Journal of Educational Psychology

Manuscript version of

Improving Children's Understanding of Mathematical Equivalence via an Intervention That Goes Beyond Nontraditional Arithmetic Practice

Nicole M. McNeil, Caroline Byrd Hornburg, Heather Brletic-Shipley, Julia M. Matthews

Funded by:

- National Science Foundation
- U.S. Department of Education, Institute of Education Sciences

Improving Children's Understanding of Mathematical Equivalence Via an Intervention That Goes Beyond Nontraditional Arithmetic Practice

Nicole M. McNeil, Caroline Byrd Hornburg, Heather Brletic-Shipley, and Julia M. Matthews

University of Notre Dame

Accepted by *Journal of Educational Psychology*, November 2018

Author Note

Caroline B. Hornburg is now at Purdue University. Julia M. Matthews is now at the University of Wisconsin-Madison. This research was supported by Grant R305A110198 to McNeil from the U.S. Department of Education's Institute of Education Sciences. Hornburg was supported by a National Science Foundation (NSF) Graduate Research Fellowship under Grant DGE-1313583 while at the University of Notre Dame. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education, or the NSF. Portions of this work were presented at the 2015 Annual Meeting of the American Educational Research Association. Thanks to Alex Boehm, Cristina Carrazza, Erin Celeste, Brianna Devlin, Emily Geiger-Medina, Juliana Johnson, Molly Knapp, Marisa Rieber, and Valerie Williams for their help with coding and administrative tasks. This research would not have been possible without the support of many administrators, teachers, parents, and children across the U.S.

Correspondence concerning this manuscript should be addressed to Nicole M. McNeil (Email: nmcneil@nd.edu), Department of Psychology, University of Notre Dame, 390 Corbett Family Hall, Notre Dame, IN 46556.

Abstract

Elementary school children (ages 7-11) struggle to understand mathematical equivalence, a foundational pre-algebraic concept. Some manipulations to the learning environment, including well-structured nontraditional arithmetic practice alone, have been shown to improve children's understanding; however, improvements have been modest. The goal of this study was to test an iteratively-developed supplemental intervention for second grade that was designed to yield widespread mastery of mathematical equivalence. The intervention included three components beyond nontraditional arithmetic practice: (a) lessons that introduce the equal sign outside of arithmetic contexts, (b) "concreteness fading" exercises, and (c) activities that require children to compare and explain different problem formats and problem-solving strategies. After the development process, a small, randomized experiment was conducted with 142 students across eight second grade classrooms to evaluate the effects of the intervention on children's solving of mathematical equivalence problems, encoding of mathematical equivalence problems, and defining of the equal sign. Classrooms were randomly assigned to the intervention or to nontraditional arithmetic practice alone. Analyses at the classroom level demonstrated that the intervention classrooms performed better than the active control classrooms both in terms of pre-to-post change in understanding of mathematical equivalence (g = 1.87) and accuracy on transfer problems at posttest (g = 2.07). Non-parametric analyses led to the same conclusions. Results suggest that the comprehensive intervention improves children's understanding of mathematical equivalence to levels that surpass equivalent levels of well-structured arithmetic practice alone, as well as business-as-usual benchmarks from previous studies.

Key words: Mathematics intervention, Elementary mathematics, Early algebra

Educational Impact and Implications Statement

This study provides evidence that a newly-developed mathematics intervention for second graders improves understanding of a foundational pre-algebraic concept—mathematical equivalence. The intervention combines well-structured nontraditional arithmetic practice with other, more conceptually-focused activities designed to change the way children think about the equal sign in arithmetic problems. Tests of children's understanding of mathematical equivalence show that the intervention is more effective than a comparable amount of well-structured nontraditional arithmetic practice alone. Survey data suggest that the intervention is easy for teachers to implement and enjoyable for students.

Improving Children's Understanding of Mathematical Equivalence Via an Intervention That Goes Beyond Nontraditional Arithmetic Practice

Should elementary school mathematics classrooms increase their focus on pre-algebraic concepts, such as mathematical equivalence, or should they stick to increasing knowledge of and proficiency with arithmetic operations, measurement, and geometry? Mathematics education researchers recommend nurturing elementary children's algebraic thinking (Blanton & Kaput, 2003; Carpenter, Franke, & Levi, 2003; National Mathematics Advisory Panel, 2008; Schliemann, Carraher, & Brizuela, 2007; Stephens, Blanton, Knuth, Isler, & Gardiner, 2015), but to evaluate this recommendation we first need to develop interventions that effectively teach pre-algebraic concepts to children. Such interventions could be used and outcomes measured as children progress to middle school and beyond. One intervention that improves children's understanding of mathematical equivalence is well-structured, nontraditional arithmetic practice (McNeil, Fyfe, & Dunwiddie, 2015). However, this intervention has not produced widespread mastery in children's understanding of mathematical equivalence (nor has any other). In the present study, we expanded a nontraditional arithmetic practice intervention to include more conceptually-focused components designed to explicitly change the way children think about the equal sign. The goal was to test if children benefit from spending time on these conceptually-focused tasks, or whether the increased time would be better spent on nontraditional arithmetic practice alone.

Mathematical equivalence is the relation between two quantities that are equal and interchangeable (Kieran, 1981), and its symbolic form (i.e., the equal sign) specifies that the two sides of an equation are equal and interchangeable. It is a "big idea" in mathematics that is widely seen as important for facilitating children's mathematical achievement and algebra readiness (e.g., Blanton & Kaput, 2005; Charles, 2005; Falkner, Levi, & Carpenter, 1999; Kieran, 1992; Knuth, Stephens, McNeil, & Alibali, 2006; Matthews & Fuchs, 2018; McNeil, Hornburg, Devlin, Carrazza, & McKeever, 2017). Unfortunately, children between the ages of 7-11 in the U.S. struggle to understand this fundamental concept (Baroody & Ginsburg, 1983; Carpenter et al., 2003; McNeil & Alibali, 2005a; National Mathematics Advisory Panel, 2008), and misconceptions continue into middle school and beyond (Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007; Fyfe, Matthews, Amsel, McEldoon, & McNeil, 2018; Knuth et al., 2006; McNeil & Alibali, 2005b; McNeil, Rittle-Johnson, Hattikudur, & Petersen, 2010). Elementary school children's misconceptions are apparent when they are asked to solve "mathematical equivalence problems," which are equations with operations on both sides of the equal sign (e.g., 3 + 4 = 5 + _). Instead of solving such problems correctly (e.g., 2), they typically add all the numbers in the problem (e.g., add 3 + 4 + 5 and write 12 in the answer blank), or add the numbers before the equal sign (e.g., add 3 + 4 and write 7 in the answer blank).

McNeil and Alibali (2005b) advanced a "change-resistance" account of children's difficulties with mathematical equivalence, arguing that children's difficulties are due to children's overly narrow experience with traditional arithmetic in school (see also McNeil, 2014). In the U.S., children learn arithmetic in a procedural fashion for years before they learn to reason about equations relationally, as expressions of mathematical equivalence. Traditionally, arithmetic problems are presented in either a left-to-right problem format with the operations to the left of the equal sign and the "answer" to the right (e.g., $3 + 4 = \underline{7}$; McNeil et al., 2006; Powell, 2012; Seo & Ginsburg, 2003), or a top-to-bottom format with the operations on top of an equal bar and the "answer" underneath (e.g., $\begin{array}{r} 3 \\ +4 \\ \hline \end{array}$). Neither of these traditional formats highlights the interchangeable nature of the two sides of an equation, and equivalence is symbolized using only the equal sign or the equal bar. Moreover, when arithmetic problems are organized into practice sets, they are often initially grouped iteratively according to a traditional arithmetic table (i.e., all the ones [1 + 1, 1 + 2… 1 + n], followed by all the twos [2 + 1, 2 + 2… 2 + n], and so on).

These narrow experiences with arithmetic lead children to construct incorrect understandings of the equal sign and mathematical equivalence that do not generalize beyond traditional arithmetic (McNeil, 2014). First, children learn a perceptual pattern related to the format of mathematics problems, namely the “operations on left side” format (McNeil & Alibali, 2004). Second, children learn to interpret the equal sign operationally as a “do something” symbol (Baroody & Ginsburg, 1983; Behr, Erlwanger, & Nichols, 1980; McNeil & Alibali, 2005a). Third, children learn the strategy “perform all given operations on all given numbers” (McNeil & Alibali, 2005b). These three narrow patterns have been referred to as *operational patterns* in previous research (e.g., McNeil & Alibali, 2005b) because they are based on experience with arithmetic operations and reflect operational rather than relational thinking (cf. Jacobs, Franke, Carpenter, Levi, & Battey, 2007). Children’s representations of such patterns gain strength during the first few years of formal schooling and become most entrenched around age nine (McNeil, 2007), at which time they become the default representations that children activate when encoding, interpreting, and solving mathematics problems (McNeil, 2014). Although children’s reliance on these patterns may be helpful when solving traditional arithmetic problems (e.g., 3 + 4 = __), it is a hindrance when children have to encode, interpret, or solve problems, like mathematical equivalence problems, that do not adhere to the patterns.

A key prediction of the change-resistance account is that modifying arithmetic practice to be less narrow and more in line with the underlying concepts will reduce children’s reliance on the operational patterns, which will in turn help children construct a better understanding of mathematical equivalence. This prediction is consistent with the recommendations of mathematics educators, who have long called for more diverse, richer exposure to a variety of problem types from the beginning of formal schooling (e.g., Blanton & Kaput, 2005; Hiebert et al., 1996; National Council of Teachers of Mathematics [NCTM], 2000). Several of these experts have suggested that children may benefit from seeing the equal sign in a variety of nontraditional problem formats (Baroody & Ginsburg, 1983; Carpenter et al., 2003; Denmark, Barco, & Voran, 1976; MacGregor & Stacey, 1999; Seo & Ginsburg, 2003; Weaver, 1973).

Well-controlled experiments have confirmed that modifications to traditional arithmetic practice improve children's understanding of mathematical equivalence (Chesney, McNeil, Petersen, & Dunwiddie, 2018; McNeil et al., 2012; McNeil et al., 2011). For example, in one lab-based experiment, children developed a better understanding of mathematical equivalence after practicing arithmetic problems presented in a nontraditional format with operations on the right side of the equal sign (e.g., __ = 9 + 8) than after practicing problems presented in the traditional format (e.g., 9 + 8 = __) or after receiving no extra practice (McNeil et al., 2011). The intervention was deemed effective according to What Works Clearinghouse Standards (d = 0.76 for solving mathematical equivalence problems, d = 0.45 for encoding mathematical equivalence problems, d = 1.36 for defining the equal sign, when compared to the traditional format; What Works Clearinghouse, 2014). In other lab-based experiments, using relational phrases such as "is the same amount as" and "is equal to" in place of the equal sign was found to improve children's encoding of mathematical equivalence problems more than using only the traditional equal sign/equal bar symbols (d = 0.40; Chesney et al., 2018), and organizing practice problems into practice sets by equivalent values (e.g., 5 + 2 = __, 3 + 4 = __, 1 + 6 = __) was found to be better than grouping problems iteratively according to a traditional arithmetic table, resulting in medium-to-large effects on solving, encoding, and defining the equal sign (McNeil et al., 2012).

McNeil, Fyfe, and Dunwiddie (2015) incorporated all three of these successful lab-based modifications into a “nontraditional” arithmetic practice workbook and then tested it experimentally against a traditional arithmetic practice workbook in elementary classrooms. The nontraditional workbook led to immediate and lasting improvements in children’s understanding of mathematical equivalence (when compared to the traditional workbook) after only being used as part of regular classroom instruction for 15 minutes per day, two days per week, for 12 weeks ($d = 0.50$ for solving mathematical equivalence problems, $d = 0.60$ for encoding mathematical equivalence problems, $d = 0.96$ for defining the equal sign). Moreover, a mediation analysis showed that the nontraditional workbook worked as designed based on the change-resistance account’s (McNeil & Alibali, 2005b) theory of change—by decreasing children’s reliance on the operational patterns.

Although well-structured nontraditional arithmetic practice alone has been shown to improve children’s understanding of mathematical equivalence, it is not a panacea. For example, of the children who participated in McNeil et al.’s (2015) nontraditional arithmetic practice intervention, only 40% demonstrated basic understanding (as defined in Table 1), and a mere 4% exhibited mastery (as defined in Table 1). It is important to note here that studies of children’s understanding of mathematical equivalence often report outcomes in terms of the percentage of children who meet certain benchmarks (e.g., solve at least one problem correctly, solve at least 75% of problems correctly; Hornburg, Rieber, & McNeil, 2017; McNeil et al., 2015). This is because performance on measures of understanding of mathematical equivalence tend to be bimodal, with a plurality of children solving either zero or most problems correctly (McNeil &

Alibali, 2004). One goal of the present project was to expand beyond well-structured nontraditional arithmetic practice alone to develop a more comprehensive intervention to decrease children's reliance on the operational patterns and to test if it would help more children achieve mastery than the same amount of time spent on well-structured nontraditional arithmetic practice alone. However, we report accuracy data, as is more standard in mathematics program evaluation.

