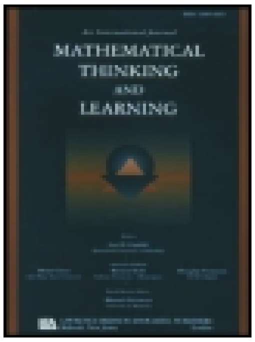

Mathematical Thinking and Learning

ISSN: 1098-6065 (Print) 1532-7833 (Online) Journal homepage: https://www.tandfonline.com/loi/hmtl20

Features of a pan balance that may support students’ developing understanding of mathematical equivalence

Neet Priya Bajwa & Michelle Perry

MATHEMATICAL THINKING AND LEARNING
https://doi.org/10.1080/10986065.2020.1700587

Features of a pan balance that may support students' developing understanding of mathematical equivalence

Neet Priya Bajwa[a] and Michelle Perry[b]

[a]Department of Mathematics, Illinois State University, Normal, Illinois, USA; [b]Department of Educational Psychology, University Of Illinois at Urbana-Champaign, Champaign, Illinois

ABSTRACT

Elementary school students struggle in interpreting the equal sign as a symbol denoting equivalence. Although many have advocated using a pan-balance scale to help students develop this understanding, less is known about what features associated with this model support learning. To attempt to control and examine these features, the investigators developed a digital, pan-balance computer applet. This allowed for experimentally manipulating three relatively grounded instructional conditions (involving the core principle of making two sides the same; a balancing analogy; or both, along with a dynamic demonstration), compared to a relatively idealized control condition. Results indicated that the relatively more grounded conditions promoted a relational understanding of mathematical equivalence among 148 second- and third-grade students and further suggest that providing dynamic, grounded support may not be as optimal as less-enriched supports to promote students' learning.

ARTICLE HISTORY
Received 6 November 2018
Revised 29 November 2019
Accepted 1 December 2019

KEYWORDS
Mathematical equivalence; equal sign; balance analogy; digital manipulative

Understanding mathematical equivalence denoted by the equal sign – the principle that two sides of an equation represent the same quantity – in symbolic problems is vexing for many young students. For decades, educational researchers (e.g. Behr, Erlwanger, & Nichols, 1980; Kieran, 1980; Kilpatrick, Swafford, & Findell, 2001; Molina & Ambrose, 2008; Perry, 1991; Renwick, 1932; Stephens, Ellis, Blanton, & Brizuela, 2017; Stephens et al., 2013) have documented young students' difficulty with this concept. Although there has been some remarkable work with children and adolescents on understanding the equal sign in the last decade, figuring out the unknown in open equations continues to be a hurdle for many students (Chesney et al., 2014).

It appears that students' difficulties in the United States typically stem from the problem that they perceive the equal sign as having an *operational* meaning (i.e. "to compute" or "to find the total") instead of a *relational* one (i.e. indicating a relation of equality or quantitative sameness between the two sides of the symbol, see, e.g. Baroody & Ginsburg, 1983; Behr et al., 1980; Blanton et al., 2018; Carpenter, Franke, & Levi, 2003; Falkner, Levi, & Carpenter, 1999; Kieran, 1981; McNeil & Alibali, 2005b; Rittle-Johnson & Alibali, 1999; Seo & Ginsburg, 2003). This notion of the equal sign as a command to compute makes it difficult for young students when they see problems in nonstandard equation formats, for example, in finding the unknown in equations where an unknown does not appear alone in the final position (e.g. $2 + 3 + 4 = \_\_ + 4$; e.g. Li, Ding, Capraro, & Capraro, 2008). Students with an operational view of the equal sign tend to write 9 or 13 in the blank when asked to find the missing number in this problem. Such nonstandard problems – with operations

CONTACT Neet Priya Bajwa nbajwa@ilstu.edu Department of Mathematics, Illinois State University, USA
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hmtl.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

appearing on both sides of the equal sign – have reliably been found to reveal students' problematic views of the equal sign and are collectively known as *equivalence problems*.

Although the equal sign can denote multiple kinds of relational meanings (including symmetry, identity, substitutive, reflexive, and transitive; see, e.g. Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Skemp, 1986), *symmetry* – or the *principle of sameness of the quantities on each side of the equal sign* – is one of the fundamental relational meanings in arithmetic and typically appears early in students' schooling. However, without this understanding, children oftentimes perceive the equal sign as an operational symbol – as a call to compute – which leads to difficulties not only in solving a variety of arithmetic problems, but also in limiting the development of algebraic thought. In fact, acquiring a relational view of the equal sign has been found to be a predictor of student success with solving equations in middle school (Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007) and thus perceiving the equal sign as a relational symbol may be crucial for the development of algebraic thinking (e.g. Byrd, McNeil, Chesney, & Matthews, 2015; Carpenter et al., 2003; Falkner et al., 1999; Kieran, 1981; Knuth, Stephens, McNeil, & Alibali, 2006; National Council of Teachers of Mathematics, 2000; Seo & Ginsburg, 2003). Given that understanding the equal sign indicates a relation of equivalence in arithmetic and in algebra, it is with good reason that the *Common Core State Standards Initiative* (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, 2012) has asserted the need for attending to mathematical equivalence and understanding the equal sign, from as early as the first grade. Although researchers have conducted remarkable work with children and adolescents on understanding the equal sign as an indicator of sameness and equivalence in the last decade, figuring out the unknown in open equations continues to be a hurdle for many students. For our current study, we focused specifically on a key idea of mathematical equivalence: quantitative symmetry or sameness (see, e.g. Rittle-Johnson & Alibali, 1999).

The pan balance as a concrete representation to support understanding of equivalence

Prior work on remedying the narrow operational view of the equal sign has emphasized, among other things, the benefits of non-symbolic (e.g. real-life scenarios) and concrete contexts (e.g. objects) to instantiate the meaning of equivalence (e.g. Barlow & Harmon, 2012; Caglayan & Olive, 2010; Seo & Ginsburg, 2003; Sherman & Bisanz, 2009; Stacey & MacGregor, 1997). This work has found that non-symbolic and concrete contexts allow students to understand a *core principle* underlying mathematical equivalence – *that both sides need to be the same*.

The success of non-symbolic and concrete contexts can be attributed theoretically, at least in part, to the extent with which they *ground* the idea to be learned (Belenky & Schalk, 2014). In the simplest form, *groundedness* can be seen as the extent to which the context offers external support for conceptual understanding, allowing learners to draw a parallel between the external support provided by the context and the intended concept to be learned. The groundedness may be afforded by the visual representation included in the support (e.g. objects, graphics, pictures), actions observed or taken (e.g. observing the balance scale teeter-totter up and down or manipulating the objects), verbal narrative that provides a conceptual link to something familiar (e.g. reminding the student of a common action or object that links to the underlying mathematics), or a combination of any of these cues.

One concrete manipulative has garnered unequal attention for supporting student understanding of mathematical equivalence: *the pan-balance scale*. Hiebert and Carpenter (1992) suggested that associating the fulcrum of the balance with the equal sign could act as a model for developing the conceptual understanding of the equal sign. Moreover, seeing the equal sign as a symbol that indicates quantitative sameness maps directly on the pan-balance model and thus may allow for students to read arithmetic and algebraic problems as quantitative symmetry (e.g. Alibali & Nathan, 2007; Rasmussen, 1977) or in an unbalanced state, the lack of symmetry between the two sides. In this way, the close proximity of the material and its features to the referent (equivalence of two sides) makes the pan-balance scale useful as a relatively grounded manipulative to support student learning (assuming, of course, that when trying to make the point that adding numbers on both sides comes to the same total, equivalent weights are used to balance the scale). In fact, national standards such as the NCTM (e.g. 2000; see, e.g. https://illuminations.nctm.org/Activity.aspx?id=3530) and others (e.g. *Common Core State Standards Initiative*, National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010, 2012) advocate for using this particular manipulative to build understanding of the equal sign as an indicator of symmetry.

The fact that a physical balance scale, by its very nature, offers a dynamic[1], external model of equivalence, at least in part, explains its appeal and reasons for mention in both policy documents and use as part of instruction. The pan-balance scale also seems intuitively appealing for educational purposes because of its iconicity and its closeness to students' experiences (e.g. on seesaws, see e.g. Mann, 2004). It is also widely represented in pre-service teacher guides (e.g. Baroody & Coslick, 1998; Powell, 2012; Van de Walle, 2013), research-based elementary school textbooks (e.g. *Everyday Mathematics*; also see Presmeg, 1997), and articles targeted for teachers (e.g. Leavy, Hourigan, & McMahon, 2013; Mann, 2004). Perhaps what makes a concrete, pan-balance scale so intuitive and valuable is that its mechanics represent the balancing act, allowing students to notice the balanced state (or imbalanced state) as a result of sameness (or imbalance) between the amounts on the two sides.

Not surprisingly, a number of studies have developed interventions using the pan-balance scale to situate the equal sign within a relatable context and to support students' success in solving equivalence problems. Some of these interventions have used a physical dynamic balance scale model with unitary weights in instruction (Alibali, 1999; Fyfe, McNeil, & Borjas, 2015; Mann, 2004; Perry, Berch, & Singleton, 1995), some have used a static balance scale model in pictures (Filloy & Rojano, 1989; Howden, 1994; Ketterlin-Geller, Jungjohann, Chard, & Baker, 2007; Linchevski & Herscovics, 1996; Vlassis, 2002; Warren & Cooper, 2005), some have used dynamic digital balance scales (Magruder & Mohr-Schroeder, 2013; Rojano & Martinez, 2009; Suh & Moyer-Packenham, 2007) and others have used balance gestures (Cui et al., 2017) to examine students' success with solving equivalence problems. Moreover, some of these interventions were mediated by an instructor, whereas some others were mediated digitally (e.g. using onscreen tutorials). These studies have varied in terms of participant age and the length of the intervention. Regardless, the findings from these studies highlight that demonstrating the act of balancing and employing balance-scale models aid upper elementary and middle-school students' understanding of mathematical equivalence (with the exception of work by Vlassis, 2002, who reported mixed results).

What these studies do not tell us is what feature(s) of the scale, when used in instruction, provide useful external supports for students' developing understanding of equivalence. The various aspects of external representations of equivalence all *seem* to work together during typical instruction that involves using the pan-balance scale to support learning. However, it is plausible that certain aspects – alone or in combination – may be responsible for instantiating the meaning of equivalence.

