" significance level, raising the question of switching to a traditional multiple regression; however, the decision was made to remain with an HLM approach for theoretical and conceptual grounds, as well as in order to reflect the original design and analytic plan (Raudenbush & Bryk, 2002).

The intervention condition was then added as a Level 1 predictor in the final model (see the second set of columns in Table 4). Reflecting the student-level random assignment, coefficients and statistical tests for the covariates were largely unchanged, however intervention condition was highly statistically significant (FH2T posttest: estimated mean =4.27; problem set posttest: estimated mean = 4.02), $\gamma = 0.25$, $t(443) = 3.81$, $p < .001$. In terms of group mean differences, this translated to a Hedge's *g* of 0.16 between the FH2T and active control problem set condition.

In order to help readers judge the practical importance of this intervention effect, we translated this effect size into the What Works Clearing House Improvement Index (Appendix E in What Works Clearinghouse, 2020), which can be interpreted as the expected change in percentile rank for students in an average comparison group if the students had received the intervention. To calculate this value, we first converted the effect size (Hedges' g) to Cohen's U3 index. An effect size of 0.16 corresponds to a U3 of 56.4%, indicating that an average student in the FH2T group would rank at the 56.4 percentile in the control group. To calculate the improvement index, representing the difference in percentile rank of an average FH2T student compared to an average comparison group, we then subtracted 50 from the U3 (56.4-50=6.4). Practically speaking, the improvement index of 6.4 suggests that teachers and administrators could expect an average student to improve 6.4 percent rank (from 50 to 56.4%) after using the FH2T intervention for 2 hours, compared to using other effective programs like problem sets in ASSISTments.

RQ 3: Does the Effect of Intervention Vary Depending on Students' Prior Knowledge?

A final analysis examined the possible interaction between intervention condition and students' prior knowledge on students' posttest performance. The analysis revealed that the interaction between intervention condition and pretest scores was not significant, $\gamma = -0.02$, $t(442) = -0.57$, $p = .556$, suggesting that the FH2T effects were consistent for students who began the program performing at both higher and lower levels.

Performance on Individual Assessment Items.

Because the pretest and posttest assessments only included six items, and the reliability was suboptimal, we explored changes in students' pretest and posttest performance on each item by intervention conditions (Table 5). We found that students' performance on Items 1, 2, 3, 4 showed minimal change from pretest to posttest. In both conditions, students showed the largest gains on Item 5, which focused on their understanding of equivalent expressions for addition and multiplication with the prompt, "Which of the following is equivalent to (the same as) (n + 3) + (n + 3) + (n + 3) + (n + 3)?". In particular, 27% of students in FH2T condition improved on this item, whereas only 19% of students in the problem set condition improved from pretest to posttest. Fewer students (FH2T: 9% and problem set: 5%) improved on Item 6 (order of operations and distribution), and the difference between the conditions were small. The findings suggest that some of the features in FH2T (e.g., dynamically transform expressions into visually different yet equivalent states) may have a positive influence on students' understanding of equivalent expressions.

Discussion

In summary, we found that (a) students in both conditions improved their performance of mathematical equivalence from pretest to posttest, (b) students in the FH2T condition showed higher performance on the posttest compared to students in the problem set with hints and feedback condition, even after controlling for students' prior knowledge and mathematics instruction level, and (c) the condition effects were similar for all students, regardless of their prior knowledge. Together, the findings suggest that FH2T is an effective intervention for improving middle-school students' understanding of mathematical equivalence.

FH2T and Problem Sets Conditions Improve Students' Understanding of Equivalence

In this study, we found that students in FH2T and problem set conditions both improved their understanding of mathematical equivalence after four 30-minute intervention sessions. Although the problem set condition was included to serve as a control in the current study, the ASSISTments learning platform was an established and effective educational technology that has been recommended without reservation by the What Works Clearinghouse (What Works Clearinghouse, 2020; Roschelle et al., 2016). Therefore, it was not surprising that solving textbook problems on arithmetic operations in ASSISTments with some student features of the platform, specifically hints on problem-solving and immediate correctness feedback during assignments, had positive effects on students' understanding of mathematical equivalence. This finding extends previous studies demonstrating ASSISTments efficacy (Murphy et al., 2020; Roschelle et al., 2016) and suggests that even implementing only the student aspects of ASSISTments (and not the teacher supports) in a brief intervention may lead to improvements in student learning.

We also found that students in the FH2T condition had better understanding of mathematical equivalence at posttest compared to students in the problem set condition. This aligns with previous findings demonstrating the effectiveness of From Here to There! in elementary (Ottmar et al., 2019) and middle-school classrooms (Ottmar et al., 2015) on improving procedural knowledge. This study expands on the prior work by suggesting that playing FH2T also improves students' understanding of mathematical equivalence. Importantly, the effect of the condition was significant and positive, suggesting that students in the FH2T condition had better understanding of mathematical equivalence at posttest compared to students in the problem set active control. Although the condition effect may be modest (Hedge's $g = .16$) in comparison to other extensive mathematics interventions (e.g., Blanton et al., 2015), it is worth noting that the intervention was brief (2 hours) and the comparison condition itself was a robustly effective evidence-based intervention with an Improvement Index of 7 (What Works Clearinghouse, 2020; Roschelle et al., 2016). The added benefit of FH2T above and beyond the significant improvement found in the active ASSISTments control condition suggests that the condition effect may underestimate the true efficacy of FH2T, if compared to a true business-as-usual control.