The specific goals we set for post-intervention performance of children participating in the comprehensive intervention are shown in Table 2, alongside benchmarks for "business as usual" and "best case" conditions. We set high performance goals even though research has shown that most children (ages 7-11) in the U.S. do not demonstrate even a basic understanding of mathematical equivalence under business-as-usual conditions, and few children demonstrate mastery-level performance in the best-case condition after receiving an intervention. Decades of research show that children struggle to gain a formal understanding of mathematical equivalence, even after intervention (e.g., Baroody & Ginsburg, 1983; Byrd, McNeil, Chesney, & Matthews, 2015; Chesney et al., 2018; Cook et al., 2013; Jacobs et al., 2007; Seo & Ginsburg, 2003; Sherman & Bisanz, 2009; consistent with the best-case benchmarks presented in Table 2). However, it is difficult to compare the outcomes of previous studies because they have not targeted the same population of children, used the same measures, or reported children's performance in the same way (e.g., some only reported accuracy rather than the percentage of children achieving basic or mastery understanding). Some studies have included children who already solve mathematical equivalence problems correctly prior to receiving the intervention (e.g., DeCaro & Rittle-Johnson, 2012; Fyfe, Rittle-Johnson, & DeCaro, 2012; Rittle-Johnson, 2006), and some studies have taught one narrow skill or concept in a single session and then post-tested children on their ability to duplicate that taught skill or concept immediately after the lesson (e.g., Alibali, Phillips, & Fischer, 2009; Hattikudur & Alibali, 2010; McNeil & Alibali, 2000). Such studies generally report higher performance levels, but it is difficult to interpret these effects because the type of learning that results from a single individual learning session can be shallow and temporary (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Perry, 1991). One final caveat is that the McNeil et al. (2015) intervention that we are using as our best-case benchmark (Table 2) included four problems in a nontraditional format ($a = b + c$) prior to the posttest mathematical equivalence problem assessment, so outcomes may be inflated over what would be expected when children are presented solely with problems that contain operations on both sides of the equal sign. Simple nontraditional arithmetic problems may help children access a relational approach to problem solving rather than rely on their default modes of solving in accordance with operational patterns (McNeil et al., 2011). For these reasons, it would be difficult for us to make firm conclusions about the current intervention just by comparing to the business-as-usual and best-case benchmarks. Thus, the present study included an active control condition in which children received well-structured nontraditional arithmetic practice alone (akin to the classroom RCT conducted by McNeil et al. [2015]), and classrooms were randomly assigned to conditions.

Theoretical Basis for the Comprehensive Intervention

The additional components of the comprehensive intervention were based on McNeil and Alibali's (2005b) change-resistance account of children's difficulties with mathematical equivalence (see also McNeil, 2014; McNeil, Hornburg, Fuhs, & O'Rear, 2017). As discussed above, this account suggests that children will develop a mastery understanding of mathematical equivalence if they are prevented from extracting, representing, activating, and applying the overly narrow operational patterns routinely encountered in arithmetic (i.e., the “operations on left side” problem format, the arithmetic-specific interpretation of the equal sign, and the “perform all given operations on all given numbers” strategy). Although well-structured nontraditional arithmetic practice reduces children’s reliance on the operational patterns and improves understanding of mathematical equivalence, it does not produce mastery when being used for fifteen minutes a day, twice a week, for twelve weeks.

Several sources of direct and indirect evidence in the literature (described next) indicated that three more conceptually-focused components beyond procedural nontraditional arithmetic practice may be useful in decreasing children’s reliance on the operational patterns: (a) lessons that introduce the equal sign outside of arithmetic contexts, (b) “concreteness fading” exercises, and (c) activities that require children to compare and explain different problem formats and problem-solving strategies. In contrast to nontraditional arithmetic practice, which has already been tested in RCTs, the three new components were based on ideas and evidence in the research literature, but they were not yet established in terms of being classroom-ready interventions. Thus, we designed and constructed many of our own intervention materials for these components.

Lessons that introduce the equal sign outside of arithmetic contexts. Although exposure to nontraditional arithmetic practice is helpful for preventing the entrenchment of operational patterns, it may not be enough on its own to completely overcome the operational view (Denmark et al., 1976). Children informally interpret addition as a unidirectional process even before the start of formal schooling (Baroody & Ginsburg, 1983), and they start to apply the operational patterns to arithmetic problems at least as early as first grade (e.g., Falkner et al., 1999). Thus, any time children encounter an arithmetic problem, they may activate representations of the operational patterns to some degree, regardless of whether those arithmetic problems are presented in traditional format (e.g., 3 + 4 = __) or nontraditional formats (e.g., __ = 3 + 4). To combat this tendency, it may be necessary to first expose children to the equal sign outside of arithmetic contexts (e.g., ••• = •••, 10 = 10), so they can solidify a relational concept of the equal sign before moving on to a variety of arithmetic problem formats (Baroody & Ginsburg, 1983; Denmark et al., 1976; McNeil, 2008; Renwick, 1932).

This hypothesis was supported experimentally in a study showing that children learn more from lessons on the meaning of the equal sign when those lessons are given outside of arithmetic contexts (e.g., 28 = 28, 1 foot = 12 inches) than when they are given in the context of arithmetic problems (McNeil, 2008). McNeil's effect in Experiment 1 (with second and third graders who had low performance overall) was small-to-medium (η_p^2 = .06) and the effect in Experiment 2 (with fifth graders) was large (η_p^2 = .30). The hypothesis about introducing children first to the equal sign outside the context of arithmetic also corresponds to the way the equal sign is introduced in China, where over 90% of elementary school children solve mathematical equivalence problems correctly (Capraro et al., 2010; Li, Ding, Capraro, & Capraro, 2008). In contrast to mathematics textbooks in the United States, mathematics textbooks in China often introduce the equal sign in a context of equivalence relations first and only later embed the equal sign within mathematical equations involving arithmetic operators and numbers. Li et al. argue that this difference in the format and sequence of problems that children are taught in school contributes to the difference in understanding between children in China versus the U.S. Their analysis together with the experimental finding above suggests that teaching the equal sign in the context of equivalence relations first before embedding it within arithmetic problems could be one way to reduce children's reliance on the operational patterns, so we included this component in our comprehensive intervention.

"Concreteness fading" exercises. Although elementary school children in the U.S. define the equal sign operationally and solve mathematical equivalence problems incorrectly, they seem to have no trouble defining the word "equal" correctly or identifying addend pairs that are "equal to" one another outside of the equation context (e.g., they can choose 5 + 3 from among distracters when asked to choose an addend pair that is equal to 3 + 5, Rittle-Johnson & Alibali, 1999). Moreover, children perform better on mathematical equivalence problems when the problems are presented with concrete materials (e.g., wooden blocks) than when they are written in symbolic form (Sherman & Bisanz, 2009). Thus, children's misunderstanding of the abstract symbols used to represent quantities (e.g., 1, 2, etc.), operations (e.g., +, –, etc.), and relations (e.g., =, >, <) in mathematics is one factor contributing to children's poor understanding of mathematical equivalence.

To circumvent children's difficulties with symbolic notation, educators often introduce mathematical concepts with concrete materials. For example, instead of writing "1 + __ = 3" on the chalkboard, a teacher may use a balance scale with one object on the left side and three objects on the right. The notion that concrete materials benefit learning has a long history in psychology and education, dating back to Montessori (1917), Bruner (1966), and Piaget (1970). Concrete materials are theorized to benefit learning by activating real-world knowledge (Kotovsky, Hayes, & Simon, 1985; Schliemann & Carraher, 2002), inducing physical or imagined action (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004; Martin & Schwartz, 2005), and enabling children to construct their own knowledge of abstract concepts (Brown, McNeil, & Glenberg, 2009; Martin, 2009; Smith, 1996). Concrete materials are also appealing for practical reasons. They are widely available in teaching supply stores, and they may increase children's motivation and interest in learning (Burns, 1996).

Although a number of studies have suggested that concrete materials hinder learning of abstract mathematical concepts (e.g., Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2006, 2008; McNeil, Uttal, Jarvin, & Sternberg, 2009; Sloutsky, Kaminski, & Heckler, 2005), none of these studies used concreteness fading, which is the recommended way to incorporate concrete materials into mathematics instruction (Bruner, 1966). Children should begin with concrete instantiations that are slowly faded away to more abstract representations (Bruner, 1966; Fyfe, McNeil, & Borjas, 2015; Fyfe, McNeil, Son, & Goldstone, 2014; Goldstone & Son, 2005; Miller & Hudson, 2007).

For example, an educator could first represent the concept "two" with two red apples, then with a picture of two dots representing those apples, and finally with the Arabic numeral 2. The objective of this "concreteness fading" method is to first present mathematical concepts in concrete, recognizable forms. Then, these forms are "varyingly refined" to strip away irrelevant details until they are finally presented in the most economic, abstract symbolic form. This sequence is ideal, according to Bruner (1966), because it increases learners' ability to understand, apply, and transfer what they learn. Bruner argued that it would be risky to skip the concrete form because learners who learn only through the abstract, symbolic form do not have a store of images in long-term memory to "fall back on" when they either forget or are unable to directly apply the abstract, symbolic transformations they have learned. Since Bruner's time, many researchers and educators have supported the idea of starting with concrete instantiations then explicitly "decontextualizing" or "fading" away to the more abstract (Butler, Miller, Crehan, Babbitt, & Pierce, 2003; Fyfe et al., 2015; Fyfe et al., 2014; Goldstone & Son, 2005; Gravemeijer, 2002; Lehrer & Schauble, 2002). Experimental evidence has shown that concreteness fading fosters a greater understanding of mathematical equivalence than purely concrete or abstract methods alone (Fyfe et al., 2015). Overall, the evidence suggests that concreteness fading exercises build conceptual knowledge in a way that helps to reduce children's reliance on the operational patterns, so we included this component in our comprehensive intervention.

Activities with comparison and explanation of different problem formats and strategies. Learning arithmetic in traditional classroom settings often involves individual children sitting quietly at their desks solving sets of homogeneous arithmetic problems using a single strategy that has been taught by the teacher. This type of learning can be classified as *active* because children are doing something (i.e., solving problems). However, to produce deep and lasting learning, instructors should also incorporate constructive and interactive activities into their lessons (Chi, 2009). *Constructive* activities include processes such as self-explanation (Chi, de Leeuw, Chiu, & LaVancher, 1994) and comparing and contrasting cases (Rittle-Johnson & Star, 2007; Schwartz & Bransford, 1998). They require learners to produce some outputs that go beyond the given information, and this process helps learners infer new knowledge and improve their existing knowledge. *Interactive* activities include processes such as guided instructional dialogues (Graesser, Person, & Magliano, 1995) and peer dialogues (Rittle-Johnson & Star, 2007; Roscoe & Chi, 2007). They require learners to explain and defend their ideas, while responding to a partner's corrective feedback or building on a partner's contributions to the discussion. In line with recommendations from mathematics educators (e.g., Ball, 1993, Boaler, 2000, 2002; Lambert, 1990) these activities go beyond traditional methods of teaching mathematical procedures and focus instead on teaching students to describe and reflect upon the underlying processes of mathematics through small-group and whole-class discussions. In three-year case studies of two schools, Boaler (1998) found that the classroom community created by such discussion about mathematical concepts is helpful for promoting deep and flexible learning, in comparison with the traditional approach (see also Boaler, 2000).

The evidence suggests that the constructive and interactive processes of comparing and explaining facilitate third through fifth grade children's understanding of mathematical equivalence. For example, both Siegler (2002) and Rittle-Johnson (2006) showed that children's accuracy solving mathematical equivalence problems improves after they are asked to explain why the correct solution to a mathematical equivalence problem is correct and why an incorrect solution is incorrect. This type of dual explaining produces significantly better outcomes than only asking children to explain why a correct solution is correct, presumably because dual explaining involves an implicit form of comparison (Siegler, 2002). Specifically, Rittle-Johnson's (2006) study demonstrated that children who explained why solutions were either correct or incorrect generated more correct procedures than children who did not explain, $d = 0.33$. Hattikudur and Alibali (2010) directly examined the effects of comparison on children's understanding of mathematical equivalence and showed that children construct a significantly better conceptual understanding of mathematical equivalence after receiving instruction that involves comparing the equal sign with other relational symbols (e.g. >, <) than after receiving similar instruction that involves the equal sign alone. Furthermore, Carpenter and colleagues (e.g., Carpenter et al., 2003; see also Falkner et al., 1999; Jacobs et al., 2007) have shown that engaging elementary children in whole classroom activities that require them to compare and explain various problem formats and true-false statements leads children to generate important generalizations about basic properties of arithmetic and improves their understanding of mathematical equivalence. These classroom results together with the lab results suggest that children are less reliant on the operational patterns when they engage in constructive and interactive activities that require them to compare and explain different problem formats and strategies; thus, we included this component in our comprehensive intervention.

