Research-Based Teaching Practices for Improving Students' Understanding of Mathematical Equivalence Have Not Made it into Elementary Classrooms

Elena M. Silla (esilla@wisc.edu)
University of Wisconsin-Madison
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Caroline Byrd Hornburg (chornburg@vt.edu)
Virginia Tech
Department of Human Development and Family Science, 295 W. Campus Drive
Blacksburg, VA 24061 USA

Matthew Kloser (mkloser@nd.edu)
University of Notre Dame
Center for STEM Education, 107 Carole Sandner Hall
Notre Dame, IN 46556 USA

Nicole M. McNeil (nmcneil@nd.edu)
University of Notre Dame
Department of Psychology, 390 Corbett Family Hall
Notre Dame, IN 46556 USA

Abstract

Elementary math instruction traditionally has emphasized procedures rather than concepts. Thus, students tend to lack a strong understanding of foundational concepts like mathematical equivalence. Cognitive scientists and mathematics educators have found small yet effective ways to modify traditional arithmetic instruction to promote students' conceptual understanding of math equivalence. Educational standards also now reflect this academic research. However, it is unclear whether classroom practices have caught up with research and policy. In the current study, we observed teachers' practices during arithmetic instruction. The goal was to determine if teachers are using research-based practices that promote understanding of math equivalence and if variation in use of research-based practices is associated with students' growth in understanding of math equivalence across the school year. Eight second and third grade classrooms (*M* students per classroom = 23) were observed twice during math instruction. Students completed a math test both before and after the observation period. Research-based practices were rarely observed in any classrooms, so there was not much variation in classroom use of research-based practices to predict student growth. Students improved their performance on all problem types tested, but performance on math equivalence problems was significantly lower than on other problem types. Results suggest that policies and practices designed to improve students' understanding of math equivalence may not have filtered down to affect instructional practices in classrooms.

Keywords: mathematical equivalence; mathematics instruction; conceptual understanding; pre-algebra; mathematical cognition

Deep mathematics learning involves understanding how to work with symbols, systems, and problem-solving procedures (i.e., procedural knowledge), as well as understanding the underlying relations among these symbols, systems, and problem-solving procedures (i.e., conceptual knowledge; Hiebert & Lefevre, 2013). However, students often struggle to connect their conceptual knowledge to the procedures they are performing (Carpenter, Franke, & Levi, 2003; McNeil & Alibali, 2005; Seo & Ginsburg, 2003). One possible source of students' difficulties connecting concepts to procedures is the way mathematics is traditionally taught in school. Mathematics classrooms in the United States traditionally have used a method of teaching that emphasizes *how* a problem is solved rather than *why* it is solved that way (e.g., Carpenter, Fennema, Peterson, & Carey, 1988; Cobb, 1987). This procedural emphasis leads children to treat mathematics as a series of isolated facts rather than as a coherent set of "big ideas" and associations among concepts (Charles, 2005; Cobb, 1987; Jacobs, Franke, Carpenter, Levi, & Battey, 2007). Consequently, students have trouble transferring the knowledge gained from experience with arithmetic to the novel equations they see later in algebra (e.g., McNeil & Alibali, 2005).

The disconnect between students' procedural knowledge and conceptual understanding is particularly evident when assessing students' formal understanding of mathematical equivalence (Knuth, Stephens, McNeil, & Alibali, 2006; Rittle-Johnson & Alibali, 1999). Math equivalence, or the relation between two mathematical expressions that are equal and interchangeable, is a foundational concept in mathematics (Baroody & Ginsburg, 1983; Blanton & Kaput, 2003; Charles, 2005; Knuth et al., 2006). Formal understanding of this concept predicts later math achievement (McNeil, Hornburg, Devlin, Carrazza, & McKeever, 2019) and algebra performance (Hornburg, Devlin, & McNeil, 2020; Matthews & Fuchs, 2020). Yet, this concept is difficult for many children ages 7-11 (e.g., Falkner, Levi, & Carpenter, 1999; McNeil, 2007; McNeil & Alibali, 2005; McNeil, Hornburg, Fuhs, & O'Rear, 2017). Instead of viewing the equal sign as demonstrating that both sides of an equation have the same value, many students view it as a signal to do something, such as add up all the numbers in an equation (Baroody & Ginsburg, 1983; McNeil & Alibali, 2005). This operational understanding of the equal sign reflects a larger trend of shallow mathematical understanding in the United States. Indeed, the United States was largely outperformed by its peers on a global high school math assessment (Hanushek, Peterson, & Woessmann, 2010); while performance between states varied, all states were outperformed by over a dozen countries. A more recent assessment shows a similar trend (Organisation for Economic Co-Operation and Development, 2019).

