


An Examination of Third- and Fourth-Graders' Equivalence Knowledge After Classroom Instruction

Emmanuelle Adrien¹ , Helena P. Osana¹ , Rebecca Watchorn Kong² , Jeffrey Bisanz² , Jody Sherman LeVos³

[1] Department of Education, Concordia University, Montreal, QC, Canada. [2] Department of Psychology, University of Alberta, Edmonton, AB, Canada.

[3] LeapFrog Enterprises, Inc., Emeryville, CA, USA.

Journal of Numerical Cognition, 2021, Vol. 7(2), 104–124, <https://doi.org/10.5964/jnc.6913>

Received: 2019-12-17 • Accepted: 2021-04-20 • Published (VoR): 2021-07-23

Handling Editor: Jake McMullen, Department of Teacher Education, University of Turku, Turku, Finland

Corresponding Author: Emmanuelle Adrien, 1455 Boul. de Maisonneuve Ouest, Montreal, QC, Canada, H3G 1M8. E-mail: emmanuelle.adrien@concordia.ca

Abstract

The present correlational study examined third- and fourth-graders' ($N = 56$) knowledge of mathematical equivalence after classroom instruction on the equal sign. Three distinct learning trajectories of student equivalence knowledge were compared: those who did not learn from instruction (Never Solvers), those whose performance improved after instruction (Learners), and those who had strong performance before instruction and maintained it throughout the study (Solvers). Learners and Solvers performed similarly on measures of equivalence knowledge after instruction. Both groups demonstrated high retention rates and defined the equal sign relationally, regardless of whether they had learned how to solve equivalence problems before or during instruction. Never Solvers had relatively weak arithmetical (nonsymbolic) equivalence knowledge and provided operational definitions of the equal sign after instruction.

Keywords

equivalence knowledge, equal sign, classroom instruction, learning trajectories

Long-term effects of learning algebra in high school include exposure to advanced mathematics courses, higher academic performance by the end of high school, and increased graduation rates in college (Stein et al., 2011). Given algebra's important role in students' academic achievement, increased attention has been paid to supporting children's algebraic thinking at the elementary level (Blanton et al., 2015; Kaput et al., 2008). The focus of the present study is on children's understanding of the equal sign in elementary school, which is regarded as a central component of algebraic thinking (Carpenter et al., 2005; Empson et al., 2011; Knuth et al., 2006). Young children's early understanding of the equal sign has been found to be predictive of their algebraic knowledge later in elementary school (Matthews & Fuchs, 2020). Yet, children hold deep-seated misconceptions about what the symbol means (Jacobs et al., 2007; Kieran, 1981; McNeil & Alibali, 2005; Seo & Ginsburg, 2003). Many children hold an *operator view*, which entails interpreting the symbol as a signal to “do” something rather than as a relation between the amounts on either side of it. In contrast, children have a *relational view* when they know the symbol means that the expressions on either side are the same in amount. Other related concepts have been documented, such as a *substitutive view*, characterized by interpreting the symbol as an indication that one side can replace the other (Jones et al., 2012).

One way that children's interpretations of the equal sign can be expressed is through their performance on *mathematical equivalence problems*, such as $8 + 4 + 2 = _ + 7$. Mathematical equivalence problems are open-number sentences with operations on both sides of the equal sign, thus making them non-standard or “non-canonical.” They



stand in contrast to canonical problems more commonly seen in school, which contain operations to the left of the equal sign and only one number to the right (e.g., $a + b = _$; McNeil, 2014; McNeil et al., 2012). Children who hold a relational view of the equal sign understand that the amount on the left side of the equation must be the same as that on the right and often consider how the amounts on both sides relate to each other before pursuing strategies for solving the problem (Carpenter et al., 2003; Kindrat & Osana, 2018; Rittle-Johnson et al., 2011).

Children perform poorly on non-canonical equivalence problems (for a review, see McNeil et al., 2017). On the whole, instructional interventions have been successful in increasing their performance (Alibali, 1999; Goldin-Meadow et al., 2009; Perry, 1991; Rittle-Johnson, 2006; Rittle-Johnson & Alibali, 1999), although knowledge gains may sometimes fade quickly with children reverting back to their entrenched misconceptions (e.g., Cook et al., 2008). Not all students respond to instruction in the same way, however, which can result in learning outcomes that deviate from educators' intended objectives. Our research focuses on the nature of children's interpretations of the equal sign as a function of how they respond to classroom instruction. Our goal is not to assess the effects of a specific type of instruction on children's learning; rather, we investigated how students with different learning trajectories in response to instruction on the equal sign would differ on various measures of equivalence knowledge. In this study, we defined a learning trajectory as the evolution of students' performance on non-canonical problems at three time points (before instruction, some days after instruction, some weeks after instruction). Our work contributes to the literature on how students' learning trajectories are related to their understanding of the equal sign, which has important implications for classroom practice.

Children's Knowledge of the Equal Sign

Like conceptual knowledge of mathematics more generally, equal sign understanding involves the interconnection of different facets of knowledge (Byrnes, 1992; Hiebert & Lefevre, 1986) that develop at different paces, resulting in development that is not linear (Matthews et al., 2012; Rittle-Johnson et al., 2011). Researchers have revealed the complexity of children's development by drawing on a variety of assessments, including performance on equivalence problems presented symbolically and nonsymbolically, the ability to recall or evaluate the structure of non-canonical equations, and ratings of others' definitions of the equal sign or the quality of their own definitions (e.g., Alibali et al., 2007; Freiman & Lee, 2004; Knuth et al., 2006; Li et al., 2008; Seo & Ginsburg, 2003; Sherman & Bisanz, 2009).

By the age of four, children have little trouble establishing whether two sets of concrete objects, such as blocks, are quantitatively equivalent (Mix, 1999). Their difficulties begin in school and stem from the failure to map their conceptual knowledge of equivalence to contexts in which the equal sign is presented symbolically (Seo & Ginsburg, 2003; Sherman & Bisanz, 2009). In short, children struggle to understand the symbol itself (=). This struggle is exhibited when children make conceptual errors when solving non-canonical equivalence problems and do not define the equal sign relationally. Further in their development, children can solve some non-canonical equations correctly (i.e., $c = a + b$ or $a = a$), but they still struggle to define the equal sign relationally. Even at the point when children learn how to solve a wider variety of non-canonical equations and recognize a relational definition of the equal sign, their operational views can nevertheless co-exist with relational ones (Rittle-Johnson et al., 2011). Thus, the development of children's equal sign understanding is not straightforward, and it is impacted by a number of factors, including their prior knowledge and previous school experiences (McNeil, 2014; McNeil & Alibali, 2005).

It is thus perhaps not surprising, then, that examinations of the effects of instructional interventions also point to the complexity of children's learning. In the first place, not all students (in some cases fewer than half) respond to instruction in intended ways, despite statistically significant effects on the average (Jacobs et al., 2007; Perry, 1991; Rittle-Johnson & Alibali, 1999). Recent research provides a possible explanation for these findings by showing that students follow different learning trajectories, even with similar instructional experiences. For example, Watchorn et al. (2011) delivered a lesson to second- and fourth-grade students on the meaning of the equal sign. On average, significant instructional effects were found on all measures, namely equivalence problem solving, reconstructions of non-canonical equations, ratings of others' definitions, and ratings of non-canonical equations in terms of whether or not they "made sense." Moreover, these different assessments allowed the authors to demonstrate that students' knowledge after instruction differed qualitatively: The students fell into four distinct profiles determined by patterns of

performance. Some students fell into a clear operational profile where they performed poorly on all measures. Others fell into a “good definers” profile, characterized by operational views except for high ratings of relational definitions of the equal sign. Students falling into a third profile, “poor equation raters,” performed highly on all measures except for equation rating, and students in the final profile exhibited relational thinking on all measures. In sum, not all children respond to instruction in the same way, and their response to instruction results in a multi-faceted picture of their equal sign understanding.

Because our objective is to explore students’ knowledge in relation to how they respond to classroom instruction on the equal sign, we argue that an individual differences approach can be useful for our purposes. McNeil and Alibali (2005) examined the relation between the degree to which elementary school children adhered to operational patterns before a brief lesson on the equal sign and their ability to learn from instruction. The authors found that the more deeply entrenched children’s operational views were before the lesson, the less likely they were to generate new strategies for solving equivalence problems and to accurately solve familiar and transfer equivalence problems after instruction.

McNeil et al. (2019) found similar results in their longitudinal investigation of second- and third-graders’ understanding of equivalence. The authors found that the strategies used by children, regardless of the accuracy of their problem solving, impacted the ways in which they think about problems in later grades. Specifically, while most of the sample used incorrect strategies in the second grade, the nature of their strategies predicted performance in the third grade. That is, when second-graders relied on (incorrect) traditional arithmetical strategies to solve equivalence problems (i.e., adding all the numbers in the equation or adding the numbers on the left of the equal sign), they were less likely to solve equivalence problems correctly in third grade, compared to second-graders who used other incorrect strategies. These findings suggest that children’s equivalence knowledge is dependent on different forms of reasoning.