Drawing from prior work and our theoretical framing (see below), we identified four features of external support that we hypothesized would be conducive to learning when using a pan-balance scale.[2] First, we identified *objects to concretize quantities* and connect them with symbols in equations (e.g. 2 + 3 = 4 + 1 accompanied by 2 blocks and 3 blocks on one side and 4 blocks and 1 block on the other side). Second, because previous work indicates that reference to the balance scale has long been included in mathematics curricula (e.g. Presmeg, 1997) and might be useful for students (e.g. Mann, 2004), we included the *balance context*, by including two important relational terms that typically accompany the use of the balance scale and may contribute to learning outcomes: "same" (thereby highlighting the principle of making the two sides the same) and "balance" (thereby highlighting the analogy of balancing the two pans). Third, because reference to the two sides has supported student understanding of mathematical equivalence (Crooks, Alibali, & McNeil, 2011), we varied the *visual depiction of two sides*, to highlight that these would need to be made equivalent. Fourth, and finally, we identified the presence or absence of the *dynamic movement* of a balance scale (i.e. observing the scale moving from an imbalanced state to a balanced state or vice versa). In summary, by providing different instructional interventions that included or omitted these features, we sought to determine which of these were linked to learning outcomes for students who primarily held an operational view of the equal sign. Thus, in this study, we sought to address this gap of not knowing which aspects of the pan-balance scale support developing understanding of equivalence when using the pan-balance scale in instruction by identifying certain features, which we varied.

Statement of the problem

In the present study, we tested whether the condition that most closely represented an actual, dynamic physical balance scale, and thus was most grounded in terms of external representations in our intervention, led to the most learning compared to other, relatively less-grounded, representations, presented on a computer, by isolating four features using affordances of technology.

Theoretical framings

Overview

To understand which features of the pan-balance scale serve to help young students overcome their difficulty in solving equivalence problems and seeing the equal sign as a relational symbol, we first review theoretical underpinnings of using concrete experiences to support learning. Here, we highlight the role that concrete manipulatives play in promoting learning of abstract ideas and provide a review of related theoretical accounts, which argue for a closer examination of external supports to build student understanding of mathematical concepts. Next, we briefly discuss how affordances of technology can offer concrete experiences through digital renditions of concrete manipulatives, broadly known as digital or virtual manipulatives (e.g. Goldstone & Son, 2005; Manches & O'Malley, 2012), and share some evidence on the virtual pan balance as an effective tool to support student learning of equivalence.

Concrete experiences to support learning

Mathematics educators have long utilized concrete materials to facilitate meaningful abstraction of mathematics concepts and procedures (e.g. Johnson, 2000; Sowell, 1989). The underpinnings of using concrete experiences in instructional approaches is often traced to Bruner (1966), which follows from classic theories of cognitive development, in which learning begins as a sensorimotor process (e.g. Piaget & Inhelder, 1969). Bruner formally theorized a three-step instructional approach, which specified the role of concrete materials and contexts in developing students' understandings of abstract mathematical concepts and procedures. According to Bruner, new concepts and procedures should be presented sequentially in three forms, gradually increasing the abstractness of the representation for the intended targeted mathematical idea: *enactive, iconic*, and then *symbolic* form. In the enactive form, the learner is introduced to a physical, concrete representation of the concept so that the abstract concept is embedded in a concrete, non-symbolic, informal, and relatable situation. Thereafter, the idea should be presented in an iconic form, which is typically a pictorial representation of the enactive experience. And, finally, an abstract representation of the concept or idea should be presented using symbols. Bruner explained that starting with a concrete, physical form offers familiarity to the abstract idea and gradually removing this familiarity in subsequent, more abstract, forms allows for making meaningful connections and abstraction of the intended idea, thus supporting learning. We suspect that young students have heard the term "balance" and have had opportunities to experience balancing in everyday life, even if the pan-balance scale is novel to students.

More recently, other researchers (e.g. Fyfe, McNeil, Son, & Goldstone, 2014; Goldstone & Son, 2005; Lehrer & Schauble, 2002) have advanced and empirically examined Bruner's approach, by beginning with concrete materials and then gradually removing contextual elements to reveal more abstract representations. As an example, Fyfe et al. (2014) advocated for a "concreteness-fading approach" to instruction in mathematics and science (also see Pea, 2004, for additional theorizing on this point; Goldstone & Son, 2005, for additional empirical results on this point). They argued that the progressive order, of initially introducing a concept using concrete experiences and then gradually removing the concreteness in subsequent steps of instruction, allows for flexible understanding of the intended mathematical or scientific idea.

In a related theoretical approach, on the role of concrete representations and their abstract counterparts, Belenky and Schalk (2014) discussed a *continuum* between groundedness and idealness, as a way to theorize about learning through using external representations that provide a concrete framing for abstract ideas (and see Goldstone & Son, 2005, for a relevant discussion of this issue in the digital space using simulations). They defined representations that are relatively *grounded* as those that activate prior knowledge or past experience, which is grounded in a familiar context. They defined representations that are relatively *idealized* as representations that may not be related to, or provide, less familiar experiences (also see Hiebert & Carpenter, 1992). In the case of a pan-balance scale, we chose four features – objects to concretize quantities, the balance context incorporating relational language, visual depiction of two sides, and the opportunity to observe the dynamic movement of the pan-balance scale – and varied these within a familiar context. In the case of a pan-balance scale to represent equivalence, this means that the use of a dynamic balance scale, along with a relational narrative targeting sameness, would be relatively grounded and might be an optimal starting point because this would embed the abstract concept in a concrete, non-symbolic, and relatable situation.

We drew from Belenky and Schalk's ideas of groundedness and idealized representations to design our interventions. We tested how these features, presented in a virtual environment, impacted student understanding of mathematical equivalence. We posited that our most grounded intervention, in theory, should be most supportive to students who lacked the relational understanding of the equal sign. From this theoretical position, the condition with the most grounded support should lead to more learning than other, relatively less grounded supports.

Concrete experiences and digital "manipulatives"[3]

With the advancement of technology, the term "concreteness" has been applied to digital renditions of physical objects – broadly classified as digital or virtual manipulatives (Dorward, 2002). In this way, *virtual manipulatives* are representations of concrete objects that can be manipulated on computers or other electronic devices (e.g. Moyer, Bolyard, & Spikell, 2002; Sarama & Clements, 2009, 2016). Evidence suggests that digital representations may be more flexible because they are devoid of the "external noise" that may accompany handling actual physical materials, which can be responsible for increased cognitive load (e.g. Chandler & Sweller, 1991). As a bonus, digital environments can be programmed to allow for providing specific and immediate feedback to the student (e.g. Sarama & Clements, 2009). According to Sarama and Clements (2009), "computer manipulatives can have just the mathematical features that developers wish it to have and just the actions on it that they wish it to promote – and not additional properties that may be distracting" (p.148). Additional affordances include taking advantage of the contiguity principle of multi-media learning theory (e.g. Mayer & Moreno, 2002): by displaying the (digital) model or manipulative alongside its symbolic representation, students may connect these more readily than when they are separated, which typically is the case for physical manipulatives and their symbolic referents.

With respect to the balance scale, some evidence exists suggesting that using a virtual (dynamic) balance scale can be as effective as a physical (dynamic) balance scale (Suh & Moyer-Packenham, 2007) in coming to see the equal sign as a symbol of equivalence. Given this, and the fact that the digital space permits selective inclusion of aspects of groundedness, while linking the symbolic with the concrete representations, we took advantage of these affordances. We designed our interventions, identified four features that are typically conducive to learning when using the balance scale, and combined or varied these aspects of the pan-balance scale while keeping others constant across the instructional conditions, to examine how the features we incorporated into our conditions impacted students' understanding of equivalence.

Dissenting evidence about the benefits of manipulatives and real-life contexts

Although ostensibly sensible, and previous empirical research demonstrates that concrete experiences can be advantageous when they are grounded in familiar and meaningful contexts (Baranes, Perry, & Stigler, 1989; Carraher, Carraher, & Schliemann, 1987), the notion that teaching abstract ideas should begin with real-life contexts has also been problematized. The issue of concern about teaching with concrete supports for learning concepts that are abstract is that these interventions may draw the learner's attention to the manipulative rather than the referents (Kaminski, Sloutsky, & Heckler, 2009, 2013; Uttal, Scudder, & DeLoache, 1997), or may overload the learner's working memory, thereby increasing cognitive load (Mayer & Moreno, 2003), leading to failure of these materials to support learning. Moreover, mathematics educators have cautioned about the use of concrete manipulatives in teaching, warning that manipulatives do not speak for themselves and concreteness should not be interpreted as the ability to take physical actions on objects because students may be able to act on the manipulative physically without ever making a mental connection between their actions and the intended idea (Ball, 1996; Sarama & Clements, 2009). Some researchers have even found that students were unsuccessful in making connections between different representations – failing to transfer the arithmetic they performed using manipulatives to paper-and-pencil tasks (Thompson & Thompson, 1990).

A consequence of this debate is that it leaves teachers not knowing what might make the most sense for their students. In designing our conditions, we were mindful of these contradictory findings and utilized affordances of technology to link representations of mathematical equivalence both with mathematical symbols and with audio instructions, to investigate the impact of different types of groundedness on students' success in solving equivalence problems. In particular, our primary research question was: Does using a grounded balance scale (i.e. a dynamic balance scale) in the virtual space result in more learning when compared with less grounded representations of the balance scale, and do these result in more learning compared to not experiencing a grounded representation at all?

Method

Setting and participants

A total of 171 second- and third-grade students from four public elementary schools in the Midwestern United States participated in the study. Of those students, 20 were at the ceiling or near ceiling at the pretest (solving 5 or 6 out of the 6 pretest equivalence problems correctly; see Table 1); 2 were removed from analysis due to experimenter error (these students had the same first name and were tested under the wrong condition), and 1 did not complete the intervention. Thus, a total of 148 participants, with a mean age of 8.4 years (*SD* = .73), were considered for analysis. Of these, 56% were female; 52% were White, 27% Black, 9% Hispanic, 1% Asian, and 11% multi-racial; 77% of participants were eligible for free-and-reduced lunch; 10% were English language learners; and 10% had an individualized educational plan.

Research design and procedures

Each student participated in a total of three sessions on three separate days. In Session 1, students were screened in their classrooms on simple addition and subtraction problems, using paper and pencil. From the second session onwards, each student met individually during school hours either

Table 1. Equivalence problems used in the pretest, posttest, and delayed posttest.

Problem Type	Pretest	Posttest (and Delayed Posttest)
Practiced	2 + 5 = 4 + _	2 + 4 = 5 + _
	3 + 6 = _ + 8	3 + 5 = _ + 7
	7 + 2 + 5 = _ + 5	7 + 2 + 4 = _ + 4
	3 + 2 + 5 = 3 + _	5 + 4 + 3 = 5 + _
	4 + 2 + 7 = _ + 7	8 + 2 + 4 = _ + 4
	5 + 3 + 4 = 5 + _	2 + 3 + 5 = 2 + _
Transfer		6 + 2 + 3 = 7 + _
		8 + 3–2 = 8 + _
		5 + 1 + _ = 6 + 2

with the first author or with a trained research assistant (referred to hereonin as a *trainer*) and all tasks, unless otherwise noted, were carried out on Apple MacBooks.

Session 2 had three phases: (a) pretest, (b) intervention on the computer using a game context, and (c) immediate posttest. Before taking the pretest, students were randomly assigned to one of four interventions: (1) Control (n = 32), (2) Core Principle (n = 38), (3) Balance Analogy (n = 38), or (4) Dynamic Balance Scale (n = 40); we describe the interventions in detail in Materials. This session lasted approximately 60 minutes, including a brief break.