To address children's woeful understanding of mathematical equivalence and underperformance in mathematics, math educators have called for more emphasis on pre-algebraic concepts in the early grades (e.g., Blanton & Kaput, 2003). This has led to changes to both national and state standards in the United States, such as the Common Core State Standards (NGA Center & CCSSO, 2010), that have been designed to incorporate understanding of the equal sign and math equivalence into the early grades. These changes are now several years old.

Researchers also have identified specific instructional practices that teachers can use to help children meet these standards. One practice is to use a variety of problem formats when teaching arithmetic, including non-traditional problems (e.g., $c = a + b$, $a = a$), as opposed to merely using traditional problem formats (e.g., $a + b = c$, $\begin{array}{r} \text{a} \\ \text{+ b} \\ \hline \text{c} \end{array}$) (McNeil, 2008; McNeil, Fyfe, & Dunwiddie, 2015; McNeil, Fyfe, Petersen, Dunwiddie, & Brletic-Shipley, 2011). Another is using a "concreteness fading" technique, by demonstrating equivalence first with physical objects before slowly fading them away to create a bridge to the abstract symbols (Fyfe, McNeil, & Borjas, 2015). Two other research-based practices have focused specifically on improving children's understanding of the equal sign: (a) using relational language that describes the equal sign as representing two equal and interchangeable quantities (e.g., "is the same amount as") and (b) explicitly referring to the equal sign verbally or through gesture (Carpenter et al., 2003; Cook, Duffy, & Fenn, 2013). These practices yield a deeper understanding of mathematical equivalence, particularly for students who struggle in mathematics (Powell & Fuchs, 2010).

We now have specific standards for students in the early grades that promote understanding of mathematical equivalence, as well as a set of small yet impactful changes to arithmetic instruction that can help children meet these standards. Yet, it is unclear whether these advances in policy and research have translated into changes in teaching practices. Do teachers use instructional strategies that promote understanding of math equivalence, or do they still use traditional methods for teaching arithmetic? Traditional practices, which tend to promote procedural understanding over conceptual understanding, may hinder students' understanding of math equivalence. For example, writing problems with the operations on the left side and the equal sign and answer blank at the end promotes a unidirectional, procedural view of equations. Using arithmetic-specific, operational language (e.g., saying "the total" or "makes") in reference to the equal sign can also be detrimental, given that students with arithmetic-specific interpretations of the equal sign are less likely to learn from instruction on early algebra concepts (Byrd, McNeil, Chesney, & Matthews, 2015). Classroom-specific case studies and textbook analyses have been conducted to observe teaching practices related to the equal sign (see Powell, 2012; Seo and Ginsburg, 2003), but there have been no studies examining arithmetic teaching practices across several diverse classrooms since the change in standards.

The primary goal of this study was to determine what arithmetic instruction looks like in early elementary classrooms. Another goal was to analyze whether individual differences in teachers' use of research-based practices was correlated with students' growth in understanding of math equivalence over the course of a school year. If classroom practice is a source of students' difficulties in understanding math equivalence, then differences in input should predict student growth. We hypothesized that teacher practices would remain largely unchanged from the traditional practices, but teachers who had begun to shift to research-based practices would have greater classroom-wide improvement in understanding of math equivalence.

Method

Design

The study was a pretest, observation, posttest design, with two classroom observations occurring between pretest and posttest administration. Two classrooms administered the posttest prior to the second observation. However, administration was still within one week of the observation and thus did not change our approach for analysis.

Participants

Eight second and third grade classrooms ($M = 23$ students per classroom, min = 16, max = 26) within four schools participated. Schools included both public and parochial schools recruited through discussions with principals and included schools at both ends of the socioeconomic spectrum. There were 186 students across the eight classrooms, but only 174 students completed the pretest and posttest. Table 1 shows the demographics of each school.

Procedure

Teachers administered a pretest (described below) within a week prior to the first observation. Each classroom was then observed twice, once in the fall (October or November) and once in the spring (February).

Table 1: Demographic characteristics by school.

Classrooms	Race/Ethnicity of Students in the School	Percentage of Students with Free/Reduced Lunch	Grade(s)	Number of Students Completing Pre and Post
A, B	2.2% Black, 6.8% Hispanic, 87.1% White, 3.9% Other	0.0%	2, 2	26, 26
C, D	27.3% Black, 8.5% Hispanic, 48.3% White, 15.9% Other	87.5%	2, 2[a]	20, 22
E, F	0.4% Black, 93.1% Hispanic, 2.6% White, 3.9% Other	98.3%	2, 3	23, 21
G, H	1.6% Black, 13.3% Hispanic, 81.9% White, 3.2% Other	32.5%	2, 3	16, 20

[a]Classroom D was a mixed classroom of 2nd graders and high-achieving 1st graders.