Overview of the Present Study

One goal of the present study was to develop a comprehensive intervention that is better than a comparable amount of well-structured nontraditional arithmetic practice at improving children's understanding of mathematical equivalence, while still being reasonably affordable, portable, and easy to administer. A second, broader goal, was to refine our theory of the process by which children construct a deep, mastery-level understanding of mathematical equivalence, together with instructional practices and activities that can support that learning (cf. Cobb, McClain, de Silva Lamberg, & Dean, 2003). We developed and refined a comprehensive intervention in two iterative design experiments (Byrd, McNeil, Brletic-Shipley, & Matthews, 2013). Findings suggested that the comprehensive intervention could be implemented in a classroom setting in a way that leads children to construct an understanding of mathematical equivalence that surpasses the best-case and business-as-usual benchmarks (described above). However, design experiments are not RCTs, and the number of plausible alternative explanations for these findings is large. Moreover, the intervention was administered by highly knowledgeable tutors (in the first design experiment) and by a teacher who had worked with us for several years (in the second design experiment), which calls into question the generalizability of the results.

Here we further evaluated the benefits and feasibility of the comprehensive intervention by conducting a small, randomized experiment in the classrooms of inexperienced teachers that we had never worked with previously. We compared the comprehensive intervention to an active control condition in which children received an equivalent amount of well-structured nontraditional arithmetic practice alone. Well-structured nontraditional arithmetic practice has been shown to improve children's mathematical equivalence understanding over traditional arithmetic practice in both a lab-based RCT (McNeil et al., 2011) and a classroom-based RCT (McNeil et al., 2015). Thus, it might be argued that children's intervention time is best spent working on this empirically-supported activity. Moreover, well-structured arithmetic practice is likely one of the easiest interventions to administer because it does not require explicit instruction. Educators may prefer it over other interventions that require more effort, especially if outcomes are similar. However, some mathematics educators argue that procedural practice alone—no matter how well structured—will be inferior to more conceptually-focused activities, especially those that include explicit, discursive justification processes (e.g., Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Schliemann et al., 2007). For this reason, we hypothesized that the comprehensive intervention would be better than the same amount of nontraditional arithmetic practice alone at promoting understanding of mathematical equivalence.

Method

Participants

Participants were eight second-grade classrooms from eight different under-resourced parochial schools across the U.S. (see Table 3 for characteristics of each classroom and Table 4 for the performance of each classroom at pretest). The teachers in the classrooms were all inexperienced teachers in a two-year "service through teaching" program that places college graduates in under-resourced parochial schools throughout the United States. This program is a member program of the Corporation of National and Community Service, and its teachers are AmeriCorps members. Note that there were 153 children across all classrooms at pretest, but that number fell to 142 children at posttest (7% attrition). The children lost to attrition were similar to the maintained children on all three components of understanding of mathematical equivalence, as well as in terms meeting criteria for basic and mastery understanding of mathematical equivalence (all p-values > .10). Moreover, the attrition was not significantly different between the comprehensive intervention and active control conditions, $\chi^2(1, N = 153) = 1.13, p = 0.25$.

Design

The design was a pretest-intervention-posttest-transfer design with matched-groups random assignment to intervention condition.

Conditions

Four classrooms were assigned to the comprehensive intervention and four classrooms were assigned to the active control condition. The active control condition was well-structured nontraditional arithmetic practice with addition facts similar to McNeil et al. (2015). An active control condition like this, which uses an intervention that has already been shown to be effective, provided a rigorous bar for testing the impact of the comprehensive intervention, certainly more rigorous than the “business as usual” control conditions that are often used in these kinds of experiments. See Table 5 for the intervention elements that were consistent across both conditions. The active control was also matched to the comprehensive intervention in terms of total number of sessions, suggested timeline for completion of all sessions, class time allotted to each session, time during each session where the teacher gave direct feedback, relative progression of sums practiced (i.e., smaller sums like 5, 6, and 7 in the beginning and larger sums towards the end), and even the cover design of the workbooks. Thus, the two conditions were matched in all ways except for the three more conceptually-focused components designed to further reduce reliance on the operational patterns. See Table 6 for the elements that were unique to the comprehensive intervention. As part of incorporating these elements, teachers assigned to the comprehensive intervention were given concrete manipulatives (e.g., balance scales, stickers, equation symbol cards, and student dry erase boards) in addition to the teacher manuals and workbooks that were provided to teachers in both conditions.

Measures of understanding of mathematical equivalence

Children completed McNeil et al.'s (2011) measure of understanding of mathematical equivalence. The measure includes three tasks: (a) equation solving, (b) equation encoding, and (c) defining the equal sign. According to McNeil and Alibali (2005b), these three tasks tap into a system of three distinct, but theoretically related constructs involved in children's understanding of mathematical equivalence. This measure of understanding of mathematical equivalence has exhibited high levels of reliability and validity in previous work (e.g., Knuth et al., 2006; McNeil & Alibali, 2005a, 2005b).

To assess *equation solving*, children completed a paper-and-pencil test consisting of eight mathematical equivalence problems (e.g., 3 + 5 + 6 = 3 + __). Teachers instructed children to try their best to solve each problem and write the number that goes in each blank. After these initial instructions children solved the problems independently. Children's problem-solving strategies were coded as correct or incorrect based on a system used in previous research (e.g., McNeil & Alibali, 2004; Perry, Church, & Goldin-Meadow, 1988). For most problems, the correctness of the strategy used could be inferred from the response itself (e.g., for the problem 3 + 5 + 6 = 3 + __, a response of 17 indicates an incorrect "Add All" strategy and a response of 11 indicates a correct strategy). As in prior work involving paper-and-pencil assessment of strategy use on these types of equations (McNeil, 2007; McNeil et al., 2015), responses were coded as reflecting a particular strategy as long as they were within ±1 of the response that would be achieved with that particular strategy because we were most interested in whether children conceptualized the problem correctly or operationally, so we did not wish to penalize children for simple calculation errors. However, conclusions do not change if we use exactly correct as the equation solving criterion. Cronbach's alpha for the equation solving task across pretest and posttest was .87.

To assess *equation encoding*, children were asked to reconstruct four mathematical equivalence problems after viewing each for five seconds (McNeil & Alibali, 2004; Rittle-Johnson & Alibali, 1999; cf. Chase & Simon, 1973; Siegler, 1976). Teachers projected each problem using a SMARTboard or overhead projector and left it visible for five seconds. Children's reconstructions were coded as correct if they were exactly correct (as in McNeil et al., 2015). Overall, 15 percent of children's reconstructions contained errors only involving misencoding the problem structure like converting the problem to a traditional arithmetic problem in the "operations equals answer" format (e.g., reconstructing 2 + 3 + 6 = 2 + __ as "2 + 3 + 6 + 2 = __"), omitting the plus sign on the right side of the equal sign (e.g., reconstructing 2 + 3 + 6 = 2 + __ as "2 + 3 + 6 = 2 __"), and omitting the equal sign (e.g., reconstructing 3 + 5 + 4 = __ + 4 as "3 + 5 + 4 __ + 4"). These are typically referred to as conceptual errors. Twelve percent contained only number errors, which involve misencoding the order of the numbers in the problem, or excluding a number in the problem (e.g., reconstructing 3 + 5 + 4 = __ + 4 as 3 + 5 = __ + 4). Thirty-five percent had both conceptual and number errors. One percent were other errors (e.g., no response). Note that slightly more children in each classroom meet the basic and mastery standards for understanding of mathematical equivalence (described below) if encoding performance is coded based only on children's conceptual correctness. However, the main results (i.e., the observed differences between classrooms in the comprehensive and active control interventions) hold regardless of this coding choice.

Cronbach's alpha for the equation encoding task across pretest and posttest was somewhat low at .63, which is a potential limitation, but it is possibly because the task draws on two (associated but distinct) processes: knowledge of equation structure (may contribute more to conceptual correctness) and basic memory processes (may contribute more to correctly encoding the specific numbers). Indeed, alpha was somewhat higher (.71) when encoding performance is coded based only on children's conceptual correctness. Callender and Osburn (1979) suggest that Guttman's $\lambda 2$ is a more appropriate measure of reliability for composite tasks, so we also calculated it for the equation encoding task, but it was not much higher at .64. We also examined inter-item correlations among the four encoding problems and found significant overall correlations among all items, ranging from r = .29 to r = .43.

To assess *defining the equal sign*, children responded to a set of questions about the equal sign. An arrow pointed to an equal sign presented alone, and the text read: (1) "What is the name of this math symbol?" (2) "What does this math symbol mean?" and (3) "Can it mean anything else?" (cf. Baroody & Ginsburg, 1983; Behr et al., 1980; Knuth et al., 2006). Teachers read each of these questions aloud, and children wrote down their responses to the questions immediately after each was read. Children's responses were categorized according to a system used in previous research (e.g., Knuth et al., 2006; McNeil & Alibali, 2005a, 2005b). We were specifically interested in whether or not children defined the equal sign relationally as a symbol of mathematical equivalence (e.g., "Two amounts are the same"). Defining the equal sign relationally has had good concurrent validity in previous research (e.g., Alibali et al., 2007;

Knuth et al., 2006). Children's definitions of the equal sign were coded as relational if they gave a relational response to either question #2 or #3.

Measure of computational fluency

Children completed a measure designed to assess their fluency with basic arithmetic facts which consisted of simple addition problems (cf. Geary, Bow-Thomas, Liu, & Siegler, 1996). This paper-and-pencil assessment includes all pair-wise combinations of the numbers 1-9, for a total of 81 problems, presented in a random order. Children were given one minute to solve as many problems as possible.

Transfer test

To assess transfer beyond the specific types of problems taught in the interventions, children completed a paper-and-pencil test (Cronbach's alpha = .79) that assessed understanding of mathematical equivalence using more advanced problems that were not strictly aligned with the intervention. Transfer items included four difficult mathematical equivalence problems with novel formats (6 – 1 = 3 + __, 2 + 5 + 3 = 14 – __, 5 + 22 + 3 = 22 + __, 13 + 18 = __ + 19) and six "far transfer" items (see Table 7).

Procedure

The research was approved by the University's Institutional Review Board. All participating teachers were told that they would administer an intervention designed to improve children's understanding of mathematical equivalence and that we were interested in testing whether one of the interventions would be better than the other. However, we did not discuss the specific differences between the two interventions, and teachers only saw the version of the intervention that they were randomly assigned to use. Moreover, the introduction to the teacher manuals in both conditions had a special note instructing teachers not to provide any additional instruction beyond what is contained in the intervention workbooks or in their regular mathematics curriculum. They were told that this was very important for us to be able to accurately evaluate the effects of the intervention.

Children in each classroom first completed a pretest to assess their understanding of mathematical equivalence. It included modified versions of the measure we had used in our design experiments (described above). Teachers administered the *equation solving*, *equation encoding*, and *defining the equal sign* measures in the classroom setting. After these measures, children also completed the computational fluency measure. After children completed the pretest, teachers mailed them to us. Inter-rater reliability was established for each measure by having a second coder evaluate the responses of a randomly selected 20% subsample. Inter-rater reliability for pretest and posttest was high for each measure; agreement between coders was 99.8% for coding whether or not a given strategy was correct (equation solving), 99.6% for coding whether or not a given reconstruction was correct (equation encoding), 98% for coding whether or not a given definition was relational (defining the equal sign), and 99.8% for coding whether or not a given addition problem solution was correct (computational fluency).

Because of the wide variability in mathematics achievement seen across classrooms in the U.S., we used the pretest data to split the classrooms into high and low performance groups based on their understanding of mathematical equivalence. We then used matched-groups random assignment to ensure that each condition contained two of the highest-performing classrooms and two of the lowest-performing classrooms. We chose this method because of our small sample size and desire to ensure classrooms in each condition were as similar as possible prior to the intervention. The only component of understanding of mathematical equivalence that differed significantly across classrooms at pretest was encoding performance, $F(7, 145) = 4.92$, p

$< .001$. Classrooms 3, 4, 5, and 7 had significantly better encoding performance than did classrooms 1, 2, 6, and 8, $t(6) = 3.61$, $p = .01$. The same group of classrooms also had significantly higher percentages of children who demonstrated basic understanding of mathematical equivalence, *Welch* $t(4.04) = 4.51$, $p = .01$. Our random assignment procedure ensured that two of these four higher-performing classrooms were assigned to each condition. Note, however, that all classrooms generally exhibited poor understanding of mathematical equivalence at pretest, and all had 0% of children demonstrating mastery (see Table 4). Performance on the computational fluency assessment was also similar across conditions (active control classrooms $M = 10.55$; comprehensive intervention classrooms $M = 10.12$). After assignment to condition, we mailed intervention materials to all classrooms. All teachers were told to devote approximately 15-20 minutes of class time twice a week for 16 weeks to implementing the supplemental sessions (32 sessions total). Recognizing that field trips and other events may disrupt scheduled plans, we requested that teachers complete all 32 sessions within no fewer than 12 weeks and no more than 20 weeks, and that sessions each occur on separate days (no two sessions on the same day). After the intervention, children completed the posttest (identical to the pretest) and the transfer test. Teachers also asked children to complete surveys about their experience with the intervention, by rating their level of agreement with four statements about the intervention on a 1-4 scale (with higher values reflecting higher agreement). The transfer test and experience survey were administered on a separate day from the posttest, on which a few children were absent (see Table 8). We requested that teachers make sure all children in their classroom completed the posttest (makeup when absent), but we did not have a strict makeup policy for the transfer and ratings assessments.