Similar findings were reported in a study by Byrd et al. (2015), who found that the conceptions students had of the equal sign before instruction predicted their early algebra performance after instruction. As expected, students in the third and fifth grade who held relational interpretations at the start of the school year (i.e., prior to instruction) were more successful on a test of early algebraic reasoning at the end of the year compared to those with non-relational interpretations. Furthermore, in the fifth grade, students who held arithmetic-specific non-relational views of the equal sign (e.g., “the number you get when you add”) before instruction were at a significant disadvantage even relative to those with other types of non-relational interpretations (e.g., “it means the end of the question”). In sum, the nature of students’ conceptions of the equal sign before instruction predicts their equivalence knowledge and other related constructs after instruction. What remains unanswered, however, is how students’ learning trajectories – defined by their knowledge of the equal sign as it evolves through classroom instruction – differ in terms of their equivalence knowledge after instruction.

Influence of Working Memory and Arithmetic Fluency

It has been well established that working memory and arithmetic fluency are important predictors of mathematical processing in a number of domains (Ashcraft et al., 1992; Geary et al., 2008; LeFevre et al., 2005; Raghubar et al., 2010; Tronsky & Royer, 2003), even when other cognitive and academic factors are taken into account (see Berg, 2008; Fuchs et al., 2005). Evidence also suggests that children’s working memory and arithmetic fluency should be taken into account when characterizing children’s algebraic reasoning more specifically, particularly in symbolic contexts (Sherman & Bisanz, 2009). Tolar et al. (2009), for example, argued that working memory is necessary for algebraic processing because students need to maintain multiple interpretations of expressions, suppress misconceptions (such as the operator view of the equal sign), and retrieve facts from long-term memory. The authors provided empirical evidence that both working memory and computational fluency accounted for the variance observed in symbolic manipulation, with the latter demonstrating the largest effects. Similarly, Lee et al. (2004) showed that a composite measure of central executive working memory contributed directly to performance on algebraic word problems. Furthermore, working memory plays an important role in how students respond to equivalence instruction. Fyfe et al. (2015) found that the working memory of second- and third-grade children moderated the effects of the feedback they received (either on their answers or their strategies) when solving equivalence problems.

Present Study

Previous research using an individual differences approach examined students' knowledge after instruction as a function of instruction type (Watchorn et al., 2011) or in relation to their conceptions of the equal sign before instruction (e.g., Byrd et al., 2015; McNeil & Alibali, 2005). In contrast, we explored how students on different learning paths (i.e., learning trajectories) differ in terms of their equivalence knowledge after instruction. We also investigated how children's knowledge on non-equivalence tasks (i.e., working memory and arithmetic fluency) was related to their understanding of the equal sign after instruction. Exploring the relationship between students' learning trajectories and the nature of their equivalence knowledge after instruction increases the ecological validity of the research and promises to yield pedagogically-useful implications.

The instruction was delivered in third- and fourth-grade classrooms by four teachers who participated in a larger professional development (PD) project on mathematical equivalence and relational thinking. During the PD, the teachers were exposed to the conceptual underpinnings of the equal sign, different ways to explain its meaning, and a variety of strategies for solving open-number sentences. After the PD, the teachers delivered one lesson on the meaning of the equal sign to their students.

We assessed students' problem solving before and after the lesson, which allowed us to identify student groups according to how they responded to the instruction. Within eight weeks of the post-assessment, we again measured students' equal sign understanding using a variety of measures, administered both in class and in individual interviews. Learning trajectories emerged at this point on the bases of their problem-solving performance assessed before instruction, at post-assessment, and on their problem solving within the eight-week period after the post-assessment. We then compared different aspects of the students' equivalence knowledge as a function of their learning trajectories. In particular, we explored how students with different learning trajectories differed in terms of their equivalence knowledge other than problem solving, namely generating their own definitions for the equal sign, evaluating the definitions of others, justifying their own strategies for solving equivalence problems, and solving equivalence problems in nonsymbolic contexts. We also examined the performance of the children in the learning trajectories on non-equivalence tasks.

Method

Participants

The study was part of a larger project (Bisanz et al., 2014) in which elementary school teachers participated in PD workshops to learn about children's thinking about mathematical equivalence and to create usable classroom activities for teaching equivalence. The present study is focused on students from four classrooms who received a combination of conceptual and procedural instruction for solving equivalence problems. The sample consisted of 26 students in third grade (13 girls, 13 boys) and 30 students in fourth grade (10 girls, 20 boys), all from three suburban elementary schools near Montréal, Canada. Participants were students of the teachers who had volunteered to be part of the PD workshops. Our sample consisted of 11 to 19 students per class (11 and 15 students for the two third-grade classes, and 11 and 19 students for the two fourth-grade classes). The participants had an average age (in years;months) of 9;10 (9;2 for third-grade students and 10;2 for fourth-grade students). We chose to focus our study on the students in the third and fourth grade because Watchorn's (2011) intervention was found to be particularly effective for students in this age range.

The province of Québec publishes an income index and a socio-economic index for all public schools in the province (Ministère de l'Éducation, du Loisir et du Sport, 2011). The income index is the proportion of children from families at or below the poverty line. The socio-economic index is a weighted proportion of (a) the number of children coming from families with mothers with no post-secondary education (two-thirds of the index) and those coming from families in which both parents were unemployed at the time of the last Canadian census (one-third of the index). Based on these indices, the schools are ranked on a scale of 1 (*least disadvantaged*) to 10 (*most disadvantaged*). The indices reported here were published by the *Gouvernement du Québec* during the year of the data collection for the present study. The income

index for School 1 was 9.61 (rank: 4) and the socio-economic index was 4.29 (rank: 2); for School 2, the indices were 17.38 (rank: 7) and 5.58 (rank: 2) on the income and socio-economic scales, respectively; and for School 3, the indices were 10.57 (rank: 4) and 4.46 (rank: 2) on the income and socio-economic scales, respectively.

In conducting this investigation, we complied with the American Psychological Association and local standards (i.e., principles of Canada's Tri-Council Policy) related to the ethical treatment of the participants. Parental permissions were obtained, and each participant provided assent. Additionally, full ethical approval from the university, the school board, and the schools' governing boards was granted.

Design and Procedure

The sequence of assessment activities and classroom instruction is presented in Table 1.

Table 1

Timeline of Assessment Activities and Classroom Instruction on the Equal Sign

Phase	Time of Year	Activity
1	January – February	In-class assessment: Equivalence Problem Solving test (Pre-instruction)
2	March – April	Classroom instruction on the meaning of the equal sign
3	April – May	In-class assessment: Equivalence Problem Solving test (Post-instruction)
4	May	In-class assessment: Fluency, Evaluating Definitions task
5	June	Interview 1: TONI-3, Numbers Reversed
6	June	Interview 2: Symbolic task, Generating Definitions task, Nonsymbolic task

Before and after instruction (Phases 1 and 3), the participants completed a test of equivalence problem solving using the Equivalence Problem Solving test (Watchorn & Bisanz, 2005), a performance measure designed to assess students' ability to solve a variety of equivalence problems. Within each classroom, about half of the students completed one version of this test before instruction and a second version after instruction, and the other half completed the versions in reverse order. Students worked independently on the test during one of their regularly scheduled mathematics classes and were given approximately 30 minutes to complete it. The pre-instruction assessment took place at the end of January or the beginning of February.

Between eight and 13 weeks after the pre-instruction equivalence problem-solving measure, the teachers delivered a single instructional lesson to their students (Phase 2). The lesson adhered to instructional principles validated in previous research on mathematical equivalence (McNeil, 2014). Specifically, the teachers presented non-canonical equations with numbers and operations on both sides of the equal sign (Bisanz et al., 2014; McNeil et al., 2011) and used language that included relational terms such as “is the same as” during the lesson (Chesney & McNeil, 2014). Teachers were introduced to the content and format of the lesson during the second PD workshop, during which time they viewed a video demonstration and practiced delivering the lesson through role-playing activities.

In their respective classrooms, the teachers began the lesson by presenting an equivalence problem (e.g., $3 + 1 + 1 = 3 + \underline{\quad}$) and saying, “the goal of a problem like this is to find a number that fits in the blank so that when you put together the numbers on the left side of the equal sign, you'll have the same amount as when you put together the numbers on the right side of the equal sign” (see Perry, 1991). Each teacher then solved either three (three teachers) or four (one teacher) equivalence problems in front of her students while emphasizing the meaning of the equal sign in each case – that both sides need to represent the same amount. All of the problems involved addition and subtraction,

and the strategy demonstrated was to add the numbers on the left side of the equal sign and then subtract the sum of the numbers on the right side. The final part of the lesson involved a comparison problem where students discussed whether an equal sign would be appropriate between two expressions (e.g., $3 + 4 + 6 \underline{\quad} 2 + 6$); the use of other symbols (i.e., not-equal “ \neq ,” greater than “ $>$,” less than “ $<$ ”) was not addressed in the lesson plan we provided to the teachers.