Students took a delayed posttest, held a week after Session 2, in Session 3.

Materials

Screening test

In Session 1, all students took a 16-item paper-and-pencil test: 6 true-false items, 2 practice items, 4 addition problems, and 4 subtraction problems (see the Appendix). The screening test included problems that are typically introduced in first grade (National Governors Association, 2012). The purpose of the true-false items in the screening was to rule out the possibility that students were already at ceiling and had a relational understanding of the equal sign. The purpose of including addition and subtraction problems was to rule out the floor effect, that a lack of competence with these basic operations may have resulted in a failure to solve mathematical equivalence problems correctly (Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011). The addition and subtraction problems were given after the true-false problems to avoid priming the students to solve all problems operationally. The test was not timed and lasted approximately 35 minutes. All participants, except one, passed at least 50% of the addition and subtraction problems and were included for the remainder of the sessions.

Pretest, immediate posttest, and delayed posttest

The pretest, immediate posttest, and delayed posttest (identical to the immediate posttest) consisted of equivalence problems with one unknown (see Table 1) and students' reasoning in finding the unknown.

The problems in the posttest were similar to those in the pretest, except that we added three transfer items (see Table 1). These transfer items required understanding that the sum on both sides of an equation had to be the same. These items were included because they differed in format; none of the addends on the left side of the equation were repeated on the right side (i.e. in 6 + 2 + 3 = 7 + _) and the other two items either used two operations (i.e. 8 + 3–2) or were missing an addend on the left side of the equation (i.e. 5 + 1 + _) (Lindvall & Ibarra, 1980).

In addition to the problems presented in Table 1, to assess students' explicit understanding of the meaning of the equal sign, students were presented with the equal sign symbol on the computer screen. The trainer pointed to the screen and asked the following questions, each separated by the student's response: (1) What is the name of this symbol? (2) What does this symbol mean? (3) Can it mean something else? Following the last question, students were presented with a piece of paper and asked (4) If you were to show or explain to a friend or someone who doesn't know what the equal sign means, how would you explain it to them? Could you draw or write something that you would use so they could understand what the equal sign means? The first three questions were based on those used by previous researchers (Baroody & Ginsburg, 1983; Knuth et al., 2006). The last question was added to elicit other potential understandings of the equal sign, if they existed.

At the posttest, to examine if students had any shifts in how they defined the equal sign after the intervention, they were asked: "After playing the game [intervention], do you think the equal sign can mean something else?"

To assess students' perceptions of the equal sign as a symbol separating the two sides of an equation (Matthews & Rittle-Johnson, 2009), students were shown four options (see the Appendix). The trainer said "Here is a number sentence and it has two sides." The trainer then pointed to each of the choices that depicted the possible separation of the two sides by a vertical line and said, "Some students thought this is how we should break the number sentence into two sides, so this (circling the side to the left of the vertical line) is one side and this (circling the side to the right of the vertical line) is the other side. Can you take a look at all of these choices and circle the choice that you think breaks the problem correctly into two sides?"

Intervention materials: overview

The intervention was delivered on a computer within the context of a game. The intervention consisted of one of three instructional conditions – each with different types of grounded support (see Table 2) – or one control condition. We designed the Control condition to be the most idealized and the Dynamic Balance-Scale condition relatively to be the most grounded. The other two in-between conditions (Core Principle and Balance Analogy) provided different sorts of groundedness.

The *Control* condition simply presented students with practice on the symbolic problems on the computer screen to find the missing number. The *Core Principle* condition focused on assessing whether the blocks that were present on the two distinct sides of the static table were the same. The *Balance Analogy* condition instructed students to make the two sides, represented by two separate pans with blocks, balanced so they became the same. The *Dynamic Balance-Scale* condition showed a dynamic balance scale, with the two sides of the scale attached through a fulcrum. The scale teetered up and down, based on which side had more blocks, then balanced when an equal number of blocks was present on both pans. The meaning of balancing a scale was discussed during the introduction phase for students in the Dynamic-Balance-Scale and Balance-Analogy conditions only.

Table 2. The features that appeared in each of the conditions.

		Features that *Varied* Across Conditions			
Condition	Symbols	Visual Presence of Blocks	Balance Context, *with relational language italicized*	Visual Depiction of 2 Sides	Movement Depicting Balancing
Control	e.g. 3 + 4 + 2 = 3 + __ (plus reading the symbolic problem)				
Core Principle	e.g. 3 + 4 + 2 = 3 + __ (plus reading the symbolic problem)	Yes	Make the 2 sides the *same*	A static table, no pans	
Balance Analogy	e.g. 3 + 4 + 2 = 3 + __ (plus reading the symbolic problem)	Yes	*Balance* the 2 sides so they *become the same*	A static table, with 2 pans	
Dynamic Balance Scale	e.g. 3 + 4 + 2 = 3 + __ (plus reading the symbolic problem)	Yes	*Balance* the scale	A pan-balance scale	The pan with the greater number of blocks goes down; if the same number in both pans, the scale teeters to balance

Empty cells indicate that the intervention condition did not contain that feature.
Also note that the symbols were consistent across all conditions.

The students in the Core-Principle and Control conditions were not shown a balance scale, nor was the term "balance" used with them.

Enacting the intervention: introducing the game, then playing the game

Introducing the game

Once the student put on the headset and the environment was launched, an animated character appeared and was introduced by the voice narration as a smart cat named Marty that likes to play with numbers. After this, students in the Balance-Analogy and Dynamic-Balance-Scale conditions saw a static pan-balance graphic as part of the introduction and were asked if they knew what a balance scale was (see Figure 1).

For the students in the Dynamic Balance-Scale condition, the narration (showing a balance scale on the screen) said that the scale gets balanced when the number of blocks in the green pan (the left pan) is the same as the number of blocks in the red pan (the right pan). The narration then continued and told the student that the green pan would be Marty's and the red pan would be the student's (with the student's name appearing on the red pan). The animation then dropped one block on the left side and three blocks on the right side, replacing the green equal sign on the fulcrum with a lock. Following this, the text and accompanying voice-over asked the student "if you put one block on Marty's side and 3

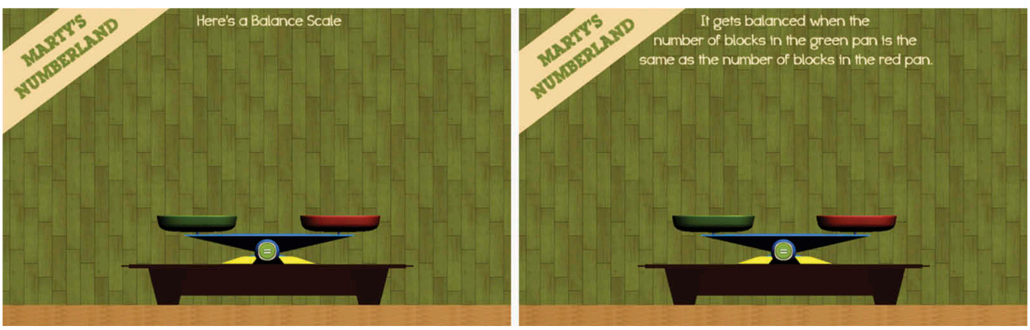

Figure 1. Introduction to the balance scale in the Dynamic Balance-Scale and the balance-Analogy conditions.

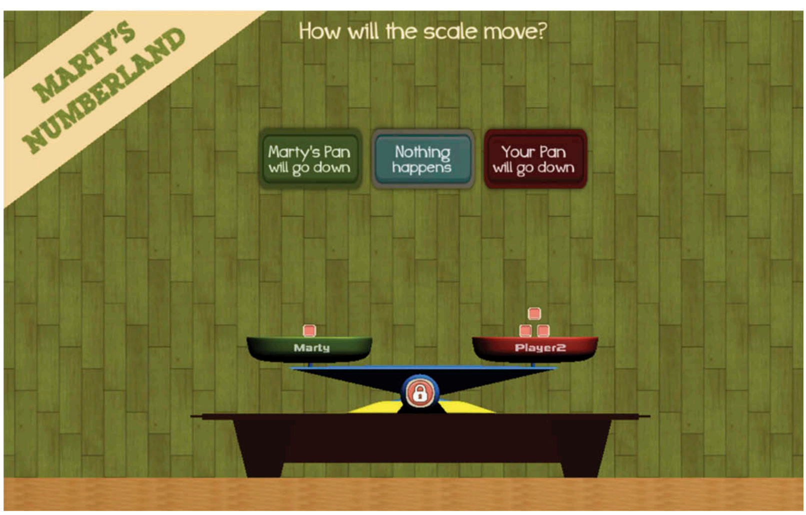

Figure 2. Initiation of the analogy in the Dynamic Balance-Scale and the Analogy conditions.

blocks on your side, how will the scale move?" The student was prompted to click on one of three options that indicated what would happen in the display (see Figure 2): Marty's pan will go down, nothing will go down, or your pan will go down. As the student moved the cursor, the student could hear the audio prompts. When students in the Dynamic Balance-Scale condition answered correctly, (1) the balance scale moved and the pan with more blocks went down as an animation, (2) the lock icon displayed at the fulcrum was replaced with a not-equal symbol (≠), and (3) the narration explained that the scale is not balanced because the green pan does not have the same number of blocks as the red pan, followed by "this side (zooming in on the left side) is not the same as this side (zooming in on the right side)." If the student clicked on an incorrect option, the narration along with accompanying visual cues offered why that choice would not make sense and prompted the student to try again. This loop continued until the student clicked on the correct option – the student's pan going down – followed by the audio and visual feedback for the correct response. Thus, the introduction phase allowed for familiarizing the students with the functioning of the virtual balance scale. This also assured that the student could follow the directions given through the audio prompt. In instances where a student stared at the screen or did not respond, the trainer restated the prompt and suggested that the student could move the mouse over the three responses and click on what they thought would happen to the scale.

Students in the Balance-Analogy condition received similar verbal information during the introduction and saw a static pan balance; however, they did not see the scale move. Students in the Core-Principle condition and the Control condition did not have access to the balance analogy and did not see the pan balance as a dynamic model or as a static graphic. All conditions, but the Control, were shown blocks. All conditions provided students with audio instructions to familiarize them with how to proceed. The students interacted with the computer screen using an external mouse to click or hover over the instructional prompts or to move to the next screen and this instruction was self-guided.

Playing the game

The intervention launched immediately after the introduction. All four versions had Marty, the animated cat, and a bingo machine, which generated the numbers for each side. Written prompts appeared on the screen; hovering the cursor over the prompt resulted in playing the audio accompaniment to the prompt. Students were told that they and Marty would play a game in which they each take turns using the bingo machine to get blocks for their sides. The prompt from the game told the students that the goal of the game was to help Marty balance the [scale, two pans, or two sides (based on the instructional condition)], or to find the number that should go in the blank (in the Control condition). The three instructional conditions (i.e. all except the Control condition) also showed two distinct sides and blocks to represent quantities on both sides during the instruction. The student and the animated character (Marty) took turns clicking on a bingo machine to draw out the numbers for their respective side of the pan-balance scale (on Marty's turn, the bingo machine rolled automatically). Marty only appeared on the screen once in the beginning and once after every turn ended. Thus, the role of Marty was simply to act as a character in charge of one side of the equation and then in the end to rejoice as the scale or the equation was balanced or made the

Table 3. Problems presented during the intervention: play session.