The Impact of FH2T on Understanding of Equivalence

Despite major efforts in research, curricula development, and policy, many students continue to struggle with understanding equivalence. The results showed that using FH2T improved performance on mathematical equivalence and conceptual assessment items compared to a control condition of online problem sets with hints and feedback in ASSISTments. Similar to previous equivalence interventions (Alibali et al., 2018; Blanton et al., 2015; McNeil et al., 2012), FH2T guides students' attention to expression structures ($24 + y + 6 + 13$ and $13 + y + 30$), and provides ample practice on equivalent expressions that build on arithmetic and mathematical properties. The findings extend previous research on equivalence interventions with elementary students and show that middle-school students may also benefit from interventions targeting equivalence understanding.

Different from the previous interventions, FH2T utilizes a dynamic algebra notation system that allows students to concretely transform expressions and receive immediate visual feedback on their actions. Although this study does not tease apart the mechanisms by which FH2T leads to improved learning, there are several plausible explanations. One possible explanation is that the perceptual learning (i.e., being able to dynamically manipulate algebraic symbols and experience math transformations) afforded in the game provides students with explicit and dynamic ways to productively explore abstract mathematics concepts. These findings are consistent with other research which found utilizing perceptual features or engaging perceptual motor systems in mathematics learning contexts had a positive influence on learning (Hulse et al., 2019; Kellman et al., 2010; Landy & Goldstone, 2007b; Ottmar et al., 2012, Ottmar et al., 2015). These are also in line with the theory that algebra reasoning comprises perceptual-motor routines (Goldstone et al., 2008; Landy & Goldstone, 2007b). In FH2T, the symbols are treated as virtual objects and are constrained to mathematically appropriate behaviors, resulting in fast feedback about possible transformations, and fluid continuous visualizations. It may be that these affordances in FH2T provide students with ample opportunities to explore and learn which dynamic actions and mathematical properties are appropriate and allowed in different mathematical contexts (De Lima & Tall, 2008; Dörfler, 2003; Goldstone et al., 2010; Landy & Goldstone, 2009). This study provides further evidence that technology-based perceptual interventions, like FH2T, may provide students with conceptually rich opportunities to explore algebraic formalisms.

The FH2T intervention may also benefit students by providing practice with various structures of expressions (e.g., transform $24 + y + 6 + 13$ to $13 + y + 30$) and equations (e.g., transform $23 + y – 13 = 10 + y$ to $3 = 3$) that deviate from a more traditional operations-equals-answer structure (e.g., $24 + 6 + 13 = 43$; Fyfe et al., 2018). Attending to mathematical relations in a variety of structures and enacting procedures appropriately are important for success in algebra (Kieran, 1989). The problems in FH2T are uniquely designed to present students with perceptually different structures of expressions and equations, and asks them to enact algebraic transformations that prove their equivalence. Giving students opportunities to explore different problem structures with varying perceptual features may improve their flexibility and expand their definitions of what equivalence means. Further, it is also plausible that the connected sensorimotor experience between the gesture-actions and resulting mathematical transformations in FH2T guide students' attention to the fluid visualization of equivalent transformations, reduce the cognitive demands of rewriting and computing complex expressions, and allow students to focus on the conceptual understanding of the links between the steps in a derivation and the high-level structure of equations.

The procedural advantages of moving symbols that seamlessly integrate with conceptually challenging expression transformation tasks in the FH2T system may help students become more familiar with algebraic notations, acquire perceptual and conceptual fluency in algebraic principles, and increase their confidence and comfort in dealing with equations. This increase in algebra familiarity and proficiency grounded in perceptual learning of procedural fluency and conceptual understanding, may improve learning outcomes in more advanced areas of algebra that assume the ability to read and manipulate equations.

Alternatively, the condition effect may be driven by other differences between FH2T and the problem set condition. For instance, the game-based mathematics puzzles, such as FH2T, may be more engaging for students and students may be more motivated to solve problems in FH2T compared to traditional answer-based online problem sets. Further, the higher posttest performance in FH2T may be due to the fact that half of the students in FH2T had the opportunity to transform linear equations as they progressed through the game, whereas students in the problem set condition only completed the assigned problem sets relevant to the four arithmetic operations and the order of operations. Although the exposure to more challenging content in the FH2T may contribute to the condition effect we observed, students made the largest improvement on Item 5 of the assessment (equivalent expressions of addition and multiplication; FH2T: 27% of students improved, problem set condition: 19%). This item did not involve linear equations and students in both conditions had the opportunity to practice solving problems relevant to this item. Therefore, it is unlikely that the differences in the content exposure between the two conditions drove the condition effect.