Fidelity

The intervention activities for both conditions were administered through the use of workbooks that were designed to be straightforward and ready to use out of the box, and all student work was recorded in these workbooks. Teachers simply read the instructions in the accompanying teacher's manual, and children completed workbooks to follow along, so the completed student workbooks are a reasonable measure of fidelity. Thus, we collected all workbooks at the end of the study as our main check on fidelity. We examined the number of intervention sessions completed. Most children in every classroom across both conditions (98% on average) completed at least 75% (24 of 32) of the intervention sessions, and conclusions were unchanged when we excluded those children who missed 9 or more sessions (n = 3, all in the comprehensive intervention).

We also maintained email contact with teachers throughout the year, and conducted surveys and phone interviews. Teacher surveys were administered at both the midpoint and endpoint of the study through Qualtrics, an online questionnaire platform, with a 100% response rate. Phone interviews were also conducted at both the midpoint and endpoint of the study with a 100% response rate. Results from the phone interviews complemented the results from the surveys. All indications were that the teachers used the intervention materials as designed. On some survey items, teachers used a Likert scale (1 = never, 5 = almost every session) to report on their perceptions of how often they used the intervention materials in particular ways in their classrooms, as well as how often children in their classrooms saw certain types of problems and activities during the intervention sessions. Questions related to components of the comprehensive intervention (but not the active control condition) were asked only at the endpoint so as not to contaminate the practices in the active control condition.

Teachers in both conditions reported using the interventions as instructed at both the midpoint and endpoint surveys. Given the similar results between timepoints, only the endpoint survey results are presented here. Teachers responded "often" (4 on the 1-5 scale) or "almost every session" (5 on the 1-5 scale) to survey items such as "I stuck closely to the teacher manual and did not provide any additional instruction beyond what was contained in the workbooks" (*M* active control = 4.75, *M* comprehensive intervention = 4.5, $t(6) = -0.66$, $p = .54$), and "I used [the intervention] only as a supplement to my regular math curriculum and did not replace or alter my regular weekly math lessons" (*M* active control = 5, *M* comprehensive intervention = 4.75, $t(6) = -1.00$, $p = .36$). On another survey item teachers used a Likert scale (1 = strongly disagree, 5 = strongly agree) and reported disagreeing with the statement "I made accommodations or modifications to [the intervention] to better facilitate student outcomes" (*M* active control = 2.5, *M* comprehensive intervention = 2.75, $t(6) = 0.22$, $p = .83$).

Moreover, as expected, there was no evidence that teachers in the two conditions differed in terms of their perceptions of how often children were exposed to the following nontraditional arithmetic practices: "seeing equations that have operations on the right side of the equal sign" (*M* active control = 4.5, *M* comprehensive intervention = 4.5), "solving for a missing value in an equation" (*M* active control = 4.5, *M* comprehensive intervention = 4.25, $t(6) = -0.45$, $p = .67$), "seeing or hearing the phrase 'is equal to' in place of the equal sign" (*M* active control = 4.75, *M* comprehensive intervention = 5.0, $t(6) = -1.00$, $p = .36$), and "generating a set of equations that all have the same sum" (*M* active control = 3.0, *M* comprehensive intervention = 3.25, $t(6) = 0.24$, $p = .82$).

However, teachers in the two conditions did differ in terms of their perceptions of how often children were exposed to the following two components, which were included in the comprehensive intervention condition but not the active control: “being reminded of concrete equivalence contexts (e.g., sharing, balancing) before solving written equations” (*M* active control = 2.0, *M* comprehensive intervention = 3.5, $t(6) = 3.00$, $p = .02$) and “comparing and contrasting two equations,” (*M* active control = 2.25, *M* comprehensive intervention = 4.0, $t(6) = 2.33$, $p = .058$). These responses provide additional evidence to support fidelity of implementation of the interventions. The one place we did not find the expected condition differences on the survey was for “viewing the equal sign outside of the context of arithmetic,” (*M* active control = 4.5, *M* comprehensive intervention = 3.25, $t(6) = -1.39$, $p = .22$). However, when we asked teachers about this question in the phone interviews, it was discovered that teachers interpreted this question in a variety of different ways, including thinking that it meant “outside of the traditional arithmetic context,” which was a component that was common to both conditions.

As an additional measure of fidelity, we also asked teachers to send back their teacher manuals if they had recorded notes, observations, reflections, and recommendations about use of the intervention in the logbook sections. The teacher logbook pages were provided within each session in the teacher manuals. Five teachers returned the teacher manuals to us with some comments, but two teachers in the comprehensive intervention and one teacher in the active control condition did not. The teachers in the comprehensive intervention who returned their logbooks made comments on 16 sessions on average and those in the active control made comments on 14 sessions on average. Comments primarily included notes about their perceptions of student engagement and understanding (e.g., “most of the students seem to be getting the concept,” “most students enjoyed this session and seemed engaged.”) or session timing (e.g., “this was a quicker session,” or “didn’t finish everything because of the 20-minute guideline."). One observation mentioned the nontraditional problem format: "students seem to do better when the addend boxes are on the left and the sum is to the right." Taken together, evidence from student workbooks, teacher surveys and phone interviews, as well as teacher logbooks point to correct implementation of the intervention in each condition. However, our evidence for fidelity is indirect. Without classroom observation we do not have the full picture of how the teachers delivered the interventions.

Results

We analyzed the data at the classroom level because this is the most conservative approach (n = 4 per condition). For each dependent variable, we tested for evidence of violations in the homogeneity of variance assumption (Levene's test) and normality assumption (Shapiro-Wilk test), and we found no evidence of violations (all p-values > .05), other than the one violation in homogeneity reported in Table 8. Thus, we analyzed the data using unique t-tests, and we report the Welch's t-test, which does not assume equal variances, when homogeneity of variance is violated. However, given that we have low power for detecting violations in normality, we re-analyzed the data for all of the main outcomes of interest using Mann-Whitney U tests, which are non-parametric tests that do not require the normality assumption (see Table 9). Also, because small-n analyses may be associated with inflated effect size estimation and low reproducibility, we also present chi-square tests averaging across classrooms as an additional way to examine if basic understanding and mastery understanding are contingent on condition (below). Conclusions are the same across all ways of analyzing the data.

Table 8 presents the average classroom performance on each outcome by condition along with the estimated effects, table 9 presents the median classroom performance on the primary outcomes along with the Mann-Whitney results, and table 10 shows performance on the primary outcomes by individual classrooms in each condition. As shown in the tables, the benefits of the comprehensive intervention on children's understanding of mathematical equivalence were both clinically and statistically significant when compared to the active control group.

As in prior work, we combined all three of measures of understanding of mathematical equivalence to create a composite measure of mathematical equivalence understanding by summing *z*-scores across the three measures of mathematical equivalence understanding. Classrooms that received the comprehensive intervention improved significantly more than classrooms that received the active control (see Table 8). The effect size was large. Note that we report Hedges' *g* instead of Cohen's *d* because of our conservative tests at the classroom level. When breaking the composite score into its components, the mean difference favored the comprehensive intervention in each case; however, only equation solving was statistically significant on its own (see Table 8), suggesting that the intervention primarily worked through improving children's equation solving. However, the What Works Clearinghouse considers effect sizes greater than .25 as substantively important even when they are not statistically significant (Institute of Education Sciences, 2014), and the effect sizes for all components meet this criterion.

You may recall that we had three goals for the comprehensive intervention: (a) that it would exceed our "best-case" benchmarks, (b) that all of the children would demonstrate basic understanding, and (c) that at least 50% of children would reach mastery. Classrooms that received the comprehensive intervention had a higher percentage of children exhibiting basic understanding and mastery than did classrooms that received the active control, and the effect sizes for both effects were large (see Table 8). We reach the same conclusions when analyzing the data averaged across the classrooms. On average, 85% (66 of 78) of children in the comprehensive intervention demonstrated basic understanding and 21% (16 of 78) achieved mastery. This performance was significantly better than in the active control condition, in which only 30% (19 of 64) of children demonstrated basic understanding, $\chi^2(1, N = 142) = 44.14$, $p < .001$ and only 2% (1 of 64) demonstrated mastery, $\chi^2(1, N = 142) = 11.98$, $p < .001$. Thus, the comprehensive intervention met its goal for exceeding the best-case benchmarks and for exceeding the active control condition in the present experiment, but it fell short of the stated objective goals for basic understanding (100%) and mastery understanding (≥50%). Note, however, upon inspection of the data, we noticed that we were closer to achieving our goal for mastery when considering only equation solving and encoding. On average, 45% (35 of 78) of children in the comprehensive intervention solved at least 75% of the equations correctly and encoded at least 75% of the equations correctly. Compare that to only 9% (6 of 64) in the active control condition, $\chi^2(1, N = 142) = 21.57$, $p < .001$.

Importantly, the benefits of the comprehensive intervention also held for transfer problems. Classrooms that received the comprehensive intervention performed significantly better on transfer problems than did classrooms that received the active control (see Table 8). Again, the effect size was large.

Finally, we did not expect that the intervention conditions would produce differential gains in computational fluency, and we did not find statistically significant differences in computational fluency when comparing the four comprehensive intervention classrooms to the four active control classrooms (see Table 8). However, the effect size was less than -.25, which meets the What Works criterion for substantively important even when it is not statistically significant (Institute of Education Sciences, 2014), so it is possible that there may be some trade-offs in terms of computational fluency.

Survey data suggested that children in the comprehensive intervention rated their enjoyment of the intervention significantly higher on a scale from 1-4 (M = 3.05, SD = 0.08) than did children in the active control (M = 2.50, SD = 0.19), $t(6) = 5.20$, $p = .002$, g = 3.20 [1.03-5.30]. However, children in both conditions reported that the intervention was important (p = .76), not too difficult (p = .85), and different from their usual math activities (p = .56). Surveys and interviews of participating teachers similarly suggested that teachers in both conditions viewed their respective interventions as important, enjoyable, fairly easy to implement, and beneficial to the children in their classrooms.

Discussion

Our goal was to develop a supplemental, comprehensive intervention for helping children construct a mastery-level understanding of mathematical equivalence. We used the change-resistance account of children's difficulties with mathematical equivalence (McNeil, 2014; McNeil & Alibali, 2005b) as our guiding framework, and we drew on the activities and instructional practices that have been shown to decrease activation of and reliance on the operational patterns in previous research. We already had strong direct evidence from multiple studies, including a classroom-based study, that nontraditional arithmetic practice is beneficial for reducing children's reliance on the operational patterns and promoting understanding of mathematical equivalence, so we used this as our active control condition. We compared this active control to a comprehensive intervention that included nontraditional arithmetic along with three additional components designed to reduce children's reliance on the operational patterns: (a) lessons that introduce the equal sign outside of arithmetic contexts, (b) concreteness fading exercises, and (c) activities that require children to compare and explain different problem formats and problem-solving strategies. Results suggest that the comprehensive intervention helps children construct a better understanding of mathematical equivalence than either an equivalent amount of nontraditional arithmetic practice alone, or business-as-usual. They further suggest that it is feasible for naïve, inexperienced teachers to use in a regular classroom setting right out of the box with no professional development.

We have argued that the intervention works because it deters children from extracting, representing, activating, and applying the overly narrow operational patterns routinely encountered in traditional arithmetic. However, given the change-resistance account's strong focus on the role of traditional arithmetic practice in children's reliance on the operational patterns (McNeil, 2014), one might wonder why nontraditional arithmetic practice by itself (i.e., the active control condition) is not enough to carry the effect. In the following paragraphs, we discuss the potential mechanisms involved in the beneficial effects of nontraditional arithmetic practice, along with the theoretical and practical reasons for why the intervention needed to go beyond nontraditional arithmetic practice in order to fully deter children from extracting, representing, activating, and applying the overly narrow operational patterns.

Mechanisms

McNeil et al. (2015) showed that nontraditional arithmetic practice leads to lasting improvements in children's understanding of mathematical equivalence specifically by reducing children's reliance on the operational patterns. These reductions in children's reliance on the operational patterns persist for months after the practice ends and are likely the result of several factors, including: (a) the cognitive conflict that arises when children encounter operations on the right side of the equal sign (Baroody & Ginsburg, 1983; Inhelder, Sinclair, & Bovet, 1974; McNeil et al., 2006; Piaget, 1980; Seo & Ginsburg, 2003; VanLehn, 1996), (b) the insight that an addend pair can be equal to other addend pairs, which may result from exposure to multiple, consecutive examples of the same value that vary in surface details (e.g., for the group of facts that sum to 9: 3 + 6, 5 + 4, 2 + 7, etc.) (Behr et al., 1980; Kieran, 1981; cf. Gelman & Williams, 1998), (c) an increased awareness of the placement of the equal sign that may stem from seeing nontraditional problem formats (McNeil et al., 2011) and from solving problems in which the equal sign is sometimes replaced with relational words such as “is equal to” and “is the same amount as” (Chesney et al., 2018), and (d) a deepening of children’s understanding of transitivity, or bidirectional relations, through the strengthening of connections among addend pairs that have the same sum (Chesney et al., 2014).