When presenting all number sentences, the teachers followed the lesson plan’s instruction to use one color for the numbers and symbols on the left side of the equal sign and a different color for the right side of the equal sign. Three teachers used a third color to represent the equal sign. The fourth teacher used the same color for the equal sign as the color used for the left side of the equation. The lesson plan also required teachers to use sweeping hand gestures over the left and right sides of the equations. After the demonstrations, the teachers put additional problems on the board (two teachers presented two problems and the other teachers each presented three) and invited students to share their own strategies in a whole-class discussion. They then gave their students practice problems to work on in small groups or on their own. Three teachers assigned 15 practice problems and one teacher provided 8. The lengths of the lesson delivered by each teacher (excluding individual and small group practice) were 13.4 min, 22.8 min,¹ 10.1 min, and 10.8 min ($M = 14.0$, $SD = 5.4$).

Eleven to 13 days after instruction, we administered the isomorphic version of the Equivalence Problem Solving test to the students in their classrooms (Phase 3) using the same procedures as before instruction. The post-instruction assessment took place in early April for one classroom and in early May for the other three classrooms. We compared students’ performance on the equivalence problem-solving measure before and after instruction to identify three initial learning trajectories: (1) students who performed poorly both before and after instruction; (2) those whose performance increased after instruction; and (3) those who performed well both before and after instruction.

Students’ knowledge and performance were subsequently assessed on three occasions. We first returned to the students’ classrooms between three and seven weeks after the post-instruction equivalence problem-solving measure (Phase 4). In this session we administered (a) a test of arithmetic fluency, and (b) a test in which students evaluated different definitions of the equal sign (Evaluating Definitions task). Second, within one week of the classroom visit, a member of the research team met individually with the students to administer measures of nonverbal intelligence (TONI-3) and working memory (Numbers Reversed; Phase 5). Third, no more than one week later, students met individually with an interviewer who assessed additional aspects of their equivalence knowledge (Phase 6): performance on an abbreviated version of the equivalence problem-solving measure (the Symbolic task), which served as a retention measure; justifications for their solutions on the Symbolic task; the quality of the definitions they provided for the equal sign (Generating Definitions task); and knowledge of arithmetical equivalence (the Nonsymbolic task). Students’ performance on the retention measure during the second interview was used to further differentiate the learning trajectories. All individual student meetings were videotaped, and the researcher was blind to the student’s group membership. The in-class assessments took place at the end of May, and the individual interviews took place in early June. For logistical reasons related to the larger project, we were not able to reduce the amount of time between the post-instruction classroom measures and the individual interviews.

Measures and Coding

Instructional Fidelity

The teachers delivered the equivalence lesson to their students in one mathematics class period. Each lesson was video recorded to enable assessment of instructional fidelity. We created a checklist that contained 13 essential lesson components (see [Appendix](#)). The first author and a trained research assistant watched the videos of each lesson independently and used the checklist to identify which components were present during each lesson. Both raters agreed on 100% of the checklist items for each lesson. The mean percentage of checklist items addressed by the teachers was 96.2% ($SD = 4.4\%$).

1) This teacher’s teaching style was more conversational than the other three teachers, which explains why the length of her lesson was relatively longer. Instructional fidelity data showed that the content covered in each of the teachers’ lessons was similar.

Non-Equivalence Measures

We administered three tasks to measure students' non-equivalence skills: one measure of arithmetic fluency (Watchorn & Bisanz, 2005), the Test of Nonverbal Intelligence 3rd Edition (TONI-3; Brown et al., 1997), and the Woodcock-Johnson-III Numbers Reversed (Woodcock et al., 2001).

Fluency Measure — The fluency measure (Watchorn & Bisanz, 2005) assessed children's speed and accuracy in solving addition and subtraction problems. The test was a paper-and-pencil measure consisting of 39 single-digit addition and subtraction problems (e.g., $4 + 5 = \underline{\quad}$). All problems were presented horizontally in canonical form, with 31 two-term problems and 8 three-term problems. Of the three-term items, 6 involved both addition and subtraction (e.g., $6 + 4 - 3 = \underline{\quad}$) and 2 involved only addition (e.g., $4 + 5 + 6 = \underline{\quad}$). Students were asked to answer as many as they could in a given amount of time by writing their answers on blank lines provided in the equations. Third graders were given 105 seconds and fourth graders were given 90 seconds to complete the test. A student's score was the number of correct responses per minute. The minimum score possible was 0, and the maximum score possible was 22.3 for third graders and 26.0 for fourth graders.

TONI-3 — The TONI-3 (Brown et al., 1997) is a nonverbal measure of cognitive ability that requires students to solve abstract figural problems. Brown et al. reported high degrees of internal consistency and test-retest reliability. In terms of construct validity, Banks and Franzen (2010) reported a strong positive correlation with the Matrix Reasoning subtest of the Intelligence Scale for Children – Fourth Edition's Full Scale IQ.

Instructions were given to the participant nonverbally, with gestures and facial expressions (e.g., pointing to test items; looking questioningly at the participant; shaking head, "no," nodding head, "yes"; see the instruction manual in Brown et al., 1997, for more information on administration pantomimes). Five practice items were administered prior to the test. The student was presented with items printed in a picture book. There were 45 items in total, and each one was printed on a single page. Items were arranged from easiest to most difficult. Each item required the student to choose, from several response choices, the picture that best fit in the empty box of a stimulus pattern. Students were asked to respond by pointing to the answer they believed was the correct one. After the practice items, test administration began with the first item and ended when the student made three incorrect responses in five consecutive items ("ceiling") or when all 45 items were administered.

Each student's raw score was calculated by adding the number of correct responses between Item 1 and the ceiling item. The TONI-3 manual contains tables that were used to translate the raw scores into standardized scores, which were used in the analyses.

Numbers Reversed — The third non-equivalence measure was the WJ-III Numbers Reversed (Woodcock et al., 2001), a measure of working memory (LeFevre et al., 2005). The examiner read a string of random digits between 0 and 9 aloud, and students were asked to repeat the numbers in reverse order. The strings of digits became increasingly longer as the test progressed. The administration of this test ended once a child made three or more errors in a block of items or once all items had been administered.

Students received one point for each string of numbers they correctly repeated backward. The points earned were added to obtain a total score, which could range from 0 to 30. Woodcock et al. (2001) reported that the reliability of the WJ-III Numbers Reversed for 8-year-olds is .86 (as cited in Seethaler & Fuchs, 2006).

Equivalence Assessments

Two equivalence assessments were conducted as a whole group with the students in their classrooms (the Equivalence Problem Solving test and the Evaluating Definitions task) and three were administered to students in one-on-one interview settings (i.e., the Symbolic task, the Generating Definitions task, and the Nonsymbolic task).

Equivalence Problem Solving Test — This test consisted of two- and three-term single-digit addition and subtraction problems. Students were asked to solve as many as they could by writing their answer on a blank line in the equation. The first five items were canonical addition and subtraction practice problems (e.g., $3 + 4 = \underline{\quad}$). These were followed by

4 sets of 5 equivalence problems (e.g., $6 + 7 = __ + 5$), with 3 canonical problems interspersed between the sets and one as the last item on the test. This made for a total of 9 canonical problems and 20 non-canonical problems.

Five types of equivalence problems were used: (a) identity ($a + b = a + __$), (b) commutativity ($a + b = b + __$), (c) two-term part-whole ($a + b = c + __$), (d) three-term part-whole ($a + b + c = d + __$), and (e) combination ($a + b + c = a + __$; Sherman & Bisanz, 2009). The blank immediately followed the equal sign in half of the equivalence problems and was at the end of the equation in the other half.

Accuracy (percent correct) was calculated separately for canonical and non-canonical problems. For the canonical problems, Cronbach's alpha reliability estimate for the sample at pretest was .67 and .71 at posttest. These low estimates are likely due to the small number of canonical items and the near-ceiling performance on these items (Pallant, 2013). For the non-canonical problems, the Cronbach's alpha reliability estimates for the sample were .95 at pretest and .96 at posttest.

Symbolic Task — As a retention measure, students were asked to solve five equivalence problems similar to those on the Equivalence Problem Solving test. Each problem was presented on an individual index card, and students wrote their answers directly on the card. Accuracy (percent correct) was calculated to obtain an Equation Solving score. Reliability for the Equation Solving subscale of the Symbolic task was high, Cronbach's $\alpha = .98$.

After each item, the researcher asked the students to justify their answers. Each justification was coded as either relational or non-relational. Justifications coded as relational were either descriptions of the equal sign as relational (e.g., for $3 + 4 = 4 + __$, “the left side has to equal the same as the right side. The left side is 7, and if I put a 3 on the line, the right side is going to equal 7, too.”) or descriptions of a procedure that was consistent with a relational view (e.g., “I added 3 and 4 and got 7, then I subtracted 4 from it and put 3 on the line.”). Justifications that were coded as non-relational included descriptions that were operational (e.g., for $6 + 4 + 5 = 6 + __$, “you always put the answer after the equal sign and so the answer is 15”), that involved a procedure that aligned with an operational view (e.g., “The answer is 15 because 6 plus 4 plus 5 is 15”), or that featured other non-relational views (e.g., “I just put any number on the line.”).