Another potential account for the condition effect is that students in FH2T practiced transforming expressions into other equivalent forms that are visually and structurally different. The expression transformation task in FH2T may explicitly highlight mathematical equivalence for students. On the contrary, students in the problem set condition solved a variety of traditional textbook problems with the goal of simplifying to a correct answer. These textbook problems may seem more like a simplifying task to students and do not explicitly emphasize equivalence concepts. Future studies should explore these alternative hypotheses and further delineate the mechanisms through which FH2T promotes mathematical learning.

From Here to There and Prior Knowledge

The lack of an interaction effect between the intervention condition and pretest scores suggests that FH2T may be benefiting students to the same degree regardless of their prior knowledge. Different from the two competing hypotheses presented in prior work (Murphy et al., 2020; Swanson et al., 2008), we found that neither students with higher nor lower prior knowledge benefitted more from the FH2T intervention. Instead, we found that the effect of intervention was comparable for students with high and low prior knowledge. Further, extending our previous study in which we showed that elementary students with low prior knowledge improved more on mathematics performance when they completed more problems in FH2T (Hulse et al., 2019), we found that simply receiving FH2T intervention, without considering the in-game progress, did not benefit middle-school students with low prior knowledge more than students with high prior knowledge. Replicating these effects with different samples (e.g., elementary vs. middle school), content covered, analytic approaches (e.g., influences of FH2T progress vs. impact of FH2T intervention regardless of progress), and mathematical outcomes (e.g., mathematics achievement vs. perception of algebraic expressions) may help further examine the potential effects of these factors on the relations between prior knowledge, FH2T intervention, and mathematical performance.

Limitations and Future Directions

While this study suggests that FH2T is a promising intervention for improving understanding of mathematical equivalence, our current study is limited in a few ways. First, given the population of our sample, the results may not be representative or generalizable to different populations in the U.S. The sample in this study had a high number of Asian and above grade level students. While the composition of our sample based on race and prior performance is not generalizable to the demographics of the U.S. as a whole, student-level random assignment within classrooms ensures comparable performance between FH2T and control students at pretest. Despite the fact that 84% of the sample were in an advanced math class, many students were not at ceiling on the pretest and the average pretest score was only 3.89 (out of 6; 65%), providing room for improvement on equivalence understanding. Further, there was adequate distribution and variability in pretest and posttest assessments, allowing for the examination of the improvement and intervention effects. Future work should aim to replicate these findings in more diverse populations.

Next, since this study was conducted in the classroom, the administration of our pretest and posttest assessments relied on teachers allocating appropriate classroom time for the intervention and assessments. However, because the study took place immediately prior to a holiday break, several teachers did not provide their students with adequate class time to complete the full intervention sessions or posttest, which resulted in a high number of students missing data at posttest. This challenge highlights the tradeoffs of conducting applied experimental work in authentic classrooms. Future studies should include other distal measures (e.g., state standardized math assessments) so that the student outcome does not depend on the fidelity of the teachers. This would also afford testing of the sustained effects of the intervention.

Further, the results presented here, although significant, do not provide insight into what components of the FH2T intervention lead to improved learning. Future directions include adding proximal measures (e.g., an expression matching task) to delineate the effects of FH2T on equivalence understanding, and examining plausible mechanisms by which progress through the FH2T game leads to improved mathematical performance. The availability of clickstream data from the FH2T game also provides opportunities to explore the relations between student behaviors within each FH2T problem and math learning. We hypothesize that greater exposure to the problems and engagement within FH2T will lead to greater gains in perceptual learning, understanding of algebraic principles, and flexible problem solving. Additional studies will explore these hypotheses.

Finally, FH2T is designed to guide students' attention to mathematical notations and structures through perception and action. Rather than relying on language-based instruction and students' English proficiency, FH2T uses gesture-action videos to dynamically demonstrate mathematical principles and provides a fluid interface for students to experience mathematical transformations. Providing perceptual-motor pathways to mathematical learning may help bridge the achievement gap for English Language Learners. Future studies should explore the accessibility and usability of FH2T in students from more diverse backgrounds, including English Language Learners, and such findings will contribute to the efforts of promoting diversity, equity, and inclusion in mathematics education.

Conclusion

Overall, this study supports the efficacy of the FH2T intervention for improving students' mathematics performance, above and beyond solving traditional textbook problems in ASSISTments, a well-studied and effective active control. FH2T is a promising intervention that addresses a relatively untapped area of practice-focused, perceptually guided instructional technology, designed based on cognitive theories. This study has implications for educators: It provides evidence that perceptually-focused, gamified learning platforms may help students develop perceptual fluency through dynamic interactions with algebraic objects, and provide a useful learning environment for students to explore mathematical ideas and improve their mathematical understanding.

References

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

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

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

Arzarello, F., Bairral, M. A., & Dané, C. (2014). Moving from dragging to touchscreen: Geometrical learning with geometric dynamic software. *Teaching Mathematics and Its Applications*, *33*(1), 39–51. https://doi.org/10.1093/teamat/hru002

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

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

Byers, P., & Hadley, J. (2012). Traditional and novel modes of activity in touch screen math apps. In *Proceedings of the 12th International Conference on Interaction Design and Children*. New York, NY.