Although results of the McNeil et al. (2015) study demonstrated the benefits of nontraditional arithmetic practice with an RCT and provided us with key insights into the mechanisms that contribute to children’s difficulties with mathematical equivalence, it also revealed that nontraditional arithmetic practice is not enough, by itself, to help most children construct a mastery-level understanding of mathematical equivalence. The active control condition in the present experiment also supports this conclusion.

One reason that nontraditional arithmetic practice is not sufficient by itself to promote a mastery-level understanding of mathematical equivalence may be because children’s knowledge of the operational patterns is so central to their understanding of arithmetic that any form of arithmetic practice—even nontraditional arithmetic practice—activates children’s representations of the operational patterns to some degree (cf. Baroody & Ginsburg, 1983; Denmark et al., 1976). Thus, to further reduce children’s likelihood of activating and strengthening the operational patterns, the current intervention exposed children to the equal sign outside of arithmetic contexts (e.g., 7 = 7) first, so they could solidify a relational view before moving on to a variety of traditional and nontraditional arithmetic problem formats (cf. Baroody & Ginsburg, 1983; Denmark et al., 1976; McNeil, 2008; Renwick, 1932). This approach has previously been shown to help children learn the meaning of the equal sign (McNeil, 2008), and it corresponds to the way the equal sign is introduced in China (Li et al., 2008).

Another way that our intervention helped reduce activation of the operational patterns was through concreteness fading exercises that strengthen the mappings between concrete relational contexts (e.g., sharing, balancing a scale) and the corresponding abstract symbols (e.g., Arabic numerals, operators, the equal sign). Research has shown that children demonstrate a better understanding of mathematical equivalence when problems are presented with concrete materials than when they are presented with abstract symbols (Sherman & Bisanz, 2009). Thus, starting instruction with concrete materials allows children to activate meaningful, helpful relational knowledge rather than the narrow, unhelpful operational patterns. Through the concreteness fading method, this knowledge can be explicitly linked to the abstract symbols and then faded away, so the abstract symbols can be understood in terms of the familiar, concrete knowledge (Fyfe et al., 2014; Goldstone & Son, 2005). Experience with the concrete materials also allows children to create a store of images (e.g., a balanced scale) that embody the abstract symbols, and those stored representations can later be used as a means of merging novel problems into the repertoire of mastered problems (Bruner, 1966). Fyfe and colleagues (2015) experimentally tested the concreteness fading approach for teaching children how to solve mathematical equivalence problems and found that children who received a lesson in a concrete-to-abstract fading condition constructed a better understanding of mathematical equivalence than did children in concrete only, abstract only, or abstract-to-concrete progression conditions. Moreover, consistent with the idea that concreteness fading reduces children’s reliance on the operational patterns, even when children were incorrect, children in the concrete-to-abstract fading condition were least reliant on the operational patterns after instruction. Specifically, children who received concrete-to-abstract instruction were the least likely to solve transfer problems by adding all of the numbers.

The final way our intervention helped to reduce children's reliance on the operational patterns and increase relational thinking was by requiring children to compare and explain different problem formats (e.g., 7 = 5 + 2 and 5 + 2 = 7) and different problem-solving strategies (e.g., comparing a correct strategy with an incorrect "Add All" strategy). This component was crucial to the success of the intervention because it requires children to articulate and justify generalizations about the underlying structure and properties of arithmetic (Carpenter et al., 2003; Jacobs et al., 2007). Through children's comparison of true and false statements and explanations of their rationale to peers and the teacher, children infer new knowledge and generate fundamental generalizations about the properties of arithmetic and mathematical equivalence (see Carpenter et al., 2003; Jacobs et al., 2007). These processes encourage children to be mindful and attentive to the representations and strategies that apply to the current situation and potential future situations instead of being mindlessly committed to a "single, rigid perspective and…oblivious to alternative ways of knowing" (Langer, 2000, p. 220). Several studies have shown that comparison and explanation facilitate children's understanding of mathematical equivalence (e.g., Hattikudur & Alibali, 2010; Jacobs et al., 2007; Rittle-Johnson, 2006; Rittle-Johnson & Star, 2007; Siegler, 2002). Moreover, consistent with the idea that these processes reduce children's reliance on the operational patterns, even when children solve mathematical equivalence problems incorrectly, children who are given interventions in which they are asked to compare and explain tend to invent new strategies, rather than sticking rigidly to the strategy of adding all the numbers (Rittle-Johnson, 2006; Siegler, 2002).

One implication for educators is that they will need to do more than simply vary the format in which arithmetic problems are presented during arithmetic practice to help children achieve mastery. Even though the active control had the same number of intervention sessions (and thus more nontraditional arithmetic practice overall), it fell short of the comprehensive intervention in terms of improving understanding. This suggests that well-structured procedural practice with arithmetic can promote conceptual knowledge (as shown by McNeil et al., 2015), but falls short when compared to more conceptually-focused instruction that includes scaffolding to help children link the math symbols to their everyday knowledge and explicit, discursive justification processes (as recommended by mathematics educators such as Carpenter et al., 2003; Schliemann et al., 2007; Stephens et al., 2015).

Overall, the evidence suggests that our comprehensive intervention did what it was designed to do. It reduced children's tendency to extract, represent, activate, and apply the overly narrow operational patterns routinely encountered in arithmetic, and it led many children to develop a mastery-level understanding of mathematical equivalence (in support of the change-resistance account, see McNeil, 2014; McNeil & Alibali, 2005b). Performance after our intervention surpassed that of our best-case benchmarks and the active control intervention. However, we fell short of our goal of 100% of children demonstrating basic understanding and at least 50% demonstrating mastery. The primary area where we fell short of achieving our stated goals was in terms of the percentage of children who defined the equal sign relationally. Recall that 45% of children in the comprehensive intervention classrooms met mastery criteria for equation solving and encoding, but only 38% defined the equal sign relationally. Thus, the intervention needs further modifications to highlight the relational nature of the equal sign. The next iteration of the intervention will include a greater proportion of problems using relational words (e.g., “is equal to”) in place of the equal sign, and it will include lessons in which the teacher explicitly teaches the relational definition by having children gesture to the two sides of an equation while saying "I want to make one side equal to the other side” (Cook et al., 2008).

Limitations and Future Directions

Results of the present study suggest that we now have a promising intervention for teaching mathematical equivalence to elementary school children. However, our ability to make strong, generalizable causal claims about the promise of the intervention is limited for at least two reasons: (a) we did not conduct a large-scale efficacy study and (b) we did not follow children over time to see if benefits lasted. We randomly assigned a small group of classrooms in different schools to conditions. Although the intervention and active control classrooms were well matched on several dimensions (e.g., pretest performance, type of school, classroom demographics, teacher experience), the possibility that teacher and school effects are influencing our results remains. In terms of longer-term effects, we know that the benefits of nontraditional arithmetic practice alone last at least six months (McNeil et al., 2015), but it is possible that the effects of the comprehensive intervention would diminish over time.

Future work could test the efficacy of the intervention by randomly assigning a large number of classrooms to the intervention or to a control intervention and examining outcomes immediately and longitudinally. Three studies to date have shown that individual differences in understanding of mathematical equivalence in the early years predicts future mathematics achievement (Hornburg, Devlin, & McNeil, 2018; McNeil et al., 2017; Matthews & Fuchs, 2018), but the skills that prospectively predict later outcomes in correlational studies do not always show up as causal effects in experimental work (Bailey, Duncan, Watts, Clements, & Sarama, 2018), so there is reason to be cautious. However, even without the large-scale longitudinal RCT, the current evidence is suggestive. We were able to compare the performance of children to decades of work serving as business-as-usual and best-case benchmarks, and the benefits of our comprehensive intervention were also evident when compared with an active control group that received a nontraditional arithmetic practice intervention previously shown to improve understanding of mathematical equivalence.

Another potential limitation of the present study is that the sample of children that we worked with may not be representative of the broader U.S. elementary student population. Although our classrooms were from schools across the U.S., we worked with new first- and second-year teachers enrolled in a “service through teaching” M.Ed. program teaching in under-resourced parochial schools. We consider the fact that our intervention was successful in this sample as well as in the two dramatically different samples from our previous design experiments to be a major strength; however, it is possible that the teachers we worked with differ from the average teacher. All teachers were in their twenties and held bachelor’s degrees from top universities, and none of them attended a traditional teacher training program. It is possible that they have better understanding of mathematics, less math anxiety, more motivation to serve children, and/or more willingness to carry out interventions than the average teacher. It is also possible that children may be more open or receptive to instruction from these younger teachers, who they may perceive to be role models. Future research needs to compare our comprehensive intervention to an active control intervention in a large-scale randomized experiment to determine efficacy in more typical classrooms where teachers may not be as invested in the content or as strictly following our recommendations.

Finally, we developed the initial scope and sequence of our comprehensive intervention by combining component interventions that had previously been shown to improve children’s understanding of mathematical equivalence by reducing reliance on the operational patterns, so we cannot pinpoint the exact element or elements that are necessary for improving outcomes to levels that surpass nontraditional arithmetic practice. Moreover, in the time since the initial development of our intervention, additional research has been conducted on the mechanisms involved in understanding of mathematical equivalence, and additional malleable factors have been identified, such as the timing of problem-solving practice and conceptual instruction (DeCaro & Rittle-Johnson, 2012; Fyfe, DeCaro, & Rittle-Johnson, 2014; Fyfe et al., 2012; Loehr, Fyfe, & Rittle-Johnson, 2014) and inclusion of activities to teach children the substitutive meaning of the equal sign (Jones, Inglis, Gilmore, & Evans, 2013). Thus, future iterations of the intervention may benefit from incorporating these and others that future research may identify to further reduce reliance on the operational patterns and deepen children's understanding.

Mathematical equivalence is considered one of the "Big Ideas" in mathematics (Charles, 2005), and it is widely regarded as one of the most important concepts for developing children's algebraic thinking (Blanton & Kaput, 2005; Carpenter et al., 2003; Kieran, 1992; Knuth et al., 2006) and facilitating children's mathematical achievement (McNeil et al., 2017). Despite the progress the field has made over the past few decades in terms of understanding the nature of children's difficulties with mathematical equivalence, this theoretical progress has not been translated into large-scale pragmatic changes in the ways children are taught. If we expect schools to make large-scale changes, then we need to provide them with a research-based, comprehensive intervention that is effective, affordable, portable, and easy to administer. The results of this study suggest we have met this need. Our intervention produces mastery understanding of mathematical equivalence in many children at a level that far surpasses business-as-usual and component interventions. Once these results are validated with a large-scale RCT, our intervention will be ready for widespread use by elementary school teachers across the U.S., and we can see if the algebra readiness of U.S. students improves as a result. At the same time, researchers can use the intervention to test the long-term effects of having an early understanding of mathematical equivalence. This is a critical need in the literature, as researchers assume that a better understanding of mathematical equivalence in elementary school translates to improvements in algebra readiness, but this casual assumption has never been directly tested. More generally, the present work demonstrates how basic research in the learning and cognitive sciences can be combined with the advice and recommendations of expert practitioners to design interventions that improve our understanding of how to help children learn foundational mathematical concepts.

References

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A., C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning*, *9*, 221-246. doi:10.1080/10986060701360902

Alibali, M. W., & Meredith, L. (2009). Sources of change in children's encoding of mathematics problems. In B. Rittle-Johnson (Organizer), *Understanding knowledge change: Investigations on how children learn mathematics and literary skills*. Symposium presented at the biennial meeting of the Cognitive Development Society, San Antonio, TX.

Alibali, M. W., Phillips, K. M. O., & Fischer, A. D. (2009). Learning new problem-solving strategies leads to changes in problem representation. *Cognitive Development, 24,* 89-101. doi:10.1016/j.cogdev.2008.12.005

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *American Psychologist*, *73*, 81-94. doi:10.1037/amp0000146

Ball, D. L. (1993) With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *Elementary School Journal*, *93*, 373-397. doi:10.1086/461730

Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal*, *84*, 198-212. doi:10.1086/461356

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

Blanton, M. L., & Kaput, J. J. (2003). Developing elementary teachers' "algebra eyes and ears."

Teaching Children Mathematics, *10*, 70-77.

Blanton, M. L., & Kaput, J. J. (2005). Characterizing a classroom practice that promotes algebraic reasoning. *Journal for Research in Mathematics Education, 36*, 412-446.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education*, *29*, 41-62. doi:10.2307/749717

Boaler, J. (2000). Exploring situated insights into research and learning. *Journal for Research in Mathematics Education*, *31*, 113-119. doi:10.2307/749822

Boaler, J. (2002). Exploring the nature of mathematical activity: Using theory, research and ‘working hypotheses' to broaden conceptions of mathematics knowing. *Educational Studies in Mathematics*, *51*, 3-21.