Problems
5 = _ + 3
2 + 5 = _ + 3
3 + 4 + 2 = 3 + _
4 + 5 + 2 = _ + 2
8 + 3 + 2 = 8 + _
2 + 6 + 3 = _ + 3

same. Although the numbers for each problem appeared to emerge randomly from the bingo machine on the screen, they were the same for each student (see Table 3, for a list of the problems). All students had two trials to find the number that should go in the blank and were given appropriate audio feedback if they picked an incorrect number. Screen shots of each of the four conditions are shown in Figure 3. Figure 3(a) shows the initial setup for each of the four conditions. For example, in the Dynamic Balance-Scale condition, students saw blocks resting on the pans of a balance scale. The problem appeared on the screen and the accompanying audio asked the student if the two sides of the scale were balanced. Figure 3(b) shows what happened next for each of the four conditions. Following the example for the Dynamic Balance-Scale condition, students were asked to balance the scale. Correct responses received auditory and dynamic visual feedback, as the scale adjusted to demonstrate balance (Figure 3(c)). The audio track verbalized the symbolic statements, for example, stating that "3+ 4 + 2 is the same as 3 + 6," followed by counting each collection, and "9 is the same as 9. 9 equals 9." The equal sign was animated, and zoomed in and out at the moment when the audio said "the same as" and "equals."

Instruction in the Balance-Analogy condition was comparable to the Dynamic Balance-Scale condition, except that the balance *scale* was absent and was replaced with a set of side-by-side pans (see column 2 in Figure 3(a)) to suggest symmetry or asymmetry and the corresponding prompt requested that students "balance the two sides" (column 2 in Figure 3(b)).

Instruction in the Core-Principle condition was comparable to the Balance-Analogy condition, except that the two pans were absent, and the prompt used relational language by asking participants to make "both sides the same" without any reference to "balancing" the two sides, which was the relational language used in the Balance-Analogy condition.

In the Control condition, the problems were presented only symbolically. Students were prompted to find the number that should go in the blank.

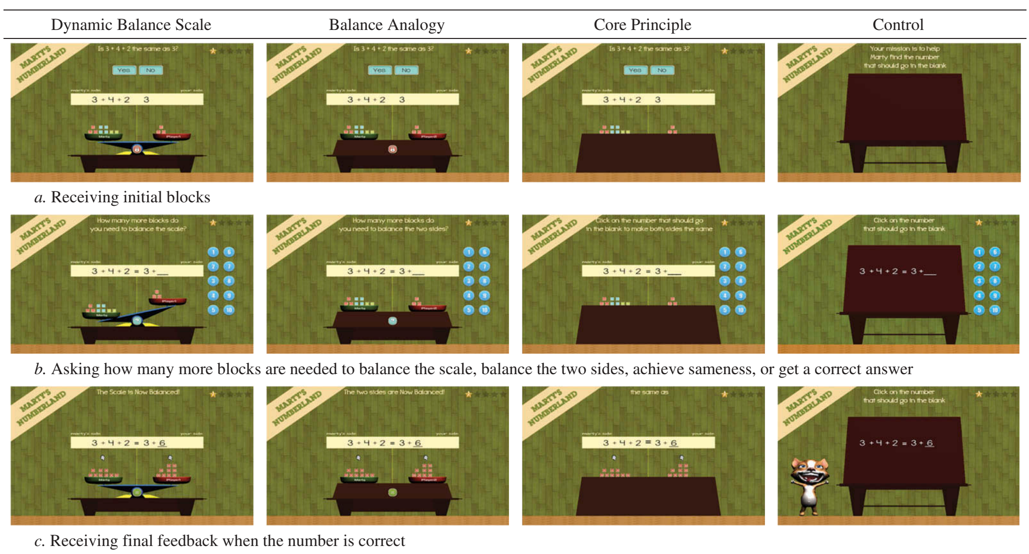

Figure 3. Instructional applet screenshots. (a) Receiving initial blocks. (b) Asking how many more blocks are needed to balance the scale, balance the two sides, achieve sameness, or get a correct answer. (c) Receiving final feedback when the number is correct. Conditions show different types of grounded support from left (most grounded) to right (least grounded, or most idealized).

Data sources and scoring

Students' responses, including their reasoning for how they arrived at their answers, served as data sources.

Scoring the screening test

The true-false, addition, and subtraction problems were scored as either correct or incorrect.

Scoring responses to equivalence problems

Equivalence problems were scored as correct if the student provided both a correct answer (we accepted answers that were ±1 from the answer as correct because we were interested in changes in relational understanding and not fluency in addition, see e.g. McNeil, 2007) and a relational strategy to justify how they attained the answer. To establish inter-rater reliability, two coders (blind to the condition) independently coded students' strategies as relational or not (for both the pretest and the posttest responses). Inter-rater agreement at the pretest (98%) and posttest (99%) was high. Because we used 6 problems involving one unknown, the scores for this type of problem ranged from 0 (no accurate response generated along with no relational reasoning) to 6 (all correct, with relational reasoning). Students also solved 3 transfer equivalence problems during both posttests (immediate and delayed). Thus, scores on transfer items ranged from 0 (for none correct) to 3.

Scoring responses for definition of the equal sign

We categorized students' responses to each of the four questions assessing their definitions of the equal sign based on a system used by McNeil and colleagues (Knuth et al., 2006; McNeil & Alibali, 2005a). Occasionally, students provided both a relational and an operational meaning; in these cases, responses were coded as relational because we had some indication that the student was working with a relational meaning. When students simply repeated the word "equal," without any additional information, to describe the equal sign, the response was coded as ambiguous because merely restating the term did not offer a clear indication of what views the student held of the equal sign. Inter-rater agreement on assigning a category to students' definition at pretest and at posttest were high (96% and 95%, respectively) and discrepancies were discussed and resolved. Table 4 provides some examples observed of students' responses.

Scoring responses for identifying two sides of an equation

Relying on Matthews and Rittle-Johnson's (2009) measure, we gave students multiple-choice items to indicate the two sides of the equation. Responses were coded as correct when the student circled the option that correctly specified the two sides. For example, in the problem 4 + 1 = 2 + 3, the correct option identified one side as 4 + 1 and the other side as 2 + 3, separated by a vertical bar (see the Appendix).

Results

Our primary research question was "Does using a grounded balance scale (i.e. a dynamic balance scale) in the virtual space result in more learning when compared with less grounded representations

Table 4. Student response categories indicating definitions of the equal sign.

Category	Examples
Relational	"It means the same as." "Same amount." "Like balanced" [with a balancing hand gesture]
Partial Relational	"It [means] the same number. So, 9 equals 9." "Like 54 and 54 are the same … you are supposed to put equal in it."
Operational	"It means total." "Where the answer goes." "It tells you the answer."
Ambiguous	"It can mean something else but I don't know how to say it … it means *eeequal* [stressing the emphasis on *e*]." "Like 2 plus 3 equals, or it could be opposite," " It is like (paused) equaling the numbers … could be equals anything"
Other	"I don't know," shook head–implying no, or gave an irrelevant response

of the balance scale, and do these result in more learning compared to symbolic problems, relatively devoid of any grounded representation?" To answer this question, first, we present results about whether the students in the four conditions were similar before the interventions. Second, we provide descriptive evidence about the effectiveness of the interventions on the outcome measures at the posttest and delayed posttest, with the groups combined. Third, to examine whether, as hypothesized, our most-grounded condition (Dynamic Balance Scale) led to more learning than the less-grounded (Balance Analogy and Core Principle) intervention conditions, we categorized students' improvement into levels of learning; we provide illustrative, qualitative data about changes in student understanding to highlight the nuances in student levels of learning; and we conducted ordered logistic regression to provide converging evidence about the varying effectiveness of the different conditions.

Comparability of the four conditions during the screening measure and the pretest

Screening measure

During the in-class screening test, students scored an average of 3.12 (SD = 1.33) correct out of 6 true-false problems, 3.17 (SD = 0.88) correct out of 4 addition problems, and 2.9 (SD = 0.96) correct out of 4 subtraction problems. One-way ANOVAs revealed that students' performance did not differ significantly across the four conditions in terms of mean scores on true-false [$F(3,144) = 0.69$, $p = .56$], addition [$F(3,144) = 2.05$, $p = .11$], or subtraction problems [$F(3,144) = 0.88$, $p = .45$].

Pretest: equivalence problems

At the pretest, students solved only 4% of equivalence problems correctly (a mean of .22 out of 6, SD = .65). Most (86% of 148 students) did not solve even one problem correctly[4]. However, students' performance on equivalence problems did not differ significantly across the four conditions, Welch's ANOVA [$F(3,100) = 0.3$, $p = .8$)].

Pretest: defining the equal sign

At the pretest, the overall percentage of students who gave a relational definition was low (8.1%), with the majority of students giving an operational definition (73%). The remaining students either provided an ambiguous definition (5.4%) or stated that they did not know what the symbol meant or gave a response that was not relevant (4.7%, categorized as "other"). Students' ability (or lack of ability) to define the equal sign relationally did not differ significantly by condition, Fisher's Exact (p = .90).

Pretest: identifying the two sides of an equation

At the pretest, only 10% of students correctly identified the two sides of the equation in both problems. The percentage of these students did not differ across conditions, Fisher's Exact (p = .1).

Table 5. Performance on each of the three components of the outcome measure, before the intervention, by instructional condition.

		Instructional Condition			
Component of the Outcome Measure	How Measured	Dynamic Balance Scale	Balance Analogy	Core Principle	Control
Equivalence problems	M correct (*SD*)	0.20 (0.52)	0.18 (0.51)	0.21 (0.74)	0.31 (0.82)
Defining the equal sign	% who defined = relationally	15.0%	15.8%	13.2%	25.0%
Identifying the two sides	% who got both correct	2.5%	7.9%	13.2%	18.8%

Table 5 provides students' pretest performances by condition on the three components of understanding the equal sign.

Summary of results for the screening measure and pretest

Taken together, these results indicate that students' performances did not differ significantly across the four conditions on any type of problems given during the screening measure or on the pretest. Thus, the four groups were comparable prior to the intervention.

Evidence of intervention effectiveness: combining across instructional conditions

To determine whether the interventions were effective, we evaluated students' performances at an immediate posttest. We also conducted a delayed posttest to determine whether student performance remained consistent or whether there was a significant drop-off after one week.