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

Chaiklin, S., & Lesgold, S. B. (1984). *Prealgebra students' knowledge of algebraic tasks with arithmetic expressions*. University of PIttsburgh: Learning Research and Development Center

Crooks, N. M., & Alibali, M. W. (2013). Noticing relevant problem features: activating prior knowledge affects problem solving by guiding encoding. *Frontiers in Psychology*, *4*(November), 884. https://doi.org/10.3389/fpsyg.2013.00884

De Lima, R. N., & Tall, D. (2008). Procedural embodiment and magic in linear equations. *Educational Studies in Mathematics*, *67*(1), 3–18. https://doi.org/10.1007/s10649-007-9086-0

Dörfler, W. (2003). Mathematics and mathematics Education: Content and people, relation and difference. *Educational Studies in Mathematics*, *54*, 147–170.

Dourish, P. (2004). *Where the action is: The fundations of embodied interaction*. Cambridge, MA: The MIT Press.

Engage, N. Y. (2014). New York State Education Department. Retrieved from https://www.engageny.org

Fyfe, E. R., Matthews, P. G., Amsel, E., McEldoon, K. L., & McNeil, N. M. (2018). Assessing formal knowledge of math equivalence among algebra and pre-algebra students. *Journal of Educational Psychology*, *110*(1), 87–101. https://doi.org/10.1037/edu0000208

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

Goldstone, R., Landy, D., & Son, J. Y. (2008). A well grounded dducation: The role of perception in science and mathematics Printed. *Symbols, Embodiment, and Meaning*

Debate, 327–355.

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*. https://doi.org/10.1007/s40593-014-0024-x

Henry, V. J., & Brown, R. S. (2008). First-grade basic facts: An investigation into teaching and learning of an accelerated, high-demand memorization standard. *Journal for Research in Mathematics Education*, *39*(2), 153–183.

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

Illustrative, M. (2017). Illustrative Mathematics Project. Retrieved from https://www.illustrativemathematics.org

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, *38*(3), 258–288.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science*, *320*(5875), 454–455. https://doi.org/10.1126/science.1154659

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

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in*

Mathematics, (12), 317–326.

Kieran, C. (1989). The early learning of algebra: A structural perspective. In S. Wagner & C. Kieran (Eds.), *Research Issues in the Learning and Teaching of Algebra: The Research Agenda for Mathematics Education, Volume 4* (pp. 33–56). Virginia, U.S.A: NCTM and Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315044378-4

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning: A project of the National Council of Teachers of Mathematics* (pp. 390–419). New York, NY: Macmillan Publishing Co, Inc.

Kieran, C. (2007). Learning and teaching of algebra at the middle school through college levels: Building meaning for symbols and their manipulation. In F. K. Lester Jr. (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (p. Vol. 2, pp. 707–762). Charlotte, NC: Information Age.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education*, *20*(3), 274. https://doi.org/10.2307/749516

Kirshner, D., & Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, *35*(4), 224–257. https://doi.org/10.2307/30034809

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.

Küchemann, D. (1980). Algebra. In K. M. Hart (Ed.), *Children's Understanding of Mathematics: 11-16* (pp. 103–119). John Murray.

Landy, D., & Goldstone, R. (2009). How much of symbolic manipulation is just symbol pushing? In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 1318–1323). Retrieved from http://csjarchive.cogsci.rpi.edu/Proceedings/2009/papers/253/paper253.pdf

Landy, D., & Goldstone, R. L. (2007a). Formal notations are diagrams: Evidence from a production task. *Memory and Cognition*, *35*(8), 2033–2040. https://doi.org/10.3758/BF03192935

Landy, D., & Goldstone, R. L. (2007b). 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

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Quarterly Journal of Experimental Psychology*, *63*(10), 1953–1968. https://doi.org/10.1080/17470211003787619

Lobato, J., Hohensee, C., & Rhodehamel, B. (2013). Students' mathematical noticing. *Journal for Research in Mathematics Education*, *44*(5), 809–850. https://doi.org/10.5951/jresematheduc.44.5.0809

Mcewen, R. N., & Dubé, A. K. (2015). Engaging or distracting : Children's tablet computer use in education. *Journal of Educational Technology & Society*, *18*(4), 9–23.

McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. S., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, *24*(3), 367–385. https://doi.org/10.1207/s1532690xci2403

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

McNeil, Nicole M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*(2), 171–184. https://doi.org/10.1016/j.learninstruc.2008.03.005

McNeil, Nicole M, & Alibali, M. W. (2000). Learning mathematics from procedural instruction: Externally imposed goals influence what is learned. *Journal of Educational Psychology*, *92*(4), 734–744. https://doi.org/10.1037/0022-0663.92.4.734

McNeil, Nicole M, & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*(4), 883–899. https://doi.org/10.1111/j.1467-8624.2005.00884.x

Mix, K. S. (1999). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology*, *74*(4), 309–332. https://doi.org/10.1006/jecp.1999.2533

Murphy, R., Roschelle, J., Feng, M., Mason, C. A., Murphy, R., Roschelle, J., … Mason, C. A. (2020). Investigating efficacy, moderators and mediators for an online mathematics homework intervention. *Journal of Research on Educational Effectiveness*, *0*(0), 1–36. https://doi.org/10.1080/19345747.2019.1710885

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

Ottmar, E. R., Landy, D., & Goldstone, R. L. (2012). Teaching the perceptual structure of algebraic expressions: Preliminary findings from the pushing symbols intervention. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, *34*, 2156–2161. Retrieved from http://www.indiana.edu/~pcl/papers/pushingsymbols.pdf

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 Annual Conference of the Cognitive Science Society*, 1793–1798.

Patsenko, E. G., & Altmann, E. M. (2010). How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General*, *139*(1), 95–116. https://doi.org/10.1037/a0018268

Perry, M., Breckinridge Church, R., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*(4), 359–400. https://doi.org/10.1016/0885-2014(88)90021-4

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*(3), 146–159. https://doi.org/10.1016/j.appdev.2011.02.005

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd Editio). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2019). *HLM8: Hierachical linear and nonlinear modeling*. Skokie, IL: Scientific Software International.

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

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, *2*(4), 233285841667396. https://doi.org/10.1177/2332858416673968

Siew, N. M., Geofrey, J., & Lee, B. N. (2016). Students' algebraic thinking and attitudes towards algebra: The effects of game-based learning using Dragonbox 12 + App. *The Research Journal of Mathematics and Technology*, *5*(1), 66–79. Retrieved from http://www.academia.edu/download/46188339/2dragonbox.pdf

Sinclair, N., & Heyd-Metzuyanim, E. (2014). Learning number with TouchCounts: The role of emotions and the body in mathematical communication. *Technology, Knowledge and Learning*, *19*(1–2), 81–99. https://doi.org/10.1007/s10758-014-9212-x

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2014). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

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

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

U.S. Department of Education. (2008). The final report of the national mathematics advisory panel. *Foundations for Success*.

Usiskin, Z. (1988). Conceptions of school algebra and uses of variables. *Algebraic Thinking, Grades K-12*, (5), 8–19.

Utah, M. (2016). The Utah Middle School Math Project. Retrieved from http://utahmiddleschoolmath.org/%0A

Uttal, D. H., Gentner, D., Liu, L. L., & Lewis, A. R. (2008). Developmental changes in children's understanding of the similarity between photographs and their referents. *Developmental Science*, *11*(1), 156–170. https://doi.org/10.1111/j.1467-7687.2007.00660.x

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

What Works Clearinghouse (2020). *What Works Clearinghouse standard handbook, Version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Table 1
Students' demographic information by condition, and their pretest scores.

	All (N = 475)		FH2T (n = 227)		Problem set (n =248)	
	n	%	*n*	%	*n*	%
Gender						
male	261	54.9	127	55.9	134	54.0
female	214	45.1	100	44.1	114	46.0
Race						
White	165	34.7	81	35.7	84	33.9
Asian	260	54.7	120	52.9	140	56.5
Hispanic	23	4.8	10	4.4	13	5.2
African American	10	2.1	6	2.6	4	1.6
Native American	5	1.1	1	0.4	4	1.6
Pacific Islander	1	0.2	1	0.4	0	0.0
Multi-racial	11	2.3	8	3.5	3	1.2
Grade						
Sixth	453	95.4	217	95.6	236	95.2
Seventh	22	4.6	10	4.4	12	4.8
Class						
Advanced	400	84.2	192	84.6	208	83.9
On-Level	34	7.2	14	6.2	20	8.1
Support	41	8.6	21	9.3	20	8.1
Student Achievement Level						
Above grade	248	52.2	119	52.4	129	52.0
Not above grade	227	47.8	108	47.6	119	48.0
Pretest scores (*M*, *SD*)	3.80	1.61	3.86	1.60	3.75	1.62

Table 2.

The mathematical equivalence items in pretest

Item	Question	Correct answer	Reference
1	8+__=8+6+4. Enter the number that goes in the blank.	10	Rittle-Johnson et al., 2011
2	898 + 13 = 896 + __ What number goes in the blank? You can try to find a shortcut, so you don't have to do all the adding.	15	Rittle-Johnson et al., 2011
3	3 + 4 = 7 ↑ What does this symbol mean? a. the total b. two quantities on either side have the same value c. what the answer is d. the problem has been solved	b	Star et al., 2014
4	If 10x + 12 = 17, which of the following must also be true? a. 10x+12-12=17-12 b. 10x-10+12-10=17 c. -10x-12=17 d. 5x+6=17	a	Star et al., 2014
5	Which of the following is equivalent to (the same as) (n + 3) + (n + 3) + (n + 3) + (n + 3)? a. n+12 b. 4n+3 c. n^4+12 d. 4(n+3)	d	Star et al., 2014
6	Which of the following is NOT equivalent to 19(73 - 15)? a. 19(58) b. 19(73) - 19(15) c. 19(-15 + 73) d. 19(73) - 15	d	Star et al., 2014