The student was awarded 1 point for each relational justification and 0 points for justifications coded as non-relational. The points were summed across all 5 items to yield a Justification score ranging from 0 to 5. Reliability for the Justification subscale was high, Cronbach's $\alpha = .96$. The first author coded all the justifications. A trained rater was asked to code 10% of the justifications, and an inter-rater agreement of 96% was obtained.

Generating Definitions Task — The researcher asked the students to generate their own definition of the equal sign. The child was presented with an index card on which the equation $3 + 4 = 2 + 5$ was written. The researcher then pointed to the equal sign in the equation and asked, “What is the name of this symbol? Can you explain to me what this symbol means?”

Students' definitions were coded as Relational, Operational, or Mixed. The interviewers were instructed to prompt students with follow-up questions whenever they provided answers that were difficult to classify (e.g., “It means equal.”). Prompting stopped when the interviewer felt she had determined the students' interpretations of the equal sign. The definitions were coded as Relational when students explained that both sides of the equal sign needed to represent the same amount. Definitions were coded as Operational when descriptions contained a misconception, such as “add all the numbers” (e.g., answering 17 to solve the problem $8 + 4 = __ + 5$) or “the answer comes next” (e.g., answering 12 to solve the problem $8 + 4 = __ + 5$). In some cases, students provided both relational and operational aspects in their definitions, such as “The equal sign always means different things. Sometimes it means ‘the same as’ [relational], and sometimes it means you need to say the answer of the problem [operational].” We decided to place such definitions in a separate category, Mixed, because of how such responses could be placed in the construct map presented by Rittle-Johnson et al. (2011). In the construct map, children at the Basic Relational level demonstrate emergent relational thinking because they can recognize and generate relational definitions of the equal sign, but Matthews et al. (2012) argued that they cannot be placed at the highest level (Comparative Relational) because their relational views coexist with operational ones. The first author coded all the definitions. A trained rater was asked to code 10% of the generated definitions, and an inter-rater agreement of 100% was obtained.

Evaluating Definitions Task — In this task (McNeil & Alibali, 2005), students were presented with the equation $3 + 4 + 1 = 3 + \underline{\quad}$ at the top of the test paper, with an arrow pointing to the equal sign. Students were asked to rate six different definitions of the equal sign by circling one of three faces (unhappy, neutral, happy) corresponding to their evaluation of the definition (“not so smart,” “kind of smart,” “very smart,” respectively). A member of the research team read the instructions aloud to the students, who were in their regular classrooms: “At the top of the sheet, you see a number sentence. There is an arrow pointing to a math symbol. Do you see it? I asked 6 students from another school what that math symbol meant. I’m going to tell you what each one of them said, and I want you to circle whether you think it is very smart, kind of smart, or not so smart.” Each definition was read out loud for the students, after which they were given time to rate it.

Two definitions reflected a relational understanding of the equal sign (e.g., “both sides of the equal sign should have the same amount”), two reflected an operational understanding of the equal sign (e.g., “the answer goes next”), and two were not related to equivalence (e.g., “all the numbers after it are small”). Using the same scoring procedure as Rittle-Johnson and Alibali (1999), we awarded students 2 points for each “very smart” rating, 1 point for each “kind of smart” rating, and 0 points for each “not so smart” rating. The Evaluating Definitions score was computed by subtracting the mean rating for the two operational definitions from the mean rating for the two relational definitions to get a difference score that could range from -2 (when students rated the operational definitions as “smarter” than the relational definitions) to 2 (when students rated the relational definitions as “smarter” than the operational definitions).

Nonsymbolic Task — The Nonsymbolic task (Sherman & Bisanz, 2009) served to assess students’ knowledge of arithmetical equivalence. Students were asked to solve five equivalence problems presented with wooden blocks placed on white index cards. The blocks were used to represent the amounts on either side of the equal sign, which was represented by a piece of cardboard that was folded like a tent. There were two (or three, depending on the item) index cards on the left side of the cardboard tent, and two index cards on the right of it. To make the two sides of the cardboard tent distinct, the index cards on the left side were placed on red construction paper, and the index cards on the right side were placed on green construction paper.

The child was instructed to put blocks on the empty card “so that when you put together these on this side of the blue tent (interviewer gestured to the blocks on the left), you’ll have the same number as when you put together these on this side of the blue tent” (interviewer gestured to the blocks and the empty card on the right; Sherman & Bisanz, 2009, p. 91). Testing began when the interviewer was sure the student understood the task, which was after he or she successfully completed (i.e., put the correct number of blocks on the blank index card) two to five practice items with canonical equations (e.g., $8 + 4 = \underline{\quad}$). Accuracy (percent correct) was calculated to obtain students’ scores. Cronbach’s alpha reliability estimate for the items on the Nonsymbolic task was .77.

Results

Performance on Equivalence Problem Solving

Means and standard deviations of equivalence problem-solving scores at pre-instruction and post-instruction by grade and by equation type (canonical, non-canonical) are presented in Table 2.

Performance was analyzed with a 2(Grade: 3, 4) by 2(Time: pre-instruction, post-instruction) by 2(Equation Type: canonical, non-canonical) ANOVA with repeated measures on the last two variables. Mean performance was higher after instruction compared to before instruction, $F(1, 54) = 65.82$, $p < .001$, $\eta_p^2 = .55$, and performance on canonical problems was higher than performance on non-canonical problems, $F(1, 54) = 107.60$, $p < .001$, $\eta_p^2 = .67$. In fact, performance was near ceiling on the canonical problems, suggesting that errors on non-canonical problems cannot be attributed to computational difficulties. There was also an interaction of time and equation type, $F(1, 54) = 50.81$, $p < .001$, $\eta_p^2 = .49$, with simple effects analyses indicating an increase in performance on non-canonical problems ($p < .001$) but not on canonical problems. As evident in Table 2, mean accuracy on non-canonical problems more than doubled from pretest to posttest for children in both grades.

Table 2

Means (Percent Correct) and Standard Deviations of Equivalence Problem Solving Test Scores Pre-instruction and Post-instruction by Grade and Equation Type

Measures	Grade 3 (<i>n</i> = 26)		Grade 4 (<i>n</i> = 30)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre-instruction				
Canonical Problems	81.94	21.36	90.00	12.24
Non-Canonical Problems	28.98	34.35	25.12	41.43
Post-instruction				
Canonical Problems	87.61	17.86	92.22	14.71
Non-Canonical Problems	69.89	39.50	71.17	38.22

Only non-canonical items were included in subsequent analyses. The length of time between the pretest and the posttest was not correlated with children's difference scores on the non-canonical items ($r = -.07$, $p = .62$). Because no effects of grade were found and because grade was not correlated with any of the measures (all $ps > .05$), the levels of grade were collapsed in the analyses reported below.

Learning Trajectories

Students' learning trajectories were identified using their performance on the Equivalence Problem Solving test at pre- and post-instruction and refined using their performance on the Equation Solving component of the Symbolic task (i.e., the retention measure) administered during the interview at Phase 6 (see Figure 1). To classify a student as having some ability in solving equivalence problems, we chose a criterion of 60% accuracy on the 20 non-canonical items of the Equivalence Problem Solving test. There were four identity problems on the test and four commutative problems, which have been shown to be easier than two-term part-whole, three-term part-whole, and combination problems (Sherman & Bisanz, 2009). If a child answered all identity and commutative problems correctly, he or she would have had to have been successful on at least one-third of the more difficult problems (an additional 4 problems out of 12) to reach the 60% criterion. Using a criterion of 75% correct yielded results nearly identical to those reported below.

Using the 60% criterion, three initial learning trajectories were formed: Nonsolvers ($n = 18$) scored below 60% on the Equivalence Problem Solving test both pre- and post-instruction; Learners ($n = 24$) scored below 60% pre-instruction and 60% or higher post-instruction; Solvers ($n = 14$) scored 60% or higher on the Equivalence Problem Solving test at both time points. Thus, of the 42 students who did not perform at threshold before instruction, 18 (43%) did not reach the same threshold on the post-instruction Equivalence Problem Solving test, a rate that is not dissimilar to that reported in previous research (e.g., Jacobs et al., 2007). The learning trajectories that emerged from the application of our threshold at pre- and post-instruction time points resulted in categories that reflect classroom realities and therefore are instructionally relevant.