Classrooms were observed during direct arithmetic instruction by two trained observers for an average of 30.5 minutes. Efforts were made to observe both classrooms in the same school within the same week. Observers noted behaviors specified in an observation checklist (described below) and made a tally each time a specific behavior occurred. Teachers administered a posttest within one week of the second observation. The average time between pretest and posttest was 122 days.

Measures

Pretest and Posttest The pretest and posttest were identical paper-and-pencil assessments administered by teachers. There were four math equivalence problems, which are problems that had operations on both sides of the equal sign (e.g., 8 + 2 = __ + 6), one simple arithmetic problem presented in a nontraditional format (12 = 7 + __), and five problems assessing general math achievement appropriate for second grade (California Department of Education, 2009). Problems were intermixed in a set order, with one problem presented per page. Teachers were provided a script to ensure uniform administration, and each problem was read aloud.

Observations Observers were trained to use an observation checklist by watching publicly available videos of math instruction. Videos included example behaviors within each of four categories. Observers then compared observations to establish reliability of the checklist and among observers. The checklist consisted of four research-based practices shown to affect students' understanding of math equivalence. The first category looked at teachers' use of various *problem formats*. These included traditional problems in both left-to-right and vertical formats, non-traditional problems in both right-to-left and reflexive (e.g., $a = a$) formats, and math equivalence problems. The second category observed teachers' use of *concreteness fading* and, more globally, tracked the use of concrete (e.g., blocks) and abstract (e.g., symbols) examples in instruction. The third category focused on the *language* teachers used in reference to the equal sign. Language could either emphasize the relational nature of the equal sign (e.g., "is the same as"), or it could emphasize an operational definition of the equal sign that was specific to arithmetic (e.g., "the total") or generally procedural (e.g., "the answer"). The fourth category tracked if teachers *explicitly referred to the equal sign* verbally or through gestures. Observers also made note of anything that could potentially affect conceptual understanding of math equivalence, including wall posters, and use of run-on equations that hinder understanding (e.g., 20 + 30 = 50 + 7 = 57; Carpenter et al., 2003).

Data Analysis Approach

Pretest and Posttest Coding Correctness was determined based on the students' written responses. Given that performance historically has been so low on math equivalence problems, with students using incorrect strategies like adding all the numbers (e.g., McNeil & Alibali, 2005), we gave students credit if they were within ±1 of the response that would be achieved using a correct strategy if they also solved another math equivalence problem exactly correct (cf. Hornburg, Rieber, & McNeil, 2017; McNeil, 2007). Thus, our strategy for coding correctness on math equivalence problems should make it easier (not harder) to be correct on those problems compared to the other problems.

Observation Coding Observers observed arithmetic instruction during 13 of the 16 classroom observations, and they observed arithmetic instruction in all classrooms during at least one of the two observations. Observers made a tally mark in the specific category whenever the format or behavior occurred. Correlations between the count of behaviors noted by the two observers were calculated for each category. The average correlation between observers was .98. One observation was excluded from analyses due to lack of adherence to observation protocol; in this case, analyses were based on the other observer's observation.

Once reliability among observers was established, the two totals within each category were averaged to create classroom scores for each category. In some cases, observations of the practices were so infrequent that an average score did not make sense to compute. In these cases, a dichotomous variable was created to analyze whether the practice was or was not ever observed.

Results

Classroom Use of Research-Based Practices

Teachers used traditional practices during observations, and they rarely used research-based practices. Fewer than 1% of problems observed across classrooms were written in a format other than the traditional formats. Out of 572 observed problems across 16 classroom observations, only two non-traditional (e.g., __ = 3 + 7) equations were observed, and both occurred during the same observation of Classroom E. Although most teachers (75%) used concrete examples, this accounted for only 12% of the types of examples used overall, and none were observed using the "concreteness fading" technique to link concrete representations to abstract math symbols. Four teachers (Classrooms B, E, F, and G) used relational language, but it was rare compared to other types of language. It accounted for only 5% of teachers' language overall. The other 95% of observed language referred to the equal sign operationally.