Table 3

Descriptive statistics and correlations for the overall sample

Variable	1	2	3	4	5	6	7	8	9	10
1.Posttest	-									
2.Pretest	.65**	-								
3. Condition	.09*	.03	-							
4. Grade	-.27**	-.26**	-.01	-						
5. Class-Adv	.56**	.49**	.01	-.51**	-					
6. Class-Sup	-.47**	-.46**	.02	.68**	-.71**	-				
7. Gender	-.08	-.06	-.02	.04	-.07	.04	-			
8. Asian	.35**	.45**	-.04	-.22**	.27**	-.31**	.09*	-		
9. White	-.30**	-.38**	.02	.09*	-.21**	.22	-.08	-.80**	-	
10.Stud level	.45**	.49**	.00	-.21**	.43**	-.31**	.05	.28**	-.20**	-
Mean	4.14	3.80	.48	6.05	.84	.09	.45	.55	.35	.52
SD	1.56	1.61	.50	.21	.37	.28	.50	.50	.48	.50
Min.	.00	.00	0	6	0	0	0	0	0	0
Max.	6.00	6.00	1	7	1	1	1	1	1	1

Note. Condition (0 = problem-set, 1 = FH2T); Class-Adv: Class Instruction Level-Advanced (0 = Not advanced,1 = Advanced); Class-Sup: Class Instruction Level-Support (0 = Not support, 1 = Support); Gender (0 = Male, 1 = Female); Asian: Race-Asian (0 = Not Asian, 1 = Asian); White: Race-White (0 = Not White, 1 = White); Stud level: Student Achievement Level (0 = Not above grade level,1 = Above grade level
* indicates $p < .05$. ** indicates $p < .01$.

Table 4

Result of HLM analyses.

	Initial Model				Final Model			
Effect	Coef	*t*	df	*p*	Coef	*t*	df	*p*
Intercept	-1.61	-1.19	22	.248	-1.92	-1.37	22	.185
Grade	0.53	2.48	22	.021	0.57	2.56	22	.018
Class-Adv	1.13	5.03	22	<.001	1.11	4.61	22	<.001
Class-Sup	-0.57	-3.23	22	.004	-0.62	-3.04	22	.006
Gender	-0.14	-1.10	444	.270	-0.14	-1.10	443	.271
Asian	0.15	1.09	444	.278	0.17	1.27	443	.205
White	-0.05	-0.36	444	.719	-0.04	-0.30	443	.762
Stud Level	0.31	3.15	444	.002	0.31	3.21	443	.001
Pretest	0.39	10.77	444	<.001	0.38	-11.16	443	<.001
FH2T					0.25	3.81	443	<.001
Random Effects	Var	X^2	df	p	Var	X^2	df	p
Level 1	1.167				1.155			
Level 2	0.032	31.4	22	0.088	0.030	30.7	22	0.103

Note. Grade: Grade in school; Class-Adv: Class Instruction Level-Advanced (0 = Not advanced, 1 = Advanced); Class-Sup: Class Instruction Level-Support (0 = Not support, 1= Support); Gender: Student gender (0 = Male, 1 = Female); Asian: Race-Asian (0 = Not Asian, 1=Asian); White: Race-White (0 = Not White, 1 = White); Stud Level: Student Achievement Level (0 = Not above grade level, 1 = Above grade level); Pretest: Pretest score; FH2T: Intervention Condition-FH2T (0 = Problem-set, 1 = FH2T).

Table 5

Percent of students responded correctly on each item of the pretest and posttest, and the percent of students improved on the items by intervention condition.

	FH2T (n = 227)			Problem Sets (n = 248)		
Item	Pretest	Posttest	Gain	Pretest	Posttest	Gain
1	91%	93%	2%	91%	91%	0%
2	86%	87%	1%	84%	81%	-3%
3	54%	57%	4%	49%	54%	5%
4	57%	58%	1%	53%	52%	-1%
5	51%	78%	27%	52%	71%	19%
6	47%	56%	9%	46%	52%	5%

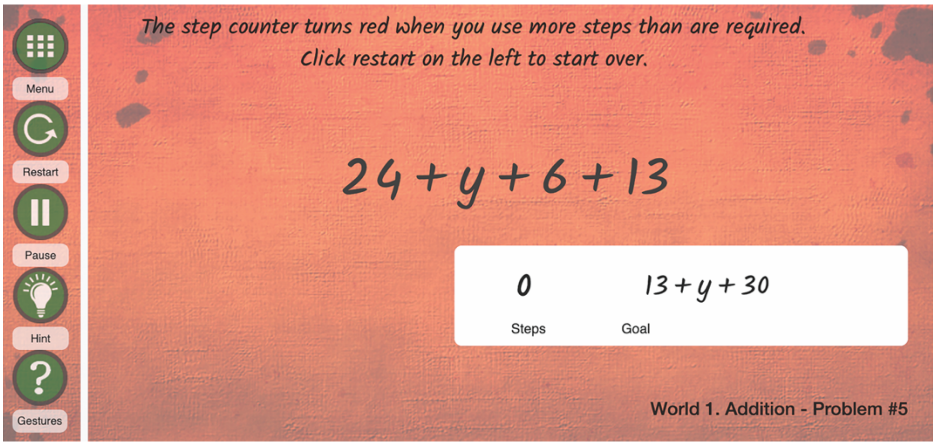

Figure 1. An example of FH2T problem consisting of a starting expression and a target goal state.