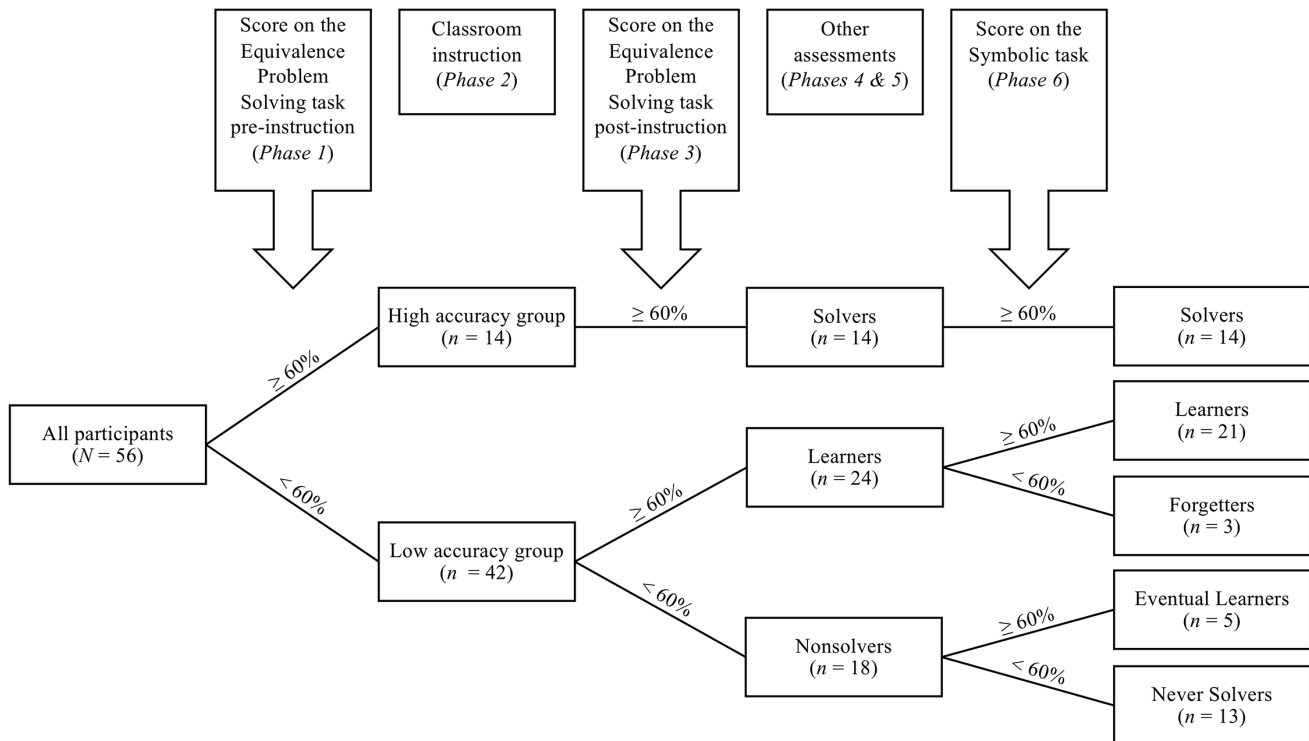
We used the same 60% threshold for the students' performance on the retention measure at Phase 6 to further refine the students' learning trajectories. Thirteen of the 18 Nonsolvers performed below the 60% cutoff on the Symbolic task, meaning their performance was comparable to that on the post-instruction Equivalence Problem Solving test. The remaining five Nonsolvers learned how to solve non-canonical problems in the time between the posttest and the individual interview with the researcher; these students were classified as "Eventual Learners." Of the 24 Learners, three students did not maintain their performance at 60% or higher after the posttest; we called these three students "Forgetters." All 14 Solvers showed above-criterion accuracy during the interview.

Separating the Eventual Learners from the Nonsolvers group and the Forgetters from the Learners, we obtained the final learning trajectories, as shown in Figure 1: (a) Solvers ($n = 14$), (b) Learners ($n = 21$), (c) Never Solvers ($n = 13$), (d) Eventual Learners ($n = 5$), and (e) Forgetters ($n = 3$). The three main learning trajectories (i.e., Solvers, Learners, and Never Solvers) were equivalent in terms of gender distribution, $\chi^2(2, N = 48) = 0.19$, $p = .91$, with 36% girls in the Solvers,

43% girls in the Learners, and 39% girls in the Never Solvers. For all the inferential tests reported below, we removed the Eventual Learners and the Forgetters from the analyses because of the small number of students in each group. We report descriptive statistics for these two learning trajectories separately.

Figure 1

Identification of Learning Trajectories



A two-way mixed ANOVA was performed with teacher (4) as the between-subjects factor and time (3: pretest, posttest, retention measure) as the within-subjects factor, and percent correct as the dependent variable. There were main effects of teacher, $F(3, 44) = 9.09$, $p < .001$, $\eta^2 = .38$, and time, $F(2, 88) = 41.06$, $p < .001$, $\eta^2 = .48$. Post-hoc analyses using Bonferroni corrections revealed that mean scores were higher on the posttest than on the pretest and that they were also higher on the Symbolic task than on the pretest, both $ps < .001$. There were no significant differences between mean scores on the posttest and on the Symbolic task. This latter result suggests that performance on equivalence problems did not change from posttest to the Symbolic task. This might provide some evidence that the teachers' equivalence instruction remained stable after the posttest, although we acknowledge that any conclusions drawn remain tentative without a more descriptive account of their classroom practices after the posttest. Lastly, performance over time did not differ by teacher, $F(6, 88) = 0.74$, $p = .62$.

Strategy Use on Equivalence Problem Solving Test

To provide a richer picture of the students within each learning trajectory, we examined the strategies used by Never Solvers, Learners, and Solvers on the non-canonical problems on the Equivalence Problem Solving test at both pretest and posttest. Each student was placed in one of five categories at pretest based on the strategies they used to solve the problems. Students who used correct strategies (i.e., leading to the correct answer) on 80% of the items were placed in the Correct category. The remaining students used incorrect strategies on more than 20% of the items. These students were placed in the Add All the Numbers category if they used this strategy on at least 80% of the items and in the Answer Comes Next category if at least 80% of their strategies were of this type. If 80% or more of strategies were not

discernable from the answers, the students were placed in the Other category. Finally, students whose strategies did not fall predominantly in any one of the above categories were placed in the Mixed category. We observed that, at pretest, Never Solvers were the only ones to be classified as Add All the Numbers, and half of the Never Solvers (6/13) were placed in either the Add All the Numbers or Answer Comes Next categories. Most of the Learners (17/21) were placed in the Mixed category, however. Finally, and as would be expected, most of the Solvers (12/14) were in the Correct category and thus used appropriate strategies to solve the items on the test. Two of the Solvers (2/14) were placed in the Mixed category.

We also looked at the change of strategies on the Equivalence Problem Solving test from pretest to posttest for each student. We found that of the 17 students who used a single arithmetic-based strategy at pretest (i.e., students who were placed either in the Add All the Numbers or Answer Comes Next categories), six (35%) moved to the Correct category at posttest – i.e., used predominantly correct strategies after instruction. Furthermore, of the 27 students who used a variety of strategies at pretest (i.e., students who were placed in the Mixed category), 18 (67%) used correct strategies at posttest. Most of the students who were not placed in the Correct category at pretest changed strategy categories from pretest to posttest; only two students who used a single arithmetic-based strategy remained in the same strategy category at both time points.

Characterizing the Learning Trajectories

In this section, we report differences between the learning trajectories on the non-equivalence tasks and equivalence knowledge measures. Means and standard deviations of the scores on the non-equivalence tasks and the equivalence knowledge measures across grades as a function of learning trajectory can be found in Table 3.

Table 3

Means and Standard Deviations of the Scores on All Measures Across Grades as a Function of Learning Trajectory

Measures	Solvers			Learners			Never Solvers			Eventual Learners			Forgetters		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Non-equivalence measures															
Fluency ^a	9.3	4.8	14	7.4	4.1	20	2.8	1.1	13	5.6	3.0	5	3.4	1.0	3
TONI-3	103.7	6.8	13	105.4	10.2	20	91.5	12.3	12	93.2	6.1	5	112.7	6.8	3
Numbers Reversed ^b	12.0	2.9	14	11.9	2.2	21	7.6	4.6	13	8.5	3.1	4	9.7	1.5	3
Equivalence knowledge															
Equation Solving ^c	97.1	7.3	14	100.0	0.0	21	3.1	11.1	13	88.0	17.9	5	0.0	0.0	3
Justification ^d	4.9	0.4	14	4.8	0.5	21	0.1	0.3	13	4.0	1.2	5	0.0	0.0	3
Evaluating Definitions ^e	0.1	0.8	14	-0.5	0.9	21	-0.5	0.8	12	0.3	0.8	5	0.0	0.9	3
Nonsymbolic ^c	100.0	0.0	14	93.3	14.6	21	70.9	35.1	11	92.0	17.9	5	93.3	11.5	3

Note. The *ns* for each task are different because of missing data, administration errors, or withdrawal of child assent.

^amin: 0, max: 26. ^bmin: 0, max: 30. ^cReported in percent. ^dmin: 0, max: 5. ^emin: -2, max: 2.

Non-Equivalence Measures

A one-way ANOVA with Fluency as the dependent measure revealed differences among the Never Solvers, Learners, and Solvers, $F(2, 44) = 10.29$, $p < .001$, $\eta^2 = .32$. Post-hoc tests with Bonferroni corrections indicated that the Never Solvers scored lower than both the Learners ($p = .004$) and the Solvers ($p < .001$) on the Fluency measure. The Learners and Solvers did not differ from each other. Similar patterns were found for the TONI-3 and the Numbers Reversed measures. On the TONI-3, the ANOVA revealed differences between the groups, $F(2, 42) = 7.83$, $p = .001$, $\eta^2 = .27$, and post-hoc tests with Bonferroni corrections again placed the Never Solvers lower on this measure than the Learners ($p = .001$) and the Solvers ($p = .012$), who did not differ from each other. On the Numbers Reversed measure, the groups differed, $F(2, 45) = 8.67$, $p = .001$, $\eta^2 = .28$, and the post-hoc tests revealed again that the Never Solvers scored lower on this measure than the other two groups (Learners, $p = .001$; Solvers, $p = .003$), between which no differences were found.