Brown, M. C., McNeil, N. M., & Glenberg, A. M. (2009). Using concreteness in education: Real problems, potential solutions. *Child Development Perspectives, 3*, 160-164. doi:10.1111/j.1750-8606.2009.00098.x

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

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research and Practice*, *18*, 99-111. doi:10.1111/1540-5826.00066

Burns, M. (1996). How to make the most of math manipulatives. *Instructor, 105*, 45-51.

Byrd, C. E., McNeil, N. M., Brletic-Shipley, H., & Matthews, J. M. (2013, September). *Development of a comprehensive intervention to improve children’s understanding of math equivalence.* Poster presented at the Society for Research on Educational Effectiveness Fall Conference (SREE), Washington, D.C.

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.lindiff.2015.01.001

Callender, J.C., Osburn, H. G. (1979). An empirical comparison of Coefficient Alpha, Guttman's Lambda – 2, and MSPLIT maximized split-half reliability estimates. *Journal of Educational Measurement, 16*, 89–99.

Capraro, R. M., Capraro, M. M., Yetkiner, A. E., Ozel, S., Kim, H. G., & Corlu, M. S. (2010). An international comparison of grade 6 students' understanding of the equal sign. *Psychological Reports, 106*, 49-53. doi:10.2466/pr0.106.1.49-53

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, *26*, 499-531. doi:10.3102/00028312026004499

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

Charles, R. I. (2005). Big ideas and understandings as the foundation for elementary and middle school mathematics. *National Council for Supervisors of Mathematics: Journal of Mathematics Education Leadership*, *7*, 9-24.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81. doi:10.1016/0010-0285(73)90004-2

Chesney, D. L., McNeil, N. M., Matthews, P. G., Byrd, C. E., Petersen, L. A., Wheeler, M. C., … 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

Chesney, D. L., McNeil, N. M., Petersen, L. A., & Dunwiddie, A. E. (2018). Arithmetic practice that includes relational words promotes understanding of symbolic equations. *Learning and Individual Differences*, *64*, 104-112. doi:10.1016/j.lindif.2018.04.013

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73-105. doi:10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477. doi:10.1207/s15516709cog1803_3

Cobb, P., McClain, K., de Silva Lamberg, T., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*, 13-24. doi:10.3102/0013189x032006013

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*, 1047-1058. doi:10.1016/j.cognition.2007.04.010

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*, 552-568. doi:10.1016/j.jecp.2012.06.009

Denmark, T., Barco, E., & Voran, J. (1976). *Final report: A teaching experiment on equality. PMDC Technical Report No. 6.* Tallahassee, FL: Florida State University.

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

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology*. *84,* 502-519. doi:10.1111/bjep.12035

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

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

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*, 1094-1108. doi:10.1037/a0028389

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development*, *67*, 2022-2044. doi:10.2307/1131607

Gelman, R., & Williams, E. (1998). Enabling constraints for cognitive development and learning: Domain specificity and epigenesis. In D. Kuhn & R. S. Siegler, (Eds.). *Handbook of child psychology: Cognition, perception and language.* (5^{th} ed., Vol. 2, pp. 575-630). New York: Wiley.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*, 424-436. doi:10.1037/0022-0663.96.3.424

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, *91*, 720-730. doi:10.1037/0022-0663.91.4.720

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414-466. doi:10.1016/s0010-0285(02)00519-4

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

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, *9*, 495-522. doi:10.1002/acp.2350090604

Gravemeijer, K. (2002). Preamble: from models to modeling. In K. Gravemeijer, R. Lehrer, B. Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 7-22). Dordrecht, The Netherlands: Kluwer. doi:10.1007/978-94-017-3194-2_2

Hattikudur, S., & Alibali, M. W. (2010). Learning about the equal sign: Does comparing with inequality symbols help? *Journal of Experimental Child Psychology*, *107*, 15-30. doi:10.1016/j.jecp.2010.03.004

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K., Human, P., Murray, H., Oliver, A., & Wearne, D. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher*, *25*, 12-21. doi:10.3102/0013189x025004012

Hornburg, C. B., Devlin, B. L., & McNeil, N. M. (2018). *Grade of acquisition of understanding of mathematical equivalence in elementary school predicts algebra readiness in middle school.* Manuscript in preparation.

Hornburg, C. B., Rieber, M. L., & McNeil, N. M. (2017). An integrative data analysis of gender differences in children's understanding of mathematical equivalence. *Journal of Experimental Child Psychology*, *163*, 140-150. doi:10.1016/j.jecp.2017.06.002

Inhelder, B., Sinclair, H., & Bovet, M. (1974). *Learning and the development of cognition.* (*Trans S. Wedgwood*). Cambridge, MA: Harvard University Press.

Institute of Educational Sciences. (2014). What works clearinghouse: Procedures and standards handbook (Version 3). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc procedures v3 0 standards handbook.pdf

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.

Jones, I., Inglis, M., Gilmore, C. K., & Evans, R. (2013). Teaching the substitutive conception of the equals sign, *Research in Mathematics Education, 15*, 1-16.

Kaminski, J., Sloutsky, V. M., & Heckler, A. F. (2006). Effects of concreteness on representation: an explanation for differential transfer. In R. Sun & N. Miyake (Eds.).

Proceedings of the 28th Annual Conference of the Cognitive Science Society (pp. 1581-1586). Austin, TX: Cognitive Science Society.

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

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

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390-419). New York: MacMillan.

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

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*, 29-63. doi:10.3102/00028312027001029

Landy, D., & Goldstone, R. L. (2007). The alignment of ordering and space in arithmetic computation. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.) *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 437-442). Austin, TX: Cognitive Science Society.

Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science*, *9*, 220-223. doi:10.1111/1467-8721.00099

Lehrer, R., & Schauble, L. (2002). Symbolic communication in mathematics and science: co-constituting inscription and thought. In E. D. Amsel, & J. Byrnes (Eds.), *Language,* *literacy, and cognitive development: The development of consequences of symbolic communication* (pp. 167-192). Mahwah, NJ: Erlbaum. doi:10.4324/9781410601452

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

Loehr, A. M., Fyfe, E. R., & Rittle-Johnson, B. (2014). Wait for it… Delaying instruction improves mathematics problem solving: A classroom study. *The Journal of Problem Solving*, *7*, 36-49. doi:10.7771/1932-6246.1166

MacGregor, M., & Stacey, K. (1999). A flying start to algebra. *Teaching Children Mathematics*, *6*, 78-85.

Martin, T. (2009). A theory of physically distributed learning: How external environments and internal states interact in mathematics learning. *Child Development Perspectives*, *3*, 140-144. doi:10.1111/j.1750-8606.2009.00094.x

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*, 587-625. doi:10.1207/s15516709cog0000_15

Matthews, P. G., & Fuchs, L. S. (2018). Keys to the gate? Equal sign knowledge at second grade predicts fourth- grade algebra competence. *Child Development*. Advance online publication. doi:10.1111/cdev.13144

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

McNeil, N. M. (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. M. (2008). Limitations to teaching children 2 + 2 = 4: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development*, *79*, 1524-1537. doi:10.1111/j.1467-8624.2008.01203.x

McNeil, N. M. (2014). A change–resistance account of children's difficulties understanding mathematical equivalence. *Child Development Perspectives*, *8*, 42-47. doi:10.1111/cdep.12062

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

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451-466. doi:10.1207/s15516709cog2803_7

McNeil, N. M., & Alibali, M. W. (2005a). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition and Development*, *6*, 285-306. doi:10.1207/s15327647jcd0602_6

McNeil, N. M., & Alibali, M. W. (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/j.1467-8624.2005.00884.x

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

McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of mathematical equivalence. *Journal of Educational Psychology, 107,* 423-436. doi:10.1037/a0037687

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & 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., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. S., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, *24,* 367-385. doi:10.1207/s1532690xci2403_3

McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C., & McKeever, M. O. (2017). Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Development*. Advance online publication. doi:10.1111/cdev.12948

McNeil, N. M., Hornburg, C. B., Fuhs, M. W., & O'Rear, C. D. (2017). Understanding children's difficulties with mathematical equivalence. In D. C. Geary, D. B. Berch, R. Ochsendorf, & K. Mann Koepke (Eds.), *Mathematical cognition and learning* (Vol. 3, pp. 167-195). San Diego, CA: Elsevier Academic Press.

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

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

Miller, S. P., & Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research and Practice*, *22*, 47-57. doi:10.1111/j.1540-5826.2007.00230.x

Montessori, M. (1917). *The advanced Montessori method.* New York: Frederick A. Stokes.

Moses, R., & Cobb, C. (2001). *Radical equations: Math literacy and civil rights*. Boston, MA: Beacon Press.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

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

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., 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. (1970). *Structuralism.* New York: Basic Books.

Piaget, J. (1980). The psychogenesis of knowledge and its epistemological significance. In M.

Piattelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 1-23). Cambridge, MA: Harvard University Press.

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal*, *112*, 627-648.

Renwick, E. (1932). Children's misconceptions concerning the symbols for mathematical equality. *British Journal of Educational Psychology*, *2*, 173-183. doi:10.1111/j.2044-8279.1932.tb02743.x

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*, 1-15. doi:10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., & Alibali, M. W. (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., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*, 561-574. doi:10.1037/0022-0663.99.3.561

Roscoe, R., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*, 534-574. doi:10.3102/0034654307309920

Schliemann, A. D., & Carraher, D. W. (2002). The evolution of mathematical reasoning: Everyday versus idealized understandings. *Developmental Review*, *22*, 242-266. doi: 10.1006/drev.2002.0547

Schliemann, A., Carraher, D., & Brizuela, B. (2007). *Bringing out the algebraic character of arithmetic: From children's ideas to classroom practice*. Mahwah, NJ: Erlbaum.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*, 475-522. doi:10.1207/s1532690xci1604_4

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: Constructing adaptive expertise* (pp. 161-187). Mahwah, NJ: Erlbaum. doi:10.5860/choice.41-1688

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

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, *8*, 481-520. doi:10.1016/0010-0285(76)90016-5

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). New York: Cambridge University. doi:10.1017/CBO9780511489709.002

Sloutsky, V. M., Kaminski, J. A., & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin and Review*, *12*, 508-513. doi:10.3758/bf03193796

Smith, J. P. (1996). Efficacy and teaching mathematics by telling: A challenge for reform. *Journal for Research in Mathematics Education*, *27*, 387-402. doi:10.2307/749874

Stephens, A., Blanton, M., Knuth, E., Isler, I., & Gardiner, A. M. (2015). Just say yes to early algebra!. *Teaching Children Mathematics*, *22*, 92-101. doi:10.5951/teacchilmath.22.2.0092

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. *Journal of Mathematical Behavior*, *32*, 173-182. doi:10.1016/j.jmathb.2013.02.001

VanLehn, K. (1996). Conceptual and meta learning during coached problem solving. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *ITS96: Proceedings of the Third International Conference on Intelligent Tutoring Systems* (pp. 29-47). New York: Springer-Verlag. doi:10.1007/3-540-61327-7_99

Weaver, J. F. (1973). The symmetric property of the equality relation and young children's ability to solve open addition and subtraction sentences. *Journal for Research in Mathematics Education*, *4*, 45-56. doi:10.2307/749023

What Works Clearinghouse (2014). WWC review of the report: Benefits of practicing 4 = 2 + 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence [single study review]. Retrieved from http://whatworks.ed.gov

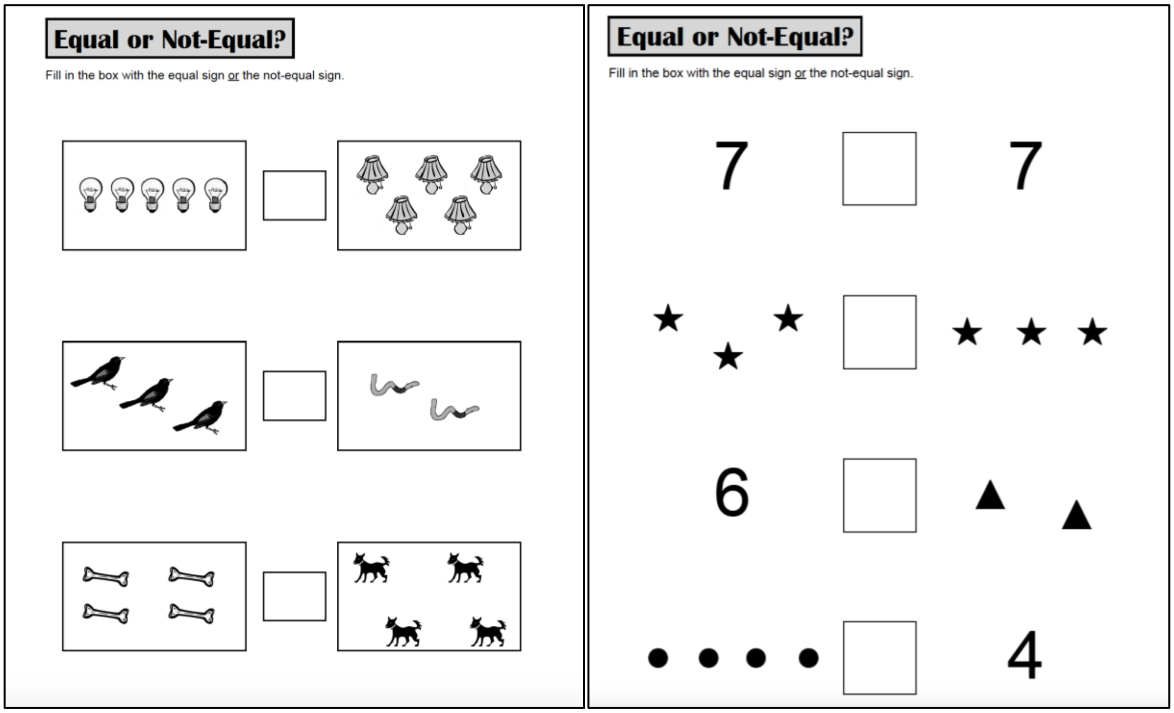


Figure 1. Examples of activities presenting “=” outside of arithmetic contexts. Children practiced using = and ≠.