Immediate posttest: equivalence problems

Overall, students performed better at the posttest. Furthermore, students' performance on transfer problems correlated significantly and positively with their performance on practice problems (*rho* = .79, $p < .001$). Before doing any planned comparisons, because of both the consistency in performance between the practice problems and transfer problems, and our interest in promoting and documenting improvement, we analyzed students' performance on gain in practice (from pretest to immediate posttest) and transfer scores, combined. For ease of reference, we refer to these combined gains as "composite gain score" (see Table 6).

Immediate posttest: defining the equal sign

Students defined the equal sign relationally in greater percentages at the posttest compared to their attempts at the pretest, from 8.1% to 28.4% at the immediate posttest, McNemar's χ^2 [(1, 148) = 49.1, $p < .001$]; likewise, the percentage of students who gave an operational definition decreased from 73% to 34%.

Immediate posttest: identifying the two sides of an equation

The percentage of students who identified the two sides of an equation relationally increased significantly from the pretest (10%) to the immediate posttest (28%), McNemar's χ^2 [(1, 148) = 12.6, $p < .001$]. These results indicate that students' performance improved at the immediate posttest.

Table 6. Performance on immediate and delayed posttest by instructional condition.

	Instructional Condition							
	Dynamic Balance Scale (n = 40)		Balance Analogy (n = 38)		Core Principle (n = 38)		Control (n = 32)	
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Immediate posttest practice problems	2.50	2.44	3.71	2.50	2.76	2.63	0.78	1.62
Immediate posttest transfer problems	0.50	0.78	0.84	0.89	0.82	0.95	0.19	0.54
Delayed posttest practice problems	2.46	2.53	3.20	2.75	2.80	2.56	1.05	1.79
Delayed posttest transfer problems	0.51	0.81	0.84	0.88	0.74	0.91	0.25	0.56
Gain on immediate posttest practice problems	2.30	2.42	3.53	2.52	2.55	2.56	0.47	1.22
Composite gain on immediate posttest*	2.80	3.05	4.37	3.19	3.37	3.39	0.66	1.56
Composite gain on delayed posttest*	2.77	3.21	4.12	3.52	3.33	3.34	1.00	2.24

*This score is a composite of gains on the practice problems plus transfer problems solved correctly

Delayed posttest

We found no significant differences between performance on the immediate posttest and performance on the delayed posttest (Wilcoxon rank-tests, all *p*s>.05), suggesting no meaningful decrement in performance after one week.

Examining the effects of condition on levels of learning

Improvement on equivalence problems at the posttest

Although, in general, students improved after the intervention, they did not do so equally across all four conditions. Descriptive results indicate that the mean score for students in the Balance-Analogy condition were the highest, followed by students in the Core-Principle condition, and then students in the Dynamic Balance-Scale condition. The mean scores of students in all three instructional conditions were higher at the posttest than those in the Control condition and statistical analyses revealed significant differences in gain scores across the conditions, Welch's $F[(3, 78.13) = 17.577, p < .001]$: students in the three instructional conditions gained significantly more than students in the Control condition (Games-Howell post-hoc comparison, $p < .05$, see Table 6).

Categorizing levels of learning on equivalence problems

The frequency distribution of students' performance on equation-solving was skewed, which is typical in studies of children's performance on such problems (McNeil, 2007). This is usually handled by categorizing students' responses binomially in terms of whether or not students solved at least one of the problems correctly. Although this allows researchers to assess whether students learned or not, it does not allow us to evaluate *gradations* of learning. Therefore, given the nature of the data and the precedent set by previous research (e.g. McNeil, 2007; McNeil, Rittle-Johnson, Hattikudur, & Peterson, 2010; Perry, Church, & Goldin-Meadow, 1988), we examined both logical and empirical ways to classify outcomes into levels of learning. We classified students' performance ordinally into three levels based on the extent to which the students correctly solved the practiced and transfer problems: (a) *none*, demonstrating no gain on the practiced problems nor any correct on the transfer problems (n = 61); (b) *emerging*, demonstrating *some gain*, but not more than a combined gain of four on practice and transfer problems (n = 42); and (c) *adequate*, demonstrating a minimum *gain* of at least five on the composite scale, with, in particular, *a gain of at least four on practice problems and at least one on transfer problems* (n = 45). We settled on this cutoff because this would likely rule out the possibility that the students guessed or that their correct responses were due to chance. In addition, the cutoffs emerged empirically; the vast majority of students who had a gain of *at least four* at the posttest (*adequate* improvement) also answered at least one transfer problem correctly (n = 45 out of 53), which was not the case for the vast majority of students who solved *up to four practice problems alone* (only 9 out of 36 students who did not answer at least four posttest problems correctly also answered a transfer problem correctly, indicating that their performance was, indeed, *emergent*). These criteria allowed for the examination of nuanced patterns of learning with respect to the different conditions[5].

Illustrative cases of student responses

Next, we provide some typical, illustrative cases of 6 students' responses, to highlight the differences across these three levels of learning (2 students at *adequate*, 2 at *emergent*, and 2 at *none*). In each case, we provide a description of students' reasoning and strategy at the pretest to the problem $7 + 2 + 5 = \_ + 5$ and then their reasoning and strategy at the posttest for the comparable problem $7 + 2 + 4 = \_ + 4$. Occasionally, we provide detail on additional problems, for clarification. We note that some students in all three instructional conditions provided relational reasonings for some problems (indicating that some instruction was better than business-as-usual, represented by the Control condition). As a reminder, the interventions did not provide instruction on how to solve the problem, leaving students to use their own strategies in making the scale balanced, making two sides balanced, or making both sides the same.

Adequate learning: Elena. At the pretest, Elena arrived at 14 for her answer (in response to the prompt “what number should go in the blank to make this number sentence true?”). In response to the follow-up prompt of “you put 14 in the blank, can you tell me how you figured out that 14 should go in the blank?” she was quick to say “because 7 plus 2 plus 5 equals 14.” At the posttest, Elena put 9 in the blank. She responded “because 7 plus 2 plus 4 equals 13. And 9 (pause) … 4 does not equal 13, so we need to add 9 to make 13 to make it the same.”

Adequate learning: Kaleb. At the pretest, Kaleb summed all the addends before the equal sign, “I added 5 plus 2 … equals 7 and 7 plus 7 equals 14, so 14.” At the posttest, he said “7 plus 2 plus 4 is 13 and … I added 7 plus 2 together … is 9. There is already 4 on both sides, so it would be … then it would be 9 + 4.” Kaleb noticed the equal addends, so he summed together the addends on the left of the equation that did not appear on the right side of the equation.

Emergent learning: Amy. At the pretest, Amy said the answer was 14 and provided the following reason: “I counted in my head … I counted these (pointing to the numbers on the left side of the equal sign).” In defining the equal sign she said, “like find out the answer … you find it and make it,” which indicated an operational view of the equal sign. At the posttest, she solved 3 of the 6 problems correctly. Amy stated that the solution was 9, but she did not provide clear evidence about why 9 should correctly solve the problem. She said “you can count these (pointing to 7 and 2) and solve the numbers by using tens.” On the next problem, 5 + 4 + 3 = 5 + _, she reverted to adding all of the numbers that preceded the blank and said that 17 should go in the blank. Her response “because … 9 is bigger than 17” did not provide a clear reasoning. She also solved all of the transfer problems incorrectly. Note that we categorized this student as emergent because she demonstrated some, but not enough to be considered adequate, relational reasoning.

Emergent learning: Camilla. Camilla did not solve any problem correctly on the pretest and demonstrated an operational understanding when interviewed about her understanding of the equal sign. On the pretest she said “because 1 counted 5 and 7 … 7 plus 5 equals 12 and plus 2 … and then plus 2 is 14.” On the posttest, she gave 13 as her solution, which was incorrect. However, she solved three other problems correctly on the posttest, when the blank was in the final position in the problem, but solved all other problems incorrectly.

Both Camilla and Amy, and the other students whom we classified as emerging, showed some gain at the posttest over their pretest performance, and their solutions, reasons, and perceptions of the equal sign showed emerging – but not solidly convincing – signs of being relational; they did not have a composite gain of at least 5 problems, including at least 1 transfer at the immediate posttest.

None (no evidence of learning): Sasha. At the pretest, Sasha gave 13 as the answer and she responded to the prompt to explain how she figured out that 13 was the answer by saying “7 8 9 10 11 12 and … .” She defined the equal sign by using an example, “5 plus 5 eeequals to 10.” During the intervention, after getting a problem incorrect using the balance scale, she figured out that, to balance the scale, both sides needed to have the same number of blocks and she was able to work through the problems presented in the game. However, at the posttest, she resorted to adding all the numbers and relied on the strategy she had used before, “7 8 9 10 11 12 13 … 13 (pointing to the blank).”

None (no evidence of learning): Abby. Like Sasha, Abby had an operational view at the pretest and responded to the problem by stating, “I did this in my head. I said 7 plus 2 equals 9 and then 9 plus 5

Table 7. Percentage of participants in each level of learning by condition at the posttest.

	Instructional Condition			
Levels of Learning	Dynamic Balance Scale (n = 40)	Balance Analogy (n = 38)	Core Principle (n = 38)	Control (n = 32)
None*	45%	16%	37%	63%
Emerging	28%	37%	24%	34%
Adequate	28%	47%	40%	3%

**Note*. None is the same as no learning equals 14." During the Dynamic Balance-Scale intervention, unlike Sasha, Abby seemed quick to get how many blocks were missing to complete the number sentence. However, at the posttest, Abby, much like Sasha, added together all of the addends preceding the equal sign. When prompted, she said "7 plus 2 equals 9 and 9 plus 4 equals 14."

As is clear from these cases, we found that students made varying degrees of progress in their understanding of equivalence. Table 7 shows the percentage of students in each of the four conditions who exhibited different levels of learning. We found that students' level of learning was related to condition, $\chi^2(6, 148) = 24.07$, $p < .001$.

To determine whether different conditions were systematically related to different levels of learning, and to identify influences of demographics on the levels of learning, we conducted an ordered logistic regression with three levels of the outcome variable (i.e. none, emerging, and adequate) and the intervention condition as a factor with four levels (i.e. Dynamic Balance Scale, Balance Analogy, Core Principle, and Control). We chose ordered logistic regression because: 1) the distance between the three categories of our outcome variable may not be the same (i.e. the distance between "none" and "emergent" might not be the same as the distance between "emergent" and "adequate"), and 2) it allows us to predict what factors influence the levels of learning. To conduct our planned comparisons, we considered two models, one with the Control condition as the reference, to contrast our most idealized condition with each of the other three conditions that were grounded in one way or another. In our second model, we used Dynamic Balance Scale condition as the reference, to contrast our most grounded condition with relatively less grounded conditions, combined, to examine our primary research question.

Ordered logistic regression, with the control condition as the reference

We first included age, gender, socio-economic status (SES, based on eligibility for free or reduced lunch; 1 = eligible, low SES, and 0 = not eligible, high SES), and English Language Learner (ELL) status (0 = not ELL; 1 = ELL) in the model as predictors. We included these demographic features to heed the call that results from research on technology and education should consider and report demographic groups in the analysis and interpretation of results (e.g. see Darling-Hammond, Zielezinski, & Goldman, 2014). Gender, age, and ELL status were not significant predictors in the base model, as assessed by the *p*-values in the model.