Equivalence Knowledge Measures

Symbolic Task: Justification — On the Justification component of the Symbolic task, Learners ($M = 4.8$, $SD = 0.5$) and Solvers ($M = 4.9$, $SD = 0.4$) obtained nearly perfect scores, whereas Never Solvers had an average score near zero ($M = 0.8$, $SD = 0.3$). Eventual Learners' mean for their justifications was 4.0 ($SD = 1.2$), and it was 0 for the Forgetters. Together with the almost perfect correlation between the two components (i.e., Equation Solving and Justification) of the Symbolic task ($r = .98$), these findings demonstrate that the justifications students provided were consistent with their responses on the Equation Solving component of the Symbolic task.

Generating Definitions Task — Students generated three types of definitions on the Generating Definitions task: operational, mixed (both relational and operational), and relational. The proportions of students in each of the five learning trajectories who provided these three definitions are presented in Table 4.

Table 4

Frequencies (and Proportions) of Definition Type by Learning Trajectory

Definition Type	Learning Trajectory				
	Never Solvers	Learners	Solvers	Eventual Learners	Forgetters
Operational	11 (84.6%)	6 (28.6%)	2 (14.3%)	2 (40.0%)	1 (33.3%)
Mixed	2 (15.4%)	11 (52.4%)	6 (42.9%)	1 (20.0%)	1 (33.3%)
Relational	0 (0.0%)	4 (19.0%)	6 (42.9%)	2 (40.0%)	1 (33.3%)
Total	13	21	14	5	3

To assess the relation between learning trajectory and definition type, the mixed and relational categories were collapsed into an “ever relational” category to represent definitions that contained at least some relational elements. A chi-square test of association was conducted between definition type (ever relational, operational) and learning trajectory (Never Solvers, Learners, and Solvers). A statistically significant association was found between definition type and learning trajectory, $\chi^2(2) = 15.84$, $p < .001$. Post-hoc analyses involved pairwise comparisons using z-tests to compare proportions with Bonferroni corrections. The proportion of students in the Never Solvers trajectory who gave an operational definition (11/13 or 85%) was significantly greater than the proportions of students in both the Learners (6/21 or 29%) and the Solvers (2/14 or 14%), $p < .05$. The proportions in the latter two groups did not differ significantly from each other. That most of the Never Solvers (85%) gave Operational definitions suggests that the students in this learning trajectory were at Level 1, Rigid Operational, of Rittle-Johnson et al.'s (2011) construct map of equivalence knowledge. More than half of the Learners (52%) gave mixed definitions, which might suggest that they were at Level 3 of the construct map. Most of the Solvers (86%) gave either mixed or relational definitions.

Evaluating Definitions Task — We conducted an ANOVA to test for group differences on the Evaluating Definitions task. Given that none of the non-equivalence measures was correlated with the Evaluating Definitions task, all $ps > .05$, we did not include any as a covariate. Differences between the Never Solvers, Learners, and Solvers groups did not reach statistical significance, $p = .09$, 95% CIs [-1.04, -0.04], [-0.93, -0.12], and [-0.39, 0.54], respectively. The means in Table 3 suggest, in fact, that while all five groups performed similarly on the Evaluating Definitions task, mean scores of all groups hovered around 0, which indicates that students rated operational and relational definitions as equally smart.

Nonsymbolic Task — An ANOVA was conducted to test for group differences on the Nonsymbolic task. Although performance on each of the TONI-3 and Numbers Reversed was correlated with the Nonsymbolic task ($r = .34$ and $r = .29$, respectively, $ps < .01$), ceiling effects on the latter measure prohibited using the two non-equivalence measures as covariates in the analysis. Results revealed significant group differences, $F(2, 43) = 7.33$, $p = .002$, $\eta^2 = .25$. Follow-up comparisons with Bonferroni corrections showed that the Learners and the Solvers outperformed the Never Solvers ($p = .01$ and $p = .002$, respectively), but did not differ from each other ($p = .99$). The mean performance of both the Even-

tual Learners and the Forgetters on the Nonsymbolic task was 92.0 ($SD = 17.9$) and 93.3 ($SD = 11.6$), respectively. Thus, despite all five groups performing above the 60% criterion on the Nonsymbolic task, the Never Solvers demonstrated relatively greater difficulty with arithmetical equivalence.

Discussion

A considerable number of students do not respond to instruction about the equal sign in intended ways, but little is known about the nature of their equivalence knowledge relative to the ways in which they respond to classroom lessons on the equal sign. We contend that characterizing students' knowledge as a function of how they respond to classroom instruction lends the ecological validity necessary for conclusions that are ultimately useful to practicing teachers. The objective of the present study, therefore, was to examine differences in the equivalence knowledge of students who responded in different ways to classroom instruction on the equal sign. Three main learning trajectories emerged from the data: those who performed poorly on a test of equation solving both before and after instruction, those who had improved, and those who performed well on the same test both before and after instruction. More tentatively, our data also revealed two additional trajectories, those who forgot what they had learned following instruction and those who at some point improved their problem-solving performance, despite not showing improvement immediately after instruction.

Students' performance at posttest was consistent with previous literature showing how children's misconceptions are resistant to change (e.g., Jacobs et al., 2007; McNeil, 2014). Even after instruction that focused on the meaning of the equal sign and included demonstrations of the procedures for solving equivalence problems, almost half of the students in our study who performed poorly at pretest (Nonsolvers) did not meet our threshold of improvement on the problem-solving measure after instruction (i.e., a score of at least 60% on the test). Our data suggest, however, that some students in this category eventually learned the meaning of the equal sign some time later, although how and why this happens is a consideration for future research.

When considering the strategies students used to solve equivalence problems at pretest, we found that, consistent with previous research, students who relied on one incorrect arithmetic-based strategy were less likely to solve the equivalence problems correctly at posttest (McNeil, 2014; McNeil & Alibali, 2005). Students who used a variety of incorrect strategies at pretest (i.e., the Learners), however, were more successful on equivalence problems at posttest. These students were perhaps less rigid in their thinking and therefore more receptive to changing their strategies to correct ones following the instruction they received from their teachers. Within-child variability in children's strategies has been associated with learning and conceptual change, which could explain these findings (Alibali, 1999; Siegler, 2007). Although many students used incorrect strategies at both pretest and posttest, most students changed strategy categories from pretest to posttest; there was a shift in their thinking, which is something that has been observed in other research (see McNeil et al., 2019).

We also found that students who failed to solve equivalence problems at both time points after instruction (i.e., the Never Solvers) had little in common with those whose problem solving improved at some point after instruction began, at least on the equivalence measures used in this study. Specifically, relative to students with different learning trajectories, the Never Solvers still struggled to solve equivalence problems, had relatively weak arithmetical (i.e., nonsymbolic) equivalence knowledge, and provided predominantly operational definitions of the equal sign immediately and several weeks after having received instruction. We observed a contrasting pattern for students in the other two primary learning trajectories, regardless of whether they knew how to solve equivalence problems before instruction (Solvers) or whether they showed improved performance afterward (Learners): Most retained their ability to solve problems, defined the equal sign relationally, and had almost no deficiencies in their arithmetical equivalence knowledge.

Despite students in the Never Solvers trajectory exhibiting such differences relative to students in the other four trajectories, more consistent findings were observed on the Nonsymbolic task, which revealed performance above 60% in all groups. Nevertheless, the Never Solvers still demonstrated relatively greater difficulty than their peers in the other four trajectories on the Nonsymbolic task. Additional research that focuses on the reasons that this group of students had greater difficulty than others is necessary for teachers to know how to respond appropriately during instruction.

Showing similar consistency across trajectories, but in the opposite direction, all groups appeared to struggle when asked to evaluate others' definitions of the equal sign. Rittle-Johnson et al. (2011) argued that recognizing relational definitions as the most appropriate for the equal sign sits at the most cognitively sophisticated level of their construct map of equivalence knowledge, and may have been more challenging than the other tasks in our battery of equivalence measures. As such, given its challenging nature, it is possible that the evaluation task we administered was not sensitive enough to differentiate the learning trajectories, but further investigation on the relation between instruction and students' evaluations is warranted.

In contrast to some previous studies (e.g., Cook et al., 2008; McNeil & Alibali, 2000), we found high retention rates for those students who learned from instruction. Except for the three "Forgetters," students did not revert to their initial incorrect strategies for solving equivalence problems weeks later. Nevertheless, our observations of the Forgetters imply that there may be benefits to immediate learning from instruction, even if their learning is forgotten several weeks later. Despite not maintaining their performance on equivalence problems on the retention measure, these students were nevertheless able to generate relational definitions of the equal sign and demonstrate understanding of arithmetical equivalence. This stands in contrast to the students who were not successful at solving equivalence problems at any point (the Never Solvers): Their understanding of arithmetical equivalence and ability to define the equal sign relationally was substantially lower than the Forgetters. More research with larger samples is clearly needed to confirm this pattern.