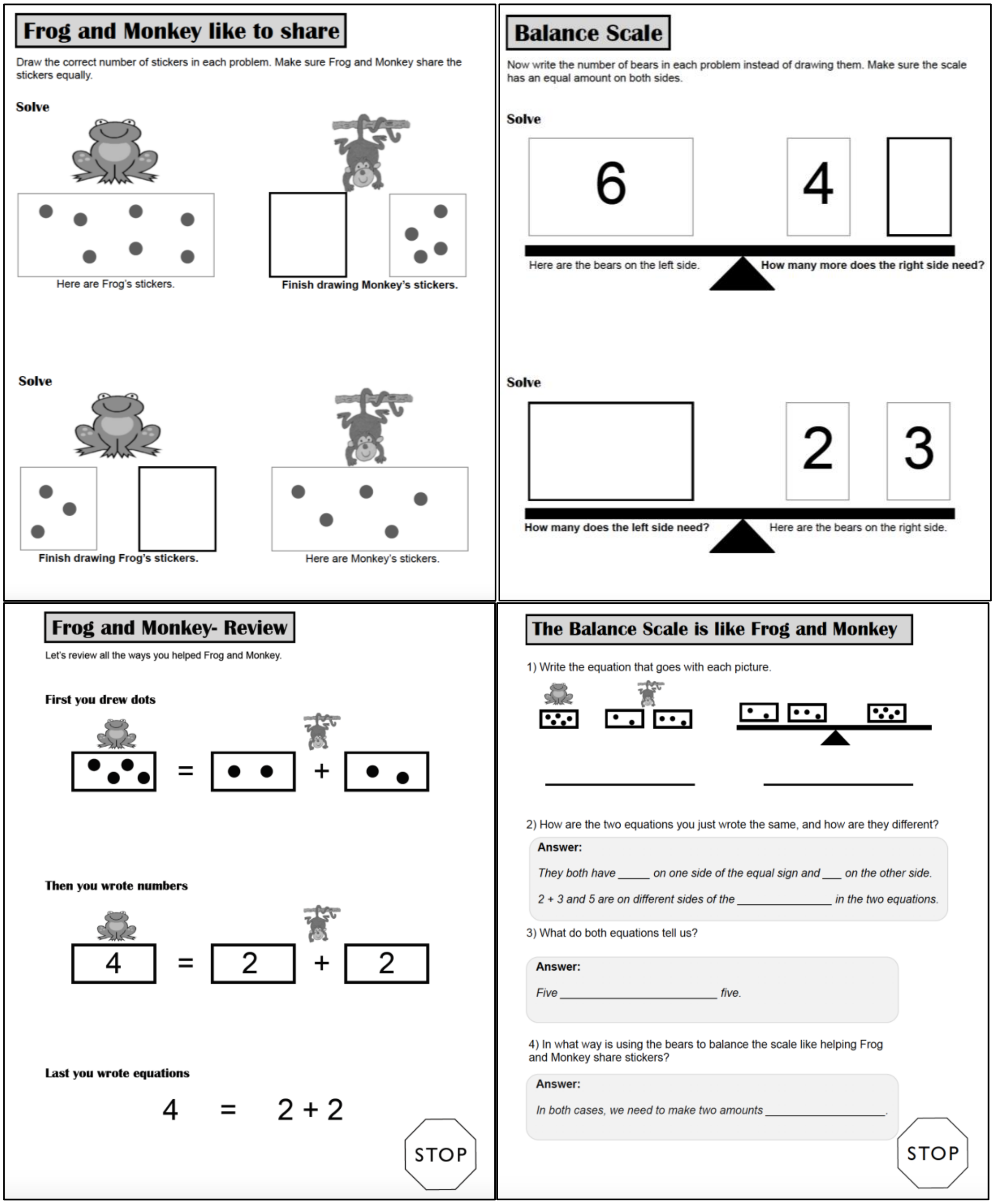


Figure 2. Examples of "concreteness fading" activities. After examples using physical objects (e.g., stuffed animals sharing stickers or bears on a balance scale) teachers would direct children to practice with workbook pages using dots or numerical symbols tied to the concrete context (top left and top right panels). Then, children would practice with only abstract symbols. At the end of a lesson teachers would review these various instantiations of equality as a whole class (bottom left panel) and discuss how various concrete contexts are similar (bottom right panel).

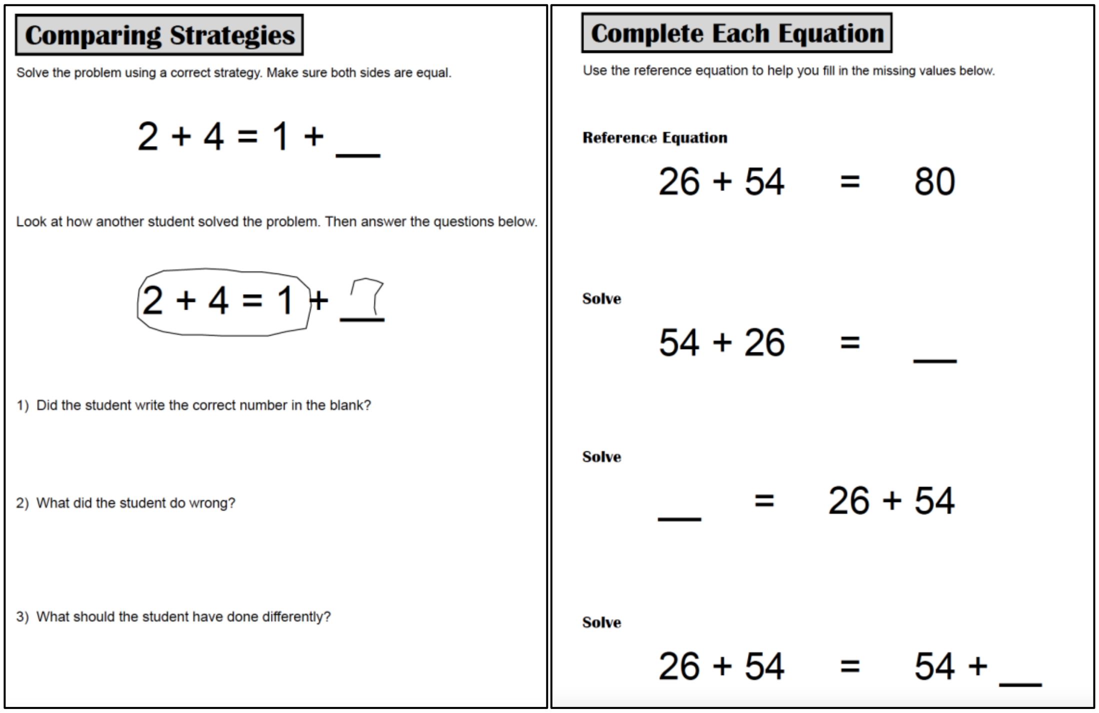


Figure 3. Examples of activities requiring children to compare and contrast problem-solving strategies (left panel) and problem formats (right panel). For the left panel, the teacher led the class through a discussion about the two children's strategies, using prompts such as "Why was her strategy correct?" and "He wasn't paying attention to where the equal sign was. What did he need to do instead to solve this problem?" For the right panel, the teacher read these instructions: "On this page you will see one complete equation and three equations you will need to complete. These equations use really big numbers, so I want you to use the completed equation at the top of your workbook page to help you figure out what number goes in the blank in each equation."

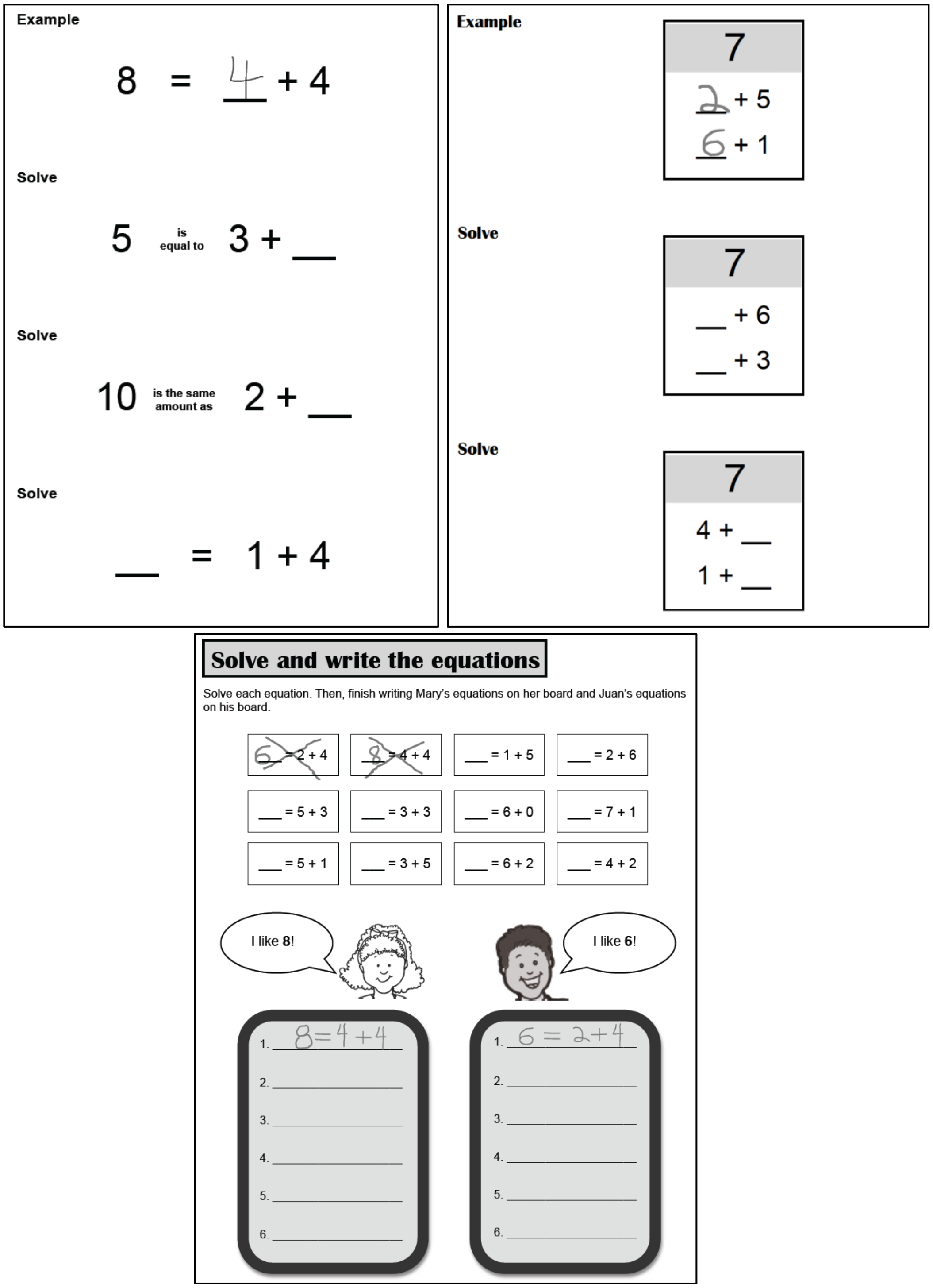


Figure 4. Examples of nontraditional arithmetic practice – relational words in place of the equal sign (top left panel), problems grouped by equivalent sums (top right panel), and problems written in the nontraditional $a = b + c$ format (bottom panel).