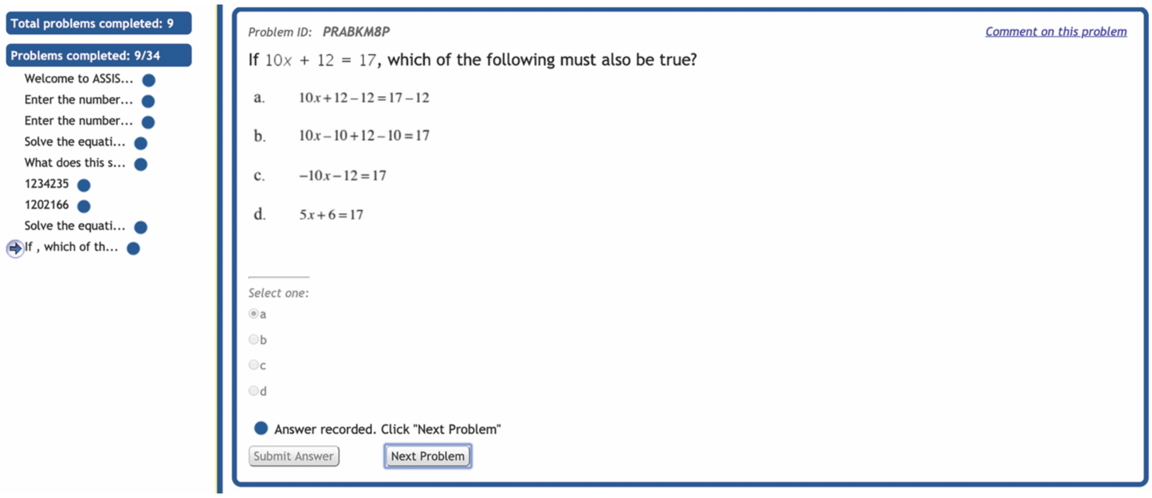

Figure 2. The basic layout and procedure in ASSISTments.