It is more difficult to explain the performance pattern for the Eventual Learners, the students who did not learn immediately after instruction but who performed above our threshold on problem solving a few weeks later. However these students eventually learned how to solve equivalence problems, their learning was accompanied by the ability to define the equal sign relationally and an understanding of arithmetical equivalence (i.e., high performance on nonsymbolic problems). One explanation for this finding is that once students learn how to solve equivalence problems, whether it is immediately after instruction or not, and whether it is retained over time, they are "primed" (e.g., Leech et al., 2008) to acquire other types of equivalence knowledge, such as generating relational definitions and demonstrating proficiency with arithmetical equivalence. An alternate explanation could be that Eventual Learners' performance on the Symbolic task was an artefact of the types of problems presented on the pretest and posttest. The canonical problems on the pretest and posttest may have suppressed their performance on the non-canonical problems on these same measures (McNeil, 2008), thus artificially inflating their relative performance on the Symbolic task.

Contributions

Our results contribute to the literature by examining the nature of children's equivalence knowledge, and their performance on non-equivalence tasks, in response to instruction on the equal sign. We found three primary trajectories – those who performed poorly before and after instruction, those who improved, and those who performed at a high level before and after instruction. The current research contributes to the literature by characterizing the nature of the equivalence and non-equivalence knowledge of the students in these three trajectories. Students who failed to perform well on equivalence problems after instruction and still struggled several weeks later showed generally weaker equivalence knowledge relative to those who showed improvement immediately after instruction, regardless of their problem-solving performance several weeks later. Another contribution of our study is that we provide suggestive evidence that there may be, in fact, two additional learning paths, those who forget and those who eventually learn. Even those students who forget what they have learned have stronger equivalence knowledge than those students who fail on equivalence problems at all time points. The robustness of these additional two trajectories, however, should be tested in future research.

The results of the present study contribute to existing literature on the relation between mathematics instruction and student learning. Theoretical and anecdotal accounts of teaching and learning in mathematics (Steffe & Thompson, 2000; Ulrich et al., 2014), and some previous research in mathematical equivalence specifically (i.e., Watchorn, 2011), identified a need to characterize students' knowledge in response to classroom lessons on the equal sign. The nature of their knowledge could generate future hypotheses about the relationship between instruction on equivalence and children's learning. For instance, the finding that students who learned how to solve equivalence problems immediately

after instruction differed from those who did not on measures of general ability, fluency, and working memory allow us to suggest that non-equivalence measures tap some of the post-instruction differences and are potentially important for delimiting the effects of equal sign instruction. These results suggest that stronger working memory and greater fluency in mathematics are likely to mediate the ways in which children learn from instruction. Indeed, given the large body of literature examining the influences of working memory on arithmetic and problem solving (Adams & Hitch, 1997; Imbo & Vandierendonck, 2007; McKenzie et al., 2003; Raghubar et al., 2010; Rasmussen & Bisanz, 2005) and arithmetic fluency on strategy development and problem-solving speed (Bull & Johnston, 1997; Carr & Alexeev, 2011; Royer et al., 1999), this mediation hypothesis is plausible and should be tested in future research.

Limitations

Certain limitations of our work should be noted. First, the number of students in each learning trajectory was small in some cases, preventing us from including other variables, such as gender, in our analyses. Our sample was also too small to help us arrive at reliable conclusions about the two smallest learning trajectories we identified (i.e., Eventual Learners and Forgetters).

Additionally, we were not able to document the instructional activities the teachers implemented in their classrooms other than the lesson that we had asked them to deliver. It is possible, for example, that some of them may have brought out the algebraic character of arithmetic in a number of other lessons (Schliemann et al., 2003) that we were not able to observe. Documenting the nature of teachers' classroom practice outside the scope of our professional development may have provided additional insight into children's responses to equivalence instruction. Relatedly, although we used a checklist to ascertain whether teachers included all key components in their equivalence lesson, we did not evaluate the quality of teachers' instruction, which may also have provided nuance to our interpretation of the data. An important direction for future research, therefore, would be to examine the interactive nature of instruction and children's developmental trajectories on their equivalence knowledge.

Conclusion

The results of the present study are informative for teachers. Students who demonstrate persistent difficulties with the equal sign are likely the ones who struggle with many, if not most, aspects of equivalence knowledge. In the context of the classroom, then, students' problem-solving performance may be a useful index for their overall equivalence knowledge. A key pedagogical implication is that students who have difficulty responding to lessons on the meaning of the equal sign may benefit from additional targeted instruction on equivalence in both symbolic and nonsymbolic contexts.

Furthermore, the added observation that students' justifications on the retention measure revealed views of the equal sign that were consistent with their responses serves to support the validity of using problem-solving performance as an indicator of overall knowledge. The conclusion that problem solving is particularly revealing of students' knowledge may only hold given the specific type of instruction delivered in the present study, however, which was focused on explicit explanations of the meaning of the equal sign and clear demonstrations on how to solve equivalence problems. Regardless, given that instruction served to constrain students' learning about equivalence suggests that frequently assessing the aspects of equivalence that are highlighted during instruction would help to identify those students with persistent difficulties.

Finally, our research provides indirect evidence that the instruction designed by Watchorn (2011) can be successfully adapted for classroom use. Although the design of the current study does not allow us to draw causal conclusions, our data imply that one classroom lesson on the meaning of the equal sign and on strategies for solving equivalence problems had a positive effect on more than half of the students in the study, who retained their performance several weeks later. Additionally, in terms of their equivalence knowledge after instruction, these students looked similar to those who performed well on equivalence problem solving before instruction. This finding is consistent with previous research showing that a single lesson can result in high levels of equivalence knowledge for some students (e.g., Cook et al., 2008; McNeil & Alibali, 2000; Sherman & Bisanz, 2009). In terms of remediation for those students who do not

respond to instruction, it may be that teaching strategies that take into account cognitive deficits, such as working memory capacity and arithmetic fluency, may be worth examining in future research.

Funding: The current research was supported by the Social Sciences and Humanities Research Council of Canada Grant 410-2009-0880.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Author Note: Rebecca Watchorn Kong is now at Research & Experimentation, Financial Consumer Agency of Canada, Ottawa, ON, Canada. Jody Sherman LeVos is now at HOMER by BEGiN, New York, NY, United States.

Data Availability: The original data for this study cannot be made publicly available because permissions for access were not obtained from the participants' parents at the time of data collection.

References

- Adams, J. W., & Hitch, G. J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, 67(1), 21-38. <https://doi.org/10.1006/jecp.1997.2397>
- Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, 35(1), 127-145. <https://doi.org/10.1037/0012-1649.35.1.127>
- Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning*, 9(3), 221-247. <https://doi.org/10.1080/10986060701360902>
- Ashcraft, M. H., Donley, R. D., Halas, M. A., & Vakali, M. (1992). Working memory, automaticity, and problem difficulty. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 301-329). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)60890-0](https://doi.org/10.1016/S0166-4115(08)60890-0)
- Banks, S. H., & Franzen, M. D. (2010). Concurrent validity of the TONI-3. *Journal of Psychoeducational Assessment*, 28(1), 70-79. <https://doi.org/10.1177/0734282909336935>
- Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, 99(4), 288-308. <https://doi.org/10.1016/j.jecp.2007.12.002>
- Bisanz, J., LeVos, J. S., Osana, H. P., Kong, R. W., & Piatt, C. (2014, May 8-9). Improving children's understanding of the equal sign: Lessons learned. In J. Bisanz (Chair), *Improving children's understanding of the equal sign* [Symposium]. Development 2014: A Canadian Conference on Developmental Psychology, Ottawa, Canada.
- Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J.-S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education*, 46(1), 39-87. <https://doi.org/10.5951/jresmetheduc.46.1.0039>
- Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence—Third Edition*. PRO-ED.
- Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, 65(1), 1-24. <https://doi.org/10.1006/jecp.1996.2358>
- 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. <https://doi.org/10.1016/j.lindif.2015.01.001>
- Byrnes, J. P. (1992). The conceptual basis of procedural learning. *Cognitive Development*, 7(2), 235-257. [https://doi.org/10.1016/0885-2014\(92\)90013-H](https://doi.org/10.1016/0885-2014(92)90013-H)
- Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school*. Heinemann.
- Carpenter, T. P., Levi, L., Franke, M. L., & Zeringue, J. K. (2005). Algebra in elementary school: Developing relational thinking. *ZDM. Zentralblatt für Didaktik der Mathematik*, 37(1), 53-59. <https://doi.org/10.1007/BF02655897>
- Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology*, 103(3), 617-631. <https://doi.org/10.1037/a0023864>