Table 1

Definitions of Basic and Mastery Understandings of Mathematical Equivalence

Outcome	Basic understanding	Mastery understanding
Solving	Solve at least one math equivalence problem in a given set correctly	Solve at least 75% of math equivalence problems in a given set correctly
Encoding	Encode at least one math equivalence problem in a given set correctly	Encode at least 75% of math equivalence problems in a given set correctly
Defining =	N/A	Provide a relational definition

Table 2

Percentage of Children Reaching Basic and Mastery Understandings of Mathematical Equivalence After Participating in Various Conditions

Condition	% children exhibiting basic understanding	% children exhibiting mastery understanding
Business-as-usual	15%	2%
Well-structured arithmetic practice only ("best-case")	40%	4%
Comprehensive intervention (our stated goals for this study)	100%	≥50%

Table 3

Characteristics of Classrooms in Each Condition

Condition	Classroom	State	Race/Ethnicity of Children in the School	Math Textbook	# Children in the Class	
					Pre	Post
Comprehensive	1	MS	1% Asian, 18% Black, 2% Hispanic, 79% White	Saxon Math	24 (13 boys, 11 girls)	24 (13 boys, 11 girls)
Comprehensive	5	FL	93% Black, 6% Hispanic, 1% Other	Progress in Mathematics	17 (9 boys, 8 girls)	16 (8 boys, 8 girls)
Comprehensive	6	TX	98% Hispanic, 2% White	Progress in Mathematics	20 (9 boys, 11 girls)	18 (8 boys, 10 girls)
Comprehensive	7	CA	1% Asian, 75% Black, 24% Hispanic	Progress in Mathematics	25 (7 boys, 18 girls)	20 (6 boys, 14 girls)
Active control	2	FL	5% Black, 85% Hispanic, 8% White, 2% Other	Everyday Math	22 (7 boys, 15 girls)	21 (7 boys, 14 girls)
Active control	3	TX	97% Hispanic, 3% White	Saxon Math	15 (6 boys, 9 girls)	13 (5 boys, 8 girls)
Active control	4	GA	5% Asian, 10% Black, 18% Hispanic, 63% White, 3% Other	Progress in Mathematics	19 (10 boys, 9 girls)	19 (10 boys, 9 girls)
Active control	8	DC	100% Black	Progress in Mathematics	11 (1 boy, 10 girls)	11 (1 boy, 10 girls)

Table 4

Performance of Classrooms in Each Condition at Pretest (N = 153)

Condition and Classroom	Problem Solving out of 8 *M* (*SD*)	Problem Encoding out of 4 *M* (*SD*)	Equal Sign Defining (% children defining relationally)	Basic Understanding of Math Equivalence (% children)	Mastery Understanding of Math Equivalence (% children)	Computational Fluency *M* (*SD*)
Comprehensive 1	0.63 (1.74)	0.46 (0.78)	4	4	0	13.71 (5.33)
Comprehensive 5	0.29 (0.59)	0.94 (1.20)	24	18	0	6.82 (4.42)
Comprehensive 6	0.30 (0.73)	0.45 (0.89)	5	5	0	8.55 (3.75)
Comprehensive 7	0.48 (0.92)	1.84 (1.46)	8	28	0	11.40 (4.39)
Condition *M*	0.44 (1.12)	0.95 (1.26)	9	14	0	10.48 (5.17)
M of Classrooms	0.42 (0.16)	0.92 (0.65)	10 (*SD* = 9)	14 (*SD* = 11)	0	10.12 (3.05)
Active control 2	0.77 (1.82)	0.36 (0.73)	14	9	0	6.73 (2.31)
Active control 3	0.67 (1.05)	1.00 (1.00)	0	27	0	7.07 (3.31)
Active control 4	0.26 (0.65)	1.21 (1.27)	11	16	0	16.84 (5.90)
Active control 8	0.27 (0.90)	0.64 (1.03)	0	9	0	11.55 (4.57)
Condition *M*	0.52 (1.26)	0.79 (1.05)	7	15	0	10.46 (5.99)
M of Classrooms	0.49 (0.26)	0.80 (0.38)	6 (*SD* = 7)	15 (*SD* = 8)	0	10.55 (4.74)

Note. Condition means were calculated by averaging across all children in the given condition. Means of classrooms were calculated by averaging the four classroom means in the given condition.

Table 5

Elements That Were Included in Both the Comprehensive and Active Control Interventions

Element	Description	References
Modified arithmetic practice	Procedural practice with well-structured nontraditional arithmetic problems, including nontraditional formats (e.g., __ = 8 + 4), problems organized by equivalent values (e.g., 5 + 2, 3 + 4, 1 + 6), and relational phrases such as "is the same amount as" and "is equal to" in place of the equal sign (see Figure 4). This began in week 3 of the comprehensive intervention (after four sessions outside the arithmetic context) and then continued throughout.	Chesney et al. (2018); McNeil et al. (2012); McNeil et al. (2015); McNeil et al. (2011)
Space around the equal sign	Early sessions included equations with substantial space around the equal sign (e.g., __ = 4 + 3), and the amount of space was gradually reduced across sessions, so that by the second half of the intervention problems had equal space around addends and the equal sign (e.g., 12 = __ + 5).	Alibali & Meredith (2009); Landy & Goldstone (2007)
Tips and suggestions in the teacher manual	We recommended: (a) leaving extra space around the equal sign whenever writing equations on the board for the first half of the intervention to help children distinguish between the two sides of an equation; (b) getting into the habit of saying "is equal to" rather than "equals" for the equal sign whenever reading a math problem aloud; (c) avoiding asking children "What's the answer?" and instead asking "What number should go in the blank?" to prevent activation of operational patterns such as "add up all the numbers."	McNeil & Alibali (2005b); Rittle-Johnson & Alibali (1999)

Table 6

Elements That Were Included in the Comprehensive Intervention Only

Element	Description	References
Equal sign outside of arithmetic context	The first four sessions included lessons/activities that introduced the equal sign outside of arithmetic contexts (e.g., 7 = 7) to enhance children's relational thinking and understanding of the equal sign. This involved comparing two amounts (represented in pictures and/or symbolically) using = or ≠ (see Figure 1). Children then moved on to seeing the = in the nontraditional arithmetic practice that paralleled the active control, as well as in problem solving exercises with operations on both sides of the equal sign (e.g., 3 + 4 = 3 + __).	McNeil (2008)
Concreteness fading	Early sessions included a high density of concrete representations (e.g., sharing, balancing a scale) relative to abstract representations (e.g., Arabic numerals). Middle sessions included a high density of concreteness fading exercises to help children make explicit links between the concrete contexts and the corresponding abstract symbols. For example, children helped two stuffed animals share stickers equally and balanced a scale (see Figure 2). Gestures were used to help ground abstract concepts. For example, during lessons with the balance scale, teachers were asked to make a level gesture when saying the word "balance," (e.g., both palms up at their sides at the same level like a balanced scale, parallel to the ground). Later sessions included a high density of abstract representations relative to concrete representations (e.g., more math symbols and words only, no physical objects or pictures).	Fyfe et al. (2015); Goldin-Meadow, Kim, & Singer (1999)
Comparing, contrasting, and explaining	Later sessions included a high density of lessons/activities that required children to compare and explain different problem formats (e.g., 7 = 5 + 2 compared to 5 + 2 = 7) and problem-solving strategies to improve children's conceptual understanding and procedural flexibility. For example, children contrasted incorrect and correct strategies for solving the same problem and used a solved problem to more easily solve another problem with the same addends arranged in a different format (see Figure 3).	Carpenter et al. (2003); Hattikudur & Alibali (2010); Rittle-Johnson (2006); Siegler (2002)

Table 7

Far Transfer Items on the Posttest Assessment

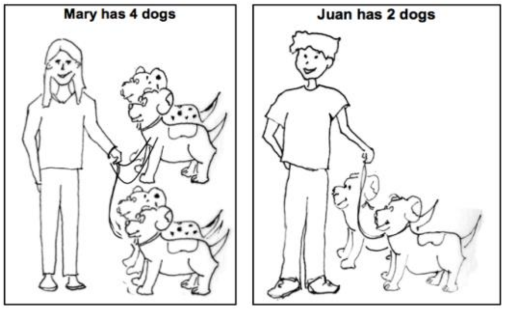
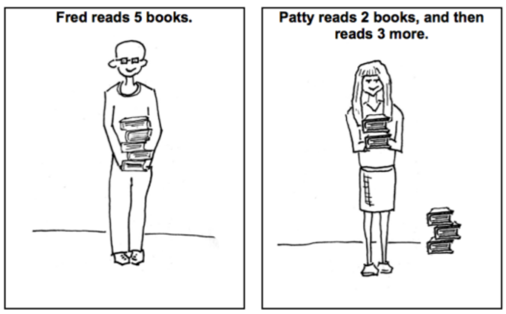
#	Item
1	Write an equation to show the relation between the two children. 
2	Write an equation to show the relation between the two children. 
3	The boys want to have the same number of stars as the girls. Sara has 5 stars. Ashley has 4 stars. Mike has 7 stars. How many stars does Dan need for the boys and girls to have the same? Circle the equation that correctly represents what is happening in the story. 5 + __ = 4 + 7 5 + 4 + 7 = __ 5 + 4 = 7 + __ 5 + 4 + 7 + 2 = __
4	(2 points) Write an equation (1 point) and solve for the missing information in the word problem (1 point). Rebecca and Megan collect sea shells. Rebecca has 2 yellow shells, 5 white shells, and 4 brown shells. Megan has 1 white shell and 2 brown shells. How many more shells does Megan need if she wants to have the same number of shells as Rebecca?
5	Is the number that goes in the $\square$ the same number in the following two equations? <u>Explain your reasoning.</u> $\square + 15 = 31$ $\square + 15 - 9 = 31 - 9$

Table 8

Average Classroom Performance on Each Outcome by Condition

	Comprehensive				Active control				Estimated effect		
Outcome	# children	*n*	*M*	*SD*	# children	*n*	*M*	*SD*	*t*	*p*	*g [95% CI]*
Pre-to-post change in composite understanding of math equivalence	78	4	4.34	1.14	64	4	1.65	1.36	3.05	.023	1.87 [0.24-3.42]
Pre-to-post change in correct solving	78	4	5.98	0.46	64	4	1.35	2.29	3.96	.007	2.44 [0.59-4.20]
Pre-to-post change in correct encoding	78	4	1.28	0.79	64	4	0.80	0.54	1.01	.35	0.62 [-0.65-1.85]
Pre-to-post change in % relational definition	78	4	29	14	64	4	16	26	0.87	.42	0.53 [-0.73-1.75]
% children exhibiting basic understanding at posttest	78	4	83	13	64	4	27	25	4.00	.007	2.46 [0.60-4.23]
% children exhibiting mastery understanding at posttest	78	4	20	9	64	4	1	3	4.22[a]	.017	2.59 [0.68-4.42]
Transfer performance	77	4	6.22	0.70	63	4	3.12	1.70	3.37	.015	2.07 [0.37-3.69]
Pre-to-post change in computational fluency	78	4	5.48	2.21	64	4	6.77	3.20	-0.67	.53	-0.41 [-1.62-0.83]

[a]Homogeneity assumption violated (p = .03), so equal variances were not assumed (df = 3.56).

Table 9

Median Classroom Performance on the Primary Outcomes by Condition

	Comprehensive			Active control			Estimated effect		
	# students	*n*	*Median*	# students	*n*	*Median*	*U*	*p*	η^2
Pre-to-post change in composite understanding of math equivalence	78	4	4.23	64	4	1.85	0	.029	.67
% children in the classroom exhibiting basic understanding at posttest	78	4	81	64	4	19	0	.029	.67
% children in the classroom exhibiting mastery understanding at posttest	78	4	21	64	4	0	0	.029	.70
Transfer performance	77	4	6.00	63	4	3.42	0	.029	.67
Pre-to-post change in computational fluency	78	4	5.39	78	4	6.37	6	.686	.04

Table 10

Performance on the Primary Outcomes by Individual Classrooms in Each Condition

Condition and Classroom	Pre-to-post Change in Composite Understanding of Math Equivalence *M* (SD)	% Children Exhibiting Basic Understanding at Posttest	% Children Exhibiting Mastery Understanding at Posttest	Transfer Performance *M* (*SD*) out of 10	Pre-to-Post Change in Computational Fluency *M* (*SD*)
Comprehensive 1	5.68 (2.16)	100	29	7.25 (1.19)	6.21 (5.93)
Comprehensive 5	3.58 (3.16)	69	25	6.00 (1.93)	4.56 (4.86)
Comprehensive 6	4.88 (2.54)	78	17	5.65 (2.98)	3.00 (3.80)
Comprehensive 7	3.24 (2.54)	85	10	6.00 (1.86)	8.15 (5.68)
Condition *M*	4.44 (2.72)	85	21	6.31 (2.08)	5.63 (5.47)
M of Classrooms	4.34 (1.14)	83 (*SD* = 13)	20 (*SD* = 9)	6.22 (0.70)	5.48 (2.21)
Active control 2	1.11 (1.69)	19	0	2.67 (1.93)	5.29 (4.29)
Active control 3	-0.02 (1.35)	8	0	0.92 (0.86)	7.46 (5.95)
Active control 4	2.89 (2.13)	63	5	4.17 (3.03)	10.89 (5.29)
Active control 8	2.60 (2.60)	18	0	4.73 (2.53)	3.45 (4.72)
Condition *M*	1.66 (2.22)	30	2	3.10 (2.60)	7.08 (5.65)
M of Classrooms	1.65 (1.36)	27 (*SD* = 25)	1 (*SD* = 3)	3.12 (1.70)	6.77 (3.20)

Note. Condition means were calculated by averaging across all children in the given condition. Means of classrooms were calculated by averaging the four classroom means in the given condition.