We then compared different models to identify goodness of fit and used the Akaike information criterion (AIC) as a measure of the goodness of fit, because it is a robust indicator for the best model. See the Appendix for detailed results from the ordered logistic regression.

The results revealed that students in the Balance-Analogy condition were 8.32 times more likely than students in the Control condition to show *adequate* versus *emerging or no* learning, when all other variables were held constant ($p < .001$). Likewise, the odds of students reaching either *emerging or an adequate level of learning* when compared to *not learning* was 8.32 times greater for students in the Balance-Analogy when compared to students in the Control condition ($p < .05$). Students in the Core-Principle condition were 4.06 times more likely than students in the Control condition to show *adequate* versus *emerging or no* learning, when all other variables were held constant ($p = .003$). Students in the Dynamic Balance-Scale condition were 2.69 times more likely than students in the Control condition to show *adequate* versus *emerging or no* learning, when all other variables were held constant ($p = .03$). Given that each of the instructional conditions was more likely to lead to learning than the Control, we next turn to examine whether our most grounded support, as provided in the Dynamic Balance-Scale Condition, led to more learning compared to the less-grounded conditions (Balance Analogy and Core Principle, combined).

Ordered logistic regression with the dynamic balance-scale condition as the reference

We conducted ordered logistic regression with the Dynamic Balance-Scale condition as the reference and compared it to the Balance-Analogy and Core-Principle conditions. The results (see the Appendix) revealed that students in the Balance-Analogy and Core-Principle conditions were 2.2 times more likely than students in the Dynamic Balance-Scale condition to show *adequate* versus *emerging or no* learning, when all other variables were held constant (p = .03). Likewise, the odds of students reaching either *emerging or an adequate level of learning* when compared to *not learning* was 2.2 times greater for students in the Balance-Analogy and Core-Principle conditions when compared to students in the Dynamic Balance-Scale condition (p = .03).

Discussion

The current study examined, in a randomized controlled experiment, whether certain features (i.e. of grounded support including dynamic movement) associated with the pan-balance manipulative, when presented on a computer screen within the context of a game, impacted young students' development of a relational understanding of the equal sign. The findings of the current study suggest that offering young students an opportunity to interact with a fully articulated, dynamic virtual pan balance was not as effective as offering young students either of the two relatively less-grounded versions. By separating out the components of the pan balance with animated, virtual versions, this study suggests that certain features associated with the manipulative are conducive to supporting young students' learning about the relational meaning of the equal sign (specifically, the visual presence of the blocks and relational language) and have advantages over the dynamic version of the virtual pan-balance model, which included observing movement. However, we also acknowledge that, in the case of this intervention, we chose dynamic movement as a distinguishing component of the most grounded condition, so we cannot make definitive claims about whether it is the amount of groundedness or the dynamic movement that might have had an impact on student learning.

In the remainder of this section, we further discuss some of the findings: (1) young children can be taught the relational meaning of the equal sign in a brief intervention with a virtual pan balance; (2) the results may have been impacted in particular ways because the instruction was delivered on a computer; and, most significantly (3) the condition providing the most grounded support with dynamic movement in our intervention did not lead to the most impressive learning outcomes. After discussion of these findings, we suggest implications for teaching and learning, note some limitations of this particular work, and end with conclusions.

Young children can be taught the relational meaning of the equal sign

Results indicated that instruction was beneficial for early elementary school students to support their relational understanding of the equal sign. We were not the first to demonstrate positive effects of instruction on students' understanding of the equal sign. Thus, the findings from the current study amplify findings from previous studies (Matthews & Rittle-Johnson, 2009; McNeil, Fyfe, Petersen, Dunwiddie, & Brletic-Shipley, 2011; Perry, 1991; Rittle-Johnson & Alibali, 1999) that have employed traditional one-on-one instructor-guided interventions and demonstrated that relatively small changes in instruction can help children understand the equal sign relationally.

Considering the effects of instruction delivered on a computer

A distinguishing feature of the current intervention was that it was delivered in a multimedia environment. According to Mayer's (Clark & Mayer, 2016; Mayer, 2005) multimedia theory of learning, information that is presented in both auditory and visual forms in close proximity supports learning by reducing cognitive load. It may be that introducing the visual symbol (=) and the relational audio message ("the same as") in sync, which was an affordance of the computer environment, likely contributed to students' improved performance. Some research has indicated that there are two particular advantages to computer renditions of physical manipulatives: (1) the flexibility available in maneuvering the manipulative, which emerges from the fact that "clumsiness" is avoided in the case of virtual manipulatives (e.g. Thompson, 1992), and (2) the connection between the symbols and the actions (on the virtual manipulative or image) because the feedback is immediate and can be directly linked to the object of study (Clements, Battista, & Sarama, 2001;

Sarama & Clements, 2009). The overall effectiveness of this brief intervention could be due in part to presenting the instruction in a multimedia format on the computer and including the presence of visual blocks and relational language. Thus, although we acknowledge that it is not the only way, at the very least, these findings suggest that the focus on the two sides and the relationship between the two sides may be crucial in shifting students' view of the equal sign as relational.

The most support may not be the best

Extending Bruner's ideas on concrete experiences in learning mathematics, coupled with recent findings on the concreteness-fading approach to instruction (Fyfe et al., 2015; McNeil & Fyfe, 2012), and prior research that has employed a pan-balance (e.g. Mann, 2004), we identified types of concreteness in a dynamic pan-balance scale. Drawing from Belenky and Schalk (2014), we intentionally combined features to examine the effect of relatively more grounded interventions with less grounded, and more idealized ones. This was important, given conflicting findings: some arguing and providing evidence that abstract instantiations are better for promoting learning and others claiming that concrete objects can be distracting or may exacerbate cognitive load (e.g. Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009).

Given the evidence that the pan-balance scale is an appropriate manipulative for teaching about the equal sign and given the successful interventions that have used this manipulative, we hypothesized that "idealized representations may make it more difficult for students to perform well … Students will likely struggle to perform because of the difficulties inherent in creating appropriate, useful, and efficient internal knowledge representations" (Belenky & Schalk, 2014, p. 34). Indeed, we found that the most idealized (Control) condition was the worst compared to the three conditions that incorporated grounding, but we did not find that the relatively idealized representations among these were the least effective nor did we find that the representation including the most grounded feature of dynamic movement was the most effective.

Because the results from our investigation did not support the prediction that the most concrete representation of the pan-balance scale – the dynamic pan balance – should provide the best support for students who struggle with equations with operations on both sides of the equal sign, we present an alternative, based on Hiebert and Grouws's (2007) construct of productive struggle. Hiebert and Grouws, relying on Vygotsky's (1978) discussion of the zone of proximal development, indicated that students should gain the most when the level of difficulty presents "enough challenge, so there is something new to figure out" (p. 388). It seems to us that, with the more grounded, Dynamic Balance-Scale condition, there was much less for students to figure out compared to the relatively more idealized (Core Principle and Balance Analogy) conditions. In other words, the students in the Dynamic Balance-Scale condition had direct access to what it means to "balance" because these students were witnessing the dynamic balancing of the scale.

Hiebert and Grouws (2007) went on to lament, "There are few principles, however, from theory or data, that speak to the degree of structure needed to facilitate (or undermine) the productive effects of students' struggle" (p. 390). At the very least, we have presented data that might shed light on a principle that speaks to the degree of structure, or groundedness, needed to facilitate the productive effects of a student's struggle: Given our results, we suspect that instruction that leads students to the core concepts should be optimal in leading students to the sort of productive struggle that results in new understanding. This implies that when we provided dynamically grounded support for students, this may have minimized the work that students had to do, thus potentially limiting them from being in the position to create new knowledge (also see Sarama & Clements 2009 work on integrated-concrete knowledge). In other words, it may be necessary for students to engage in productive struggle (Hiebert & Grouws, 2007) for them to develop deeper understandings of important mathematical concepts.

Finally, we found that any of the contextualized instructional interventions we offered was better than the minimal intervention provided in the Control condition with only the symbolic problems, where students only received feedback in terms of correctness or incorrectness of their solutions.

This finding was not surprising, as one would expect students' understanding of an abstract concept, such as equivalence denoted by the equal sign, to be facilitated by some concrete representation as opposed to solely a symbolic one (Caglayan & Olive, 2010; Sherman & Bisanz, 2009).

Implications for teaching and learning

An implication of this study is that, although models and manipulatives may be useful, educators need to evaluate particular manipulatives' pedagogical effectiveness, especially with respect to the inherent ideas of those models. Examining this issue was foreshadowed by Fyfe et al.'s (2014) concerns: "Given the widespread use and endorsement of concrete materials by researchers and teachers alike, it is pertinent that we find optimal ways to use these materials to facilitate both learning and transfer" (p. 22). Given the limited time and resources that teachers have in their classrooms, it is beneficial to know what features of instruction with manipulatives offer what sorts of potentials for learning. For teachers, the findings from our investigation indicate that, at times, providing seemingly helpful manipulatives may not allow students to reflect and gain new understanding or, worse, give the illusion that students are learning when they may not be and, thereby may potentially hinder learning. Furthermore, a model may contain different types of groundedness and some features can be leveraged to promote greater learning than others.

As practical guidance for teachers, when extended practice with a manipulative is not possible due to practical or other limitations, teachers should consider introducing the abstract concepts by providing some contextualized support, but not the most grounded (i.e. with dynamic movement), and build instruction from there to allow students to reflect on understanding the idea. Likewise, instructional designers should also take heed: It may be more effective to *not begin* with a completely embellished manipulative – even if it appears intuitively to be the most appealing. Uttal et al. (2009) and others (e.g. Kaminski & Sloutsky, 2013; Kaminski et al., 2009) have made strong arguments for bare-boned representations to remove distraction. It is also possible that the most grounded and embellished representation may add to cognitive load (e.g. Chandler & Sweller, 1991). The work presented here supports this but goes further by presenting instruction on the computer, which limited the possibility of mishandling an actual, physical manipulative (e.g. Kaminski & Sloutsky, 2013), or the possibility of missing other crucial aspects necessary to solve equivalence problems, for example the connection between the visual representation of the blocks and the translation to the symbolic form (in the equations).

Limitations

One of the limitations of this work is that we focused on the relational meaning of quantitative equivalence of the two sides of the equal sign. The complete relational meaning of the equal sign is a complex mathematical construct and it can imply different relationships based on the format of the problem. We focused on this meaning because it is crucial for elementary students' performance and understanding of arithmetic relationships and also one of the initial steps toward developing algebraic reasoning. However, we did not explore the full extent of students' understanding of the equal sign or a complete relational understanding (e.g. see Jacobs et al.'s 2007 work for a broader meaning of relational reasoning). More work on the different relational meanings would further help in demarcating what other relational ideas the dynamic pan balance successfully promotes.