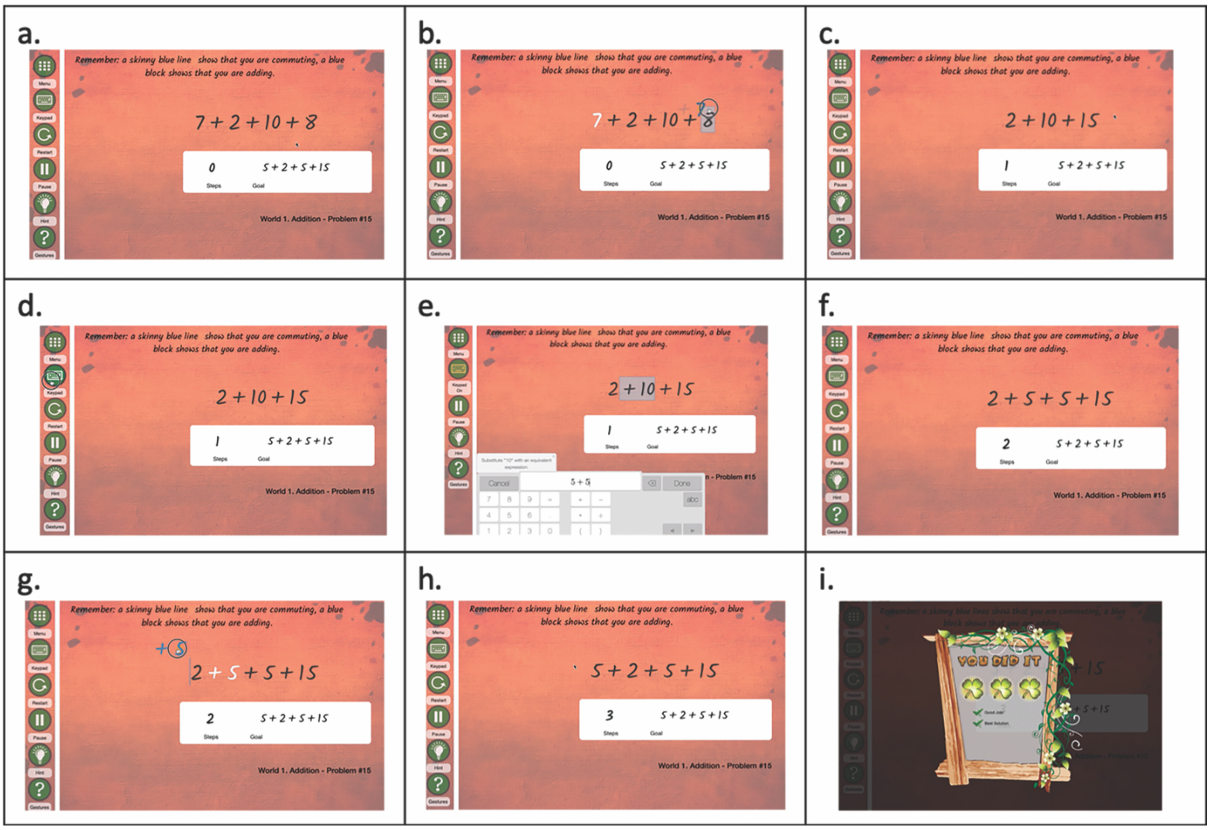

Figure 3. A sample problem in From Here to There! (a) and a potential transformation process involving three steps (b, c, d, e, f, g) to reach the goal state (h, i).

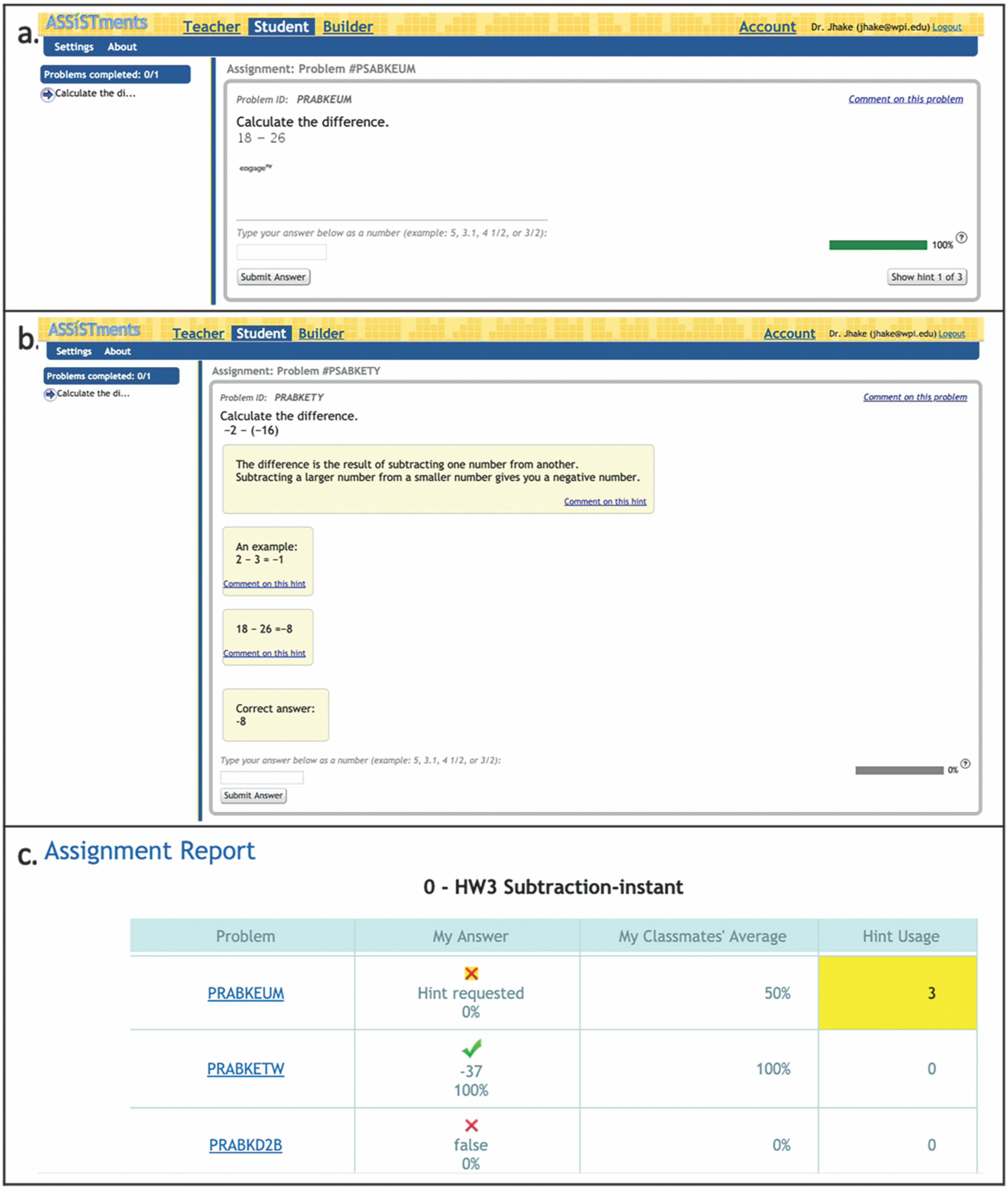

Figure 4. A sample question (a), three hints and the correct answer in yellow boxes (b), and the assignment report (c) in ASSISTments