- Chesney, D. L., & McNeil, N. M. (2014). Activation of operational thinking during arithmetic practice hinders learning and transfer. *The Journal of Problem Solving*, 7(1), 24-35. <https://doi.org/10.7771/1932-6246.1165>
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, 106(2), 1047-1058. <https://doi.org/10.1016/j.cognition.2007.04.010>
- Empson, S. B., Levi, L., & Carpenter, T. P. (2011). The algebraic nature of fractions: Developing relational thinking in elementary school. In J. Cai & E. Knuth (Eds.), *Early algebraization: A global dialogue from multiple perspectives* (pp. 409-428). Springer Berlin Heidelberg.
- Freiman, V., & Lee, L. (2004). Tracking primary students' understanding of the equal sign. In M. Johnsen & A. Berit (Eds.), *Proceedings of the 28th international group for the psychology of mathematics education* (Vol. 2, pp. 415-422). Bergen University College.
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, 97(3), 493-513. <https://doi.org/10.1037/0022-0663.97.3.493>
- Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2015). When feedback is cognitively-demanding: The importance of working memory capacity. *Instructional Science*, 43, 73-91. <https://doi.org/10.1007/s11251-014-9323-8>
- Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., & Graban, J. (2008). Report of the task group on learning processes. In National Mathematics Advisory Panel, *Foundations for success: The final report of the National Mathematics Advisory Panel* (pp. 4-i-4-221). U.S. Department of Education.
- Goldin-Meadow, S., Wagner Cook, S., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, 20(3), 267-272. <https://doi.org/10.1111/j.1467-9280.2009.02297.x>
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Erlbaum.
- Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96(4), 284-309. <https://doi.org/10.1016/j.jecp.2006.09.001>
- Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, 38(3), 258-288.
- Jones, I., Inglis, M., Gilmore, C., & Dowens, M. (2012). Substitution and sameness: Two components of a relational conception of the equals sign. *Journal of Experimental Child Psychology*, 113, 166-176. <https://doi.org/10.1016/j.jecp.2012.05.003>
- Kaput, J., Carraher, D., & Blanton, M. (Eds.). (2008). *Algebra in the early grades*. Erlbaum.
- Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, 12, 317-326. <https://doi.org/10.1007/BF00311062>
- Kindrat, A., & Osana, H. P. (2018). The relationship between mental computation and relational thinking in the seventh-grade. *Fields Mathematics Education Journal*, 3(1), Article 6. <https://doi.org/10.1186/s40928-018-0011-4>
- Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, 37(4), 297-312. <https://doi.org/10.2307/30034852>
- Lee, K., Ng, S.-F., Ng, E.-L., & Lim, Z.-Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology*, 89(2), 140-158. <https://doi.org/10.1016/j.jecp.2004.07.001>
- Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, 31(4), 357-378. <https://doi.org/10.1017/S0140525X08004469>
- LeFevre, J., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition* (pp. 361-378). Psychology Press.
- 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 in the United States. *Cognition and Instruction*, 26(2), 195-217. <https://doi.org/10.1080/07370000801980845>
- Matthews, P. G., & Fuchs, L. S. (2020). Keys to the gate? Equal sign knowledge at second grade predicts fourth-grade algebra competence. *Child Development*, 91(1), e14-e28. <https://doi.org/10.1111/cdev.13144>
- Matthews, P. G., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education*, 43(3), 316-350. <https://doi.org/10.5951/jresemetheduc.43.3.0316>
- McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology*, 20(3), 93-108.

- McNeil, N. M. (2008). Limitations to teaching children $2 + 2 = 4$: Typical arithmetic problems can hinder learning of mathematical equivalence. *Child Development*, 79(5), 1524-1537. <https://doi.org/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(1), 42-47. <https://doi.org/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. <https://doi.org/10.1037/0022-0663.92.4.734>
- McNeil, N. M., & Alibali, M. W. (2005). Knowledge change as a function of mathematics experience: All contexts are not created equal. *Journal of Cognition and Development*, 6(2), 285-306. https://doi.org/10.1207/s15327647jcd0602_6
- 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(4), 1109-1121. <https://doi.org/10.1037/a0028997>
- 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 math equivalence. *Child Development*, 82(5), 1620-1633. <https://doi.org/10.1111/j.1467-8624.2011.01622.x>
- McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C., & McKeever, M. O. (2019). Consequences of individual differences in children's formal understanding of mathematical equivalence. *Child Development*, 90(3), 940-956. <https://doi.org/10.1111/cdev.12948>
- McNeil, N. M., Hornburg, C. B., Fuhs, M. W., & O'Rear, C. (2017). Understanding children's difficulties with mathematical equivalence. In D. C. Geary, D. B. Berch, & K. Mann Koepke (Eds.), *Mathematical cognition and learning* (pp. 167-195). Elsevier. <https://doi.org/10.1016/B978-0-12-805086-6.00008-4>
- Ministère de l'Éducation, du Loisir et du Sport. (2011). *Indice de défavorisation par école: 2010-2011*. Québec, Canada: Author.
- Mix, K. S. (1999). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology*, 74(4), 309-332. <https://doi.org/10.1006/jecp.1999.2533>
- Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).
- Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, 6(4), 449-468. [https://doi.org/10.1016/0885-2014\(91\)90049-J](https://doi.org/10.1016/0885-2014(91)90049-J)
- Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20(2), 110-122. <https://doi.org/10.1016/j.lindif.2009.10.005>
- Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology*, 91(2), 137-157. <https://doi.org/10.1016/j.jecp.2005.01.004>
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77(1), 1-15. <https://doi.org/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(1), 175-189. <https://doi.org/10.1037/0022-0663.91.1.175>
- Rittle-Johnson, B., Matthews, P. G., Taylor, R. S., & McEldon, K. L. (2011). Assessing knowledge of mathematical equivalence: A construct-modeling approach. *Journal of Educational Psychology*, 103(1), 85-104. <https://doi.org/10.1037/a0021334>
- Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, 24(3), 181-266. <https://doi.org/10.1006/ceps.1999.1004>
- Schliemann, A., Carraher, D., Brizuela, B., Earnest, D., Goodrow, A., Lara-Roth, S., & Peled, I. (2003). Algebra in elementary school. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the joint meeting of the International Group for the Psychology of Mathematics Education and the North American Chapter of the International Group for the Psychology of Mathematics Education - Volume 4* (pp. 127-134). University of Hawai'i, Honolulu.
- Seethaler, P. M., & Fuchs, L. S. (2006). The cognitive correlates of computational estimation skills among third-grade students. *Learning Disabilities Research & Practice*, 21(4), 233-243. <https://doi.org/10.1111/j.1540-5826.2006.00220.x>
- Seo, K., & 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). Lawrence Erlbaum Associates.

- Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology*, *101*(1), 88-100. <https://doi.org/10.1037/a0013156>
- Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, *10*, 104-109. <https://doi.org/10.1111/j.1467-7687.2007.00571.x>
- Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In R. Lesh & A. E. Kelly (Eds.), *Research design in mathematics and science education* (pp. 267-306). Erlbaum.
- Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research*, *81*(4), 453-492. <https://doi.org/10.3102/0034654311423025>
- Tolar, T. D., Lederberg, A. R., & Fletcher, J. M. (2009). A structural model of algebra achievement: Computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology*, *29*(2), 239-266. <https://doi.org/10.1080/01443410802708903>
- Tronsky, L. N., & Royer, J. M. (2003). Relationships among basic computational automaticity, working memory, and complex mathematical problem solving: What we know and what we need to know. In J. M. Royer (Ed.), *Mathematical cognition* (pp. 117-145). Information Age.
- Ulrich, C., Tillema, E. S., Hackenberg, A. J., & Norton, A. (2014). Constructivist model building: Empirical examples from mathematics education. *Constructivist Foundations*, *9*(3), 328-339.
- Watchorn, R. (2011). *Improving children's understanding of mathematical equivalence* [Doctoral dissertation, University of Alberta]. Education and Research Archive, University of Alberta. https://era.library.ualberta.ca/items/730c2e31-3de7-4c1e-af74-898f278af0db/view/e8487031-d4dd-4f7a-9f6a-fa9d804f32c9/Watchorn_Rebecca_Spring-202011.pdf
- Watchorn, R., & Bisanz, J. (2005, October 21-22). *Enduring instructional effects on equivalence problems* [Poster presentation]. Cognitive Development Society, San Diego, CA, USA.
- Watchorn, R., Osana, H. P., Sherman, J., Taha, M., & Bisanz, J. (2011). Improving mathematical equivalence performance in Grades 2 and 4. In L. R. Weist & T. Lamberg (Eds.), *Proceedings of the thirty-third annual conference of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1811-1812). University of Nevada, Reno, United States.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Riverside.

Appendix

Instructional Fidelity Checklist

1. Wrote “=” on the board and asked students what the symbol was
2. Wrote equivalence problem on the board
3. Used different colors to differentiate both sides
4. Explained to students the goal of equivalence problems
5. Gestured to each side during the lesson
6. Emphasized that there are two sides to an equivalence problem
7. Explained the meaning of “=”
8. Solved equivalence problem on the board
9. Gave extra examples
10. Showed exercises with missing equal sign
11. Gave students practice with missing equal sign problems
12. Gave students time to practice in small groups or pairs
13. Wrap-up with reiteration of what “=” means



Journal of Numerical Cognition (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.