Furthermore, we did not allow students to experience concreteness fading: They did not receive the most enriched representation and then have this faded, which others have found to produce successful learning outcomes (e.g. Fyfe et al., 2015; Goldstone & Son, 2005). Future work in the virtual space that starts with the relatively most grounded representation and then fades through the analogy to a completely idealized symbol would add to the discussion of how to best use virtual grounded manipulatives to support student understanding.

Our sample was limited. As an example, although the students were diverse in some ways (e.g. SES), they were not diverse in others (e.g. all were from the Midwest, few were African American, and more second-grade students participated than third-grade students). Thus, there is a need to see how varied types of groundedness of a manipulative provide support for learning with a wider age range, and with other groups of students.

Additionally, we did not explicitly examine students' prior experiences with the *balance analogy*. For example, we did not have any evidence regarding students' experiences with playing on a seesaw or using the term "balance" in their day-to-day lives (we assumed that most students would have experiences balancing and playing on a see-saw). Future investigations should attempt to collect this information to examine if background knowledge or previous experience might impact receptivity to the manipulative or the analogy. We provide a cautionary note along with this suggestion: Asking students about their previous experiences with balancing may also sensitize students toward the instruction and therefore needs to be handled carefully.

Conclusion

The study reported herein presents an empirical examination of a theoretical construct – types of groundedness and their effects on learning – specifically in using a pan-balance scale to help students understand mathematical equivalence. The findings indicated that our model that provided maximum groundedness (with dynamic movement) did not promote learning as well as the models that were grounded in a non-dynamic manner. Thus, we posit that deeply grounded support, while intuitive, may not be ideal in promoting optimal learning of equivalence. Specifically, although the virtual dynamic pan-balance scale might foster learning, this more fully explicated model might take away some of the crucial intellectual work that learners need to do to make sense of a concept. By providing the core principle of sameness, along with a visual depiction of quantities, or by providing that along with the analogy of balancing, may provide just the right level of support to make sense of equivalence. In sum, providing opportunities for young students to wrestle with equivalence – without doing all of the work for them – may be just the right balance for helping students learn.

Notes

1. We use the term *dynamic* to indicate *movement* of the pan balance; this is different than how the term is used elsewhere, e.g. in dynamic environments, such as in *Geometer's Sketchpad*.
2. We note that we have identified four features that typically surface when using a pan-balance scale in instruction, which we thought were likely to have an effect on student learning, although we acknowledge that there might be other features that can have an effect on student outcomes. For example, we employ objects that look the same and weigh the same employing different weights may have a different (or similar results), as in Barlow and Harmon's (2012) work.
3. We use the term *manipulative* broadly. Although some might argue that the term *manipulative* should be reserved for things that can be handled (either with hands, or virtually, with, e.g. a mouse), following Manches and O'Malley (2012), we use the term *digital manipulative* because we wanted to be transparent about the parallels we make between the actual pan-balance scale and its digital equivalent.
4. Much of our data were skewed and not distributed normally. In those cases when data violated assumptions for normality (Shapiro-Wilk, $p < .001$) and the Levene's test for homogeneity of variances revealed unequal variances ($p < .001$), we conducted non-parametric tests, including Welch's ANOVA, McNemar's test, and Wilcoxon signed-rank test, which are appropriate (and robust) tests in such cases.
5. We also conducted analyses with data treated continuously and obtained comparable results. We report the categorical data because these illustrate the variation in the extent of student learning more clearly.
6. These questions are adapted from Matthews et al. (2012).
7. These questions are adapted from the questions by Rittle-Johnson et al. (2011). A trainer read a prompt and explained the task as described in the text.

Acknowledgments

The research reported in this article was supported by a grant from the Hardie Dissertation Fellowship program in the College of Education at the University of Illinois at Urbana-Champaign to the first author. We thank Lyn English and the reviewers whose insight helped shape and clarify this article. The authors also appreciate valuable feedback received from Sarah T. Lubienski, Robb Lindgren, Dan Hoffman and the STEM Development Research Group, with special thanks to Shuai Wang for his statistical expertise, on an earlier version of this work. Finally, we are grateful to the students who participated in this project; without them, none of this would have been possible.

References

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*, 127–145. doi:10.1037/0012-1649.35.1.127.

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 performance solving equivalent equations. *Mathematics Thinking and Learning*, *9*, 221–247. doi:10.1080/10986060701360902.

Alibali, M. W., & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 349–365). Mahwah, NJ: Erlbaum.

Ball, D. L. (1996). Teacher learning and the mathematics reform: What we think we know and what we need to learn. *Phi Delta Kappan*, *77*, 500–508.

Baranes, R., Perry, M., & Stigler, J. W. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction*, *6*, 287–318. doi:10.1207/s1532690xci0604_1.

Barlow, A. T., & Harmon, S. E. (2012). Problem contexts for thinking about equality: An additional resource. *Childhood Education*, *88*, 96–101. doi:10.1080/00094056.2012.662121.

Baroody, A. J., & Coslick, R. T. (1998). *Fostering children's mathematical power: An investigative approach to K-8 mathematics instruction*. Mahwah, NJ: Erlbaum.

Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal*, *84*, 199–212. doi:10.1086/461356.

Behr, M., Erlwanger, S., & Nichols, E. (1980). How children view the equal sign. *Mathematics Teaching*, *92*, 13–15.

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review*, *26*, 27–50. doi:10.1007/s10648-014-9251-9.

Blanton, M., Otálora, Y., Brizuela, B. M., Gardiner, A. M., Sawrey, K. B., Gibbins, A., & Kim, Y. (2018). Exploring kindergarten students' early understandings of the equal sign. *Mathematical Thinking and Learning*, *20*, 167–201. doi:10.1080/10986065.2018.1474534.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Byrd, C. E., McNeil, N. M., Chesney, D. L., & Matthews, P. G. (2015). A specific misconception of the equal sign acts as a barrier to children's learning of early algebra. *Learning and Individual Differences*, *38*, 61–67. doi:10.1016/j.lindif.2015.01.001.

Caglayan, G., & Olive, J. (2010). Eighth grade students' representations of linear equations on a cups and tiles model. *Educational Studies in Mathematics*, *74*, 143–162. doi:10.1007/s10649-010-9231-z.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic & algebra in elementary school*. Portsmouth, NH: Heinemann.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1987). Written and oral mathematics. *Journal for Research in Mathematics Education*, *18*, 83–97. doi:10.2307/749244.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332. doi:10.1207/s1532690xci0804_2.

Chesney, D. L., McNeil, N. M., Matthews, P. G., Byrd, C. E., Petersen, L. A., Wheeler, M. C., Fyfe, E. R., & Dunwiddie, A. E. (2014). Organization matters: Mental organization of addition knowledge relates to understanding math equivalence in symbolic form. *Cognitive Development*, *30*, 30–46. doi:10.1016/j.cogdev.2014.01.001.

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Hoboken, NJ: Wiley.

Clements, D. H., Battista, M. T., & Sarama, J. (2001). Logo and geometry. *Journal for Research in Mathematics Education. Monograph*, *10*, i–177. doi:10.2307/749924.

Crooks, N. M., Alibali, M. W., & McNeil, N. M. (2011, March). *Highlighting relevant problem features improves encoding of equivalence problems and leads to the generation of correct strategies*. Poster presented at the Biennial Meeting of the Society for Research in Child Development, Montreal, Canada.

Cui, J., Popescu, V., Adamo-Villani, N., Cook, S. W., Duggan, K. A., & Friedman, H. S. (2017). Animation stimuli system for research on instructor gestures in education. *IEEE Computer Graphics and Applications*, *37*(4), 72–83. doi:10.1109/MCG.2017.3271471.

Darling-Hammond, L., Zielezinski, M. B., & Goldman, S. (2014). *Using technology to support at-risk students' learning*. Stanford Center for Opportunity Policy in Education. Retrieved from https://edpolicy.stanford.edu/publications/pubs/1241

Dorward, J. (2002). Intuition and research: Are they compatible? *Teaching Children Mathematics*, *8*, 329–332.

Falkner, K. P., Levi, L., & Carpenter, T. P. (1999). Children's understanding of equality: A foundation for algebra. *Teaching Children Mathematics*, *6*, 232–236.

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics*, *9*(2), 19–25.

Fyfe, E. R., McNeil, N. M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction*, *35*, 104–120. doi:10.1016/j.learninstruc.2014.10.004.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*, 9–25. doi:10.1007/s10648-014-9249-3.

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, *14*, 69–110. doi:10.1207/s15327809jls1401_4.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 65–97). New York, NY: MacMillan.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.

Howden, H. (1994). *Algebra tiles for the overhead projector*. Vernon Hills, IL: ETA/Cuisenaire.

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*, 258–288.

Johnson, V. M. (2000, April). *An investigation of effects of instructional strategies on conceptual understanding of young children in mathematics*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Kaminski, J. A., & Sloutsky, V. M. (2013). Extraneous perceptual information interferes with children's acquisition of mathematical knowledge. *Journal of Educational Psychology*, *105*, 351–363. doi:10.1037/a0031040.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2009). Concrete instantiations of mathematics: A double-edged sword. *Journal for Research in Mathematics Education*, *40*, 90–93.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. *Journal of Experimental Psychology: Applied*, *19*, 14–29. doi:10.1037/a0031931.

Ketterlin-Geller, L. R., Jungjohann, K., Chard, D. J., & Baker, S. (2007). From arithmetic to algebra. *Educational Leadership*, *65*, 66–71.

Kieran, C. (1980). The interpretation of the equal sign: Symbol for an equivalence relation vs. an operator symbol. *Proceedings of the Fourth International Conference for the Psychology of Mathematics Education* (163–169). Berkeley, CA. . doi:10.1152/jappl.1980.48.1.163.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*, 317–326. doi:10.1007/BF00311062.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

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

Leavy, A., Hourigan, M., & McMahon, Á. (2013). Early understanding of equality. *Teaching Children Mathematics*, *20*, 246–252. doi:10.5951/teacchilmath.20.4.0246.

Lehrer, R., & Schauble, L. (2002). Symbolic communication in mathematics and science: Co-constituting inscription and thought. In E. Amsel & J. Byrnes (Eds.), *Language, literacy, and cognitive development: The development and consequences of symbolic communication* (pp. 167–192). Mahwah, NJ: Erlbaum.

Li, X., Ding, M., Capraro, M. M., & Capraro, R. M. (2008). Sources of differences in children's understandings of mathematical equality: Comparative analysis of teacher guides and student texts in China and the United States. *Cognition and Instruction*, *26*, 195–217. doi:10.1080/07370000801980845.

Linchevski, L., & Herscovics, N. (1996). Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. *Educational Studies in Mathematics*, *30*, 39–65. doi:10.1007/BF00163752.

Lindvall, C. M., & Ibarra, C. G. (1980). Incorrect procedures used by primary grade pupils in solving open addition and subtraction sentences. *Journal for Research in Mathematics Education*, *11*, 50–62. doi:10.2307/748732.

Magruder, R., & Mohr-Schroeder, M. (2013). Solving equations is all about balance: Using virtual manipulatives in the middle school classroom. In D. Polly (Ed.), *Common core mathematics standards and implementing digital technologies* (pp. 201–214). Hershey, PA: IGI Global.

Manches, A., & O'Malley, C. (2012). Tangibles for learning: A representational analysis of physical manipulation. *Personal and Ubiquitous Computing*, *16*, 405–419. doi:10.1007/s00779-011-0406-0.
Mann, R. L. (2004). Balancing act: The truth behind the equals sign. *Teaching Children Mathematics*, *11*, 65–69.
Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology*, *104*, 1–21. doi:10.1016/j.jecp.2008.08.004.
Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, *43*, 316–350. doi:10.5951/jresematheduc.43.3.0316.
Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 183–200). New York, NY: Cambridge University Press.
Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*, 107–119. doi:10.1016/S0959-4752(01)00018-4.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52. doi:10.1207/S15326985EP3801_6.
McNeil, N. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology*, *43*, 687–695. doi:10.1037/0012-1649.43.3.687.
McNeil, N., & Alibali, M. (2005a). Knowledge change as a function of mathematics experiences: All contexts are not created equal. *Journal of Cognition and Development*, *6*, 285–306. doi:10.1207/s15327647jcd0602_6.
McNeil, N., & Alibali, M. (2005b). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*, 883–899. doi:10.1111/cdev.2005.76.issue-4.
McNeil, N., Fyfe, E., Petersen, L., Dunwiddie, A., & Brletic-Shipley, H. (2011). Benefits of practicing 4 = 2 + 2; Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development*, *82*, 1620–1633. doi:10.1111/j.1467-8624.2011.01622.x.
McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction*, *22*, 440–448. doi:10.1016/j.learninstruc.2012.05.001.
McNeil, N. M., Rittle-Johnson, B., Hattikudur, S., & Petersen, L. A. (2010). Continuity in representation between children and adults: Arithmetic knowledge hinders undergraduates' algebraic problem solving. *Journal of Cognition and Development*, *11*, 437–457. doi:10.1080/15248372.2010.516421.
Molina, M., & Ambrose, R. (2008). From an operational to a relational conception of the equal sign. Thirds graders' developing algebraic thinking. *Focus on Learning Problems in Mathematics*, *30*, 61–80.
Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching Children Mathematics*, *8*, 372–377.
National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Authors.
National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors.
National Governors Association Center for Best Practices & Council of Chief State School Officers. (2012). *Common core state standards for mathematics*. Washington, DC: Authors.
Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*, 423–451. doi:10.1207/s15327809jls1303_6.
Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, *6*, 449–468. doi:10.1016/0885-2014(91)90049-J.
Perry, M., Berch, D., & Singleton, J. L. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, *6*, 213–235.
Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*, 359–400. doi:10.1016/0885-2014(88)90021-4.
Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. H. Weaver, Trans. New York, NY: Basic (Original work published 1966).
Powell, S. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal*, *112*, 627–648. doi:10.1086/665009.
Presmeg, N. C. (1997). Reasoning with metaphors and metonymies in mathematics learning. In L. D. English (Ed.), *Mathematical reasoning: Analogies, metaphors, and images* (pp. 267–279). Mahwah, NJ: Erlbaum.
Rasmussen, S. (1977). *Algebra tiles*. Emeryville, CA: Key Curriculum Press.
Renwick, E. M. (1932). Children's misconceptions concerning the symbols for mathematical equality. *British Journal of Educational Psychology*, *2*, 173–183. doi:10.1111/bjep.1932.2.issue-2.
Rittle-Johnson, B., & Alibali, M. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 175–189. doi:10.1037/0022-0663.91.1.175.
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*, 85–104. doi:10.1037/a0021334.

Rojano, T., & Martinez, M. (2009). From concrete modeling to algebraic syntax: Learning to solve linear equations with a virtual balance. In S. L. Swars, D. W. Stinson, & S. Lemons-Smith (Eds.), *Proceedings of the 31st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (vol. 5, pp. 235–243). Atlanta, GA: Georgia State University.

Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development Perspectives*, *3*, 145–150. doi:10.1111/j.1750-8606.2009.00095.x.

Sarama, J., & Clements, D. H. (2016). Physical and virtual manipulatives: What is "concrete"? In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 71–93). Cham, Switzerland: Springer.

Seo, K. H., & Ginsburg, H. P. (2003). You've got to carefully read the math sentence … : Classroom context and children's interpretations of the equals sign. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Recent research and theory* (pp. 161–187). Mahwah, NJ: Lawrence Erlbaum Associates.

Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and non-symbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology*, *101*, 88–100. doi:10.1037/a0013156.

Skemp, R. R. (1986). *The psychology of learning mathematics* (2nd ed.). London, England: Penguin.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal of Research in Mathematics Education*, *20*, 498–505. doi:10.2307/749423.

Stacey, K., & MacGregor, M. (1997). Multiple referents and shifting meanings of unknowns in students' use of algebra. In E. Pehkonen (Ed.), *Proceedings of the 21st Conference of the International Group for the Psychology of Mathematics Education* (vol. 4, pp. 190–197), Lahti, Finland. doi:10.1038/nsb0397-190.

Stephens, A. C., Ellis, A. B., Blanton, M., & Brizuela, B. M. (2017). Algebraic thinking in the elementary and middle grades. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 386-420). Reston, VA: National Council of Teachers of Mathematics.

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*, 173–182. doi:10.1016/j.jmathb.2013.02.001.

Suh, J., & Moyer-Packenham, P. S. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching*, *26*, 155–173.

Thompson, P. W. (1992). Notations, conventions, and constraints: Contributions to effective uses of concrete materials in elementary mathematics. *Journal for Research in Mathematics Education*, *23*, 123–147. doi:10.2307/749497.

Thompson, P. W., & Thompson, A. G. (1990). Salient aspects of experience with concrete manipulatives. In F. Hitt (Ed.), *Proceedings of the 14th annual meeting of the International Group for the Psychology of Mathematics* (Vol. 3, pp. 337–343). Mexico City: International Group for the Psychology of Mathematics Education.

Uttal, D., O'Doherty, K., Newland, R., Hand, L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives*, *3*, 156–159. doi:10.1111/j.1750-8606.2009.00097.x.

Uttal, D. H., Scudder, K. V., & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, *18*, 37–54. doi:10.1016/S0193-3973(97)90013-7.

Van de Walle, J. (2013). *Elementary and middle school mathematics* (8th ed.). Boston, MA: Pearson Education.

Vlassis, J. (2002). The balance model: Hindrance or support for the solving of linear equations with one unknown. *Educational Studies in Mathematics*, *49*, 341–359. doi:10.1023/A:1020229023965.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard.

Warren, E., & Cooper, T. J. (2005). Young children's ability to use the balance strategy to solve for unknowns. *Mathematics Education Research Journal*, *17*, 58–72. doi:10.1007/BF03217409.

Appendix

Screening Questions

1. For each number sentence below, decide if the number sentence is True. In other words, does it make sense[6]? After each problem, circle True, False, or Don't Know.

Examples:

3 + 4 = 7	True	False	Don't Know
3 + 4 = 12	True	False	Don't Know

a) 8 = 8	True	False	Don't Know
b) 7 + 6 = 0	True	False	Don't Know
c) 5 + 3 = 3 + 5	True	False	Don't Know
d) 8 = 5 + 10	True	False	Don't Know
e) 3 + 1 = 1 + 1 + 2	True	False	Don't Know
f) 4 = 4 + 0	True	False	Don't Know

2. Here are some addition problems. Solve them as best as you can. Write your answer in the blank.
   a) 4 + 5 = ___________
   b) 13 + 15 = ___________
   c) 24 + 19 = ___________
   d) 3 + 3 + 3 = ___________

3. Here are some subtraction problems. Solve them as best as you can. Write your answer in the blank.
   a) 6 − 4 = ___________
   b) 18 − 13 = ___________
   c) 26 − 18 = ___________
   d) 9 − 3 − 3 = ___________

Identifying the Two Sides of the Equation

Pretest

1. This problem has two sides. Circle the choice that correctly breaks the problem into **two sides**[7]:

5 = 3 + 2

a)	Side A	Side B	b)	Side A	Side B
	5 =	3		5	3 + 2
c)	Side A	Side B	d)	Side A	Side B
	5 =	3 + 2		5 = 3 + 2	3 + 2 = 5
e) ?					

2. This problem has two sides. Circle the choice that correctly breaks the problem into **two sides**:

4 + 1 = 2 + 3

a)	Side A	Side B	b)	Side A	Side B
	4 + 1 =	2		4 + 1= 2	+ 3
c)	Side A	Side B	d)	Side A	Side B
	4 + 1 = 2	4 + 1 = 2 + 3		4 + 1	2 + 3
e) ?					

Ordered Logistic Regression Analysis on Levels of Learning at the Posttest, with the Control condition as Reference

Predictor Variable	β	se	t	p	Odds Ratio	95% CI 2.5	95% CI 97.5
Dynamic Balance Scale	0.991	0.476	2.083	0.037*	2.695	1.074	7.005
Balance Analogy	2.119	0.488	4.339	0.000*	8.324	3.274	22.381
Core Principle	1.403	0.480	2.924	0.003*	4.067	1.615	10.678
Eligible For Free Lunch	-1.182	0.384	-3.081	0.002*	0.307	0.142	0.643
Age	0.127	0.239	0.532	0.595	1.136	0.710	1.819
AIC ** =	305.07						

** Akaike value for the model; * p < .05.
The Control group and higher SES (non-eligibility for reduced lunch) group are references in the model.

Ordered Logistic Regression Analysis on Levels of Learning, with the Most Grounded (Dynamic Balance Scale Condition) as the Reference

Predictor Variable	β	se	t	p	Odds Ratio	95% CI 2.5	95% CI 97.5
Balance Analogy and Core Principle, combined	0.792	0.380	2.085	0.037*	2.208	1.054	4.696
Control	-0.965	0.474	-2.035	0.042*	0.381	0.147	0.953
Eligible For Free Lunch	-1.129	0.378	-2.984	0.003*	0.323	0.151	0.672
Age	0.200	0.234	0.855	0.392	1.222	0.772	1.941
AIC** =	305.63						

** Akaike information criterion (AIC) value for the model; * p < .05.
The groups with the most support and higher SES (non-eligibility for reduced lunch) were references in this model.

Binomial Logistic Regression Analysis on Students' Relational Definition of the Equal Sign

Predictor Variable	β	se	z	p	Odds Ratio	95% CI 2.5	95% CI 97.5
Condition							
Dynamic Balance Scale	1.869	0.821	2.278	0.023*	6.484	1.541	44.911
Balance Analogy	2.787	0.813	3.428	0.001*	16.237	3.973	111.642
Core Principle	2.030	0.816	2.487	0.013*	7.615	1.831	52.427
Student Background							
Eligible For Free Lunch	-1.024	0.448	-2.287	0.022*	0.359	0.147	0.859
Age	0.228	0.284	0.804	0.421	1.256	0.720	2.204

* p<.05.