Finally, observers noted a large emphasis on math fact fluency, with few references to mathematical relations or equality. Three teachers (Classrooms B, F, and G) referred to the equal sign verbally or through gesture, but within those observations it was seldom referenced (only 12 times across those observations). Observers also noted posters on two classroom walls that showed traditional formats which may have promoted an operational view of the equal sign (e.g., addition and subtraction "doubles" facts; 5 + 5 = 10, 10 – 5 = 5), as well as practices in one classroom (Classroom G) that may have reinforced an operational view (e.g., using run-on equations like 10 × 6 = 60 – 6 = 54).

As observers visited the classrooms, they noted that half the teachers emphasized conceptual understanding of problem-solving strategies, whereas the other half focused primarily on procedural understanding. The emphasis on conceptual understanding involved behaviors such as allowing students to reason why problems were solved a certain way or asking them to generate multiple strategies. These observations inspired a post-hoc category based on the classroom's emphasis on conceptual versus procedural understanding. Classrooms A, B, D, and E were categorized as "conceptual," whereas the remaining classrooms, which exhibited no noted behaviors to promote conceptual understanding, were categorized as "procedural."

Classroom Mathematics Performance

Table 2 presents math test performance by classroom. We analyzed classroom performance on the math test from pretest to posttest based on student accuracy (%) on each problem type in each classroom. We conducted a 3 (problem type: math equivalence, nontraditional arithmetic, general mathematics) x 2 (time of observation) repeated measures ANOVA and found two large main effects. Averaging across time, classroom accuracy solving math equivalence problems (M = 37.73, SE = 6.78) was much lower than the percentage for nontraditional arithmetic (M = 71.45, SE = 6.75) or general mathematics (M = 72.26, SE = 5.11) problems, $F(2, 14) = 46.34$, $p < .001$, $\eta_p^2 = .87$. Averaging across problem type, classroom accuracy was higher at posttest (M = 68.44, SE = 5.16) than at pretest (M = 52.52, SE = 6.87), $F(1, 7) = 19.45$, $p = .003$, $\eta_p^2 = .74$. We did not find evidence of an interaction, $F(2, 14) = 1.40$, $p = .28$, $\eta_p^2 = .17$, but we recognize power for detecting an interaction was low.

Although research-based practices were not often observed, some classrooms did demonstrate use of a few research-based practices, and all classrooms improved at least some in understanding of math equivalence over the time studied (see Table 2). Descriptively, there were no obvious associations between use of research-based practices and change in understanding of math equivalence other than the fact that the least improvement in understanding of math equivalence occurred in Classroom C, which had no evidence of using research-based practices and was classified as procedural, whereas the greatest improvement occurred in Classroom B, which had evidence of two research-based practices and was classified as conceptual. Note that this did not seem to be an across-the-board "better classrooms use better practices and produce better growth" phenomenon, as growth on the general math problems for Classroom C was above the median and for Classroom B it was below the median.

Table 2: Math test performance by classroom.

	Performance on Each Problem Type (*M* % Accuracy)					
Classroom	Math Equivalence		Non-traditional Arithmetic		General Mathematics	
	Pre	Post	Pre	Post	Pre	Post
A	62.50	79.81	84.62	84.62	83.08	83.85
B	25.00	71.15	69.23	92.31	74.62	80.77
C	5.00	10.00	25.00	60.00	44.00	58.00
D	25.00	56.82	45.45	81.82	44.55	70.91
E	34.78	44.56	91.30	91.30	89.57	88.70
F*	8.33	28.57	47.62	52.38	47.62	69.52
G	37.50	48.44	93.75	93.75	81.25	88.75
H*	23.75	42.50	45.00	85.00	72.00	79.00

Note. Asterisks next to classroom letters indicate third grade.

Discussion

The arithmetic lessons we observed focused primarily on traditional practices for teaching arithmetic and did not contain much evidence of the research-based practices for promoting understanding of math equivalence that align with policy changes. This was most evident in the lack of variety in problem format, as teachers primarily presented problems in traditional formats, the format most often presented in math textbooks (Powell, 2012). Although teachers sometimes used concrete examples in instruction, we did not see evidence of them using the research-based concreteness fading technique in which they start with a concrete equivalence context, like sharing or balancing, and explicitly link to symbols while slowly fading out the context (see Fyfe et al., 2015). Furthermore, although half of the classrooms used relational language (e.g., "same as") at least once and nearly half referenced the equal sign verbally or through gesture at least once, it was rare for a teacher to engage in these behaviors more than once or twice within a lesson. Only Classroom E used relational language more than twice, and only Classroom F ever referenced the equal sign more than once. Even classrooms that used research-based practices used traditional practices far more often.

Overall, teachers tended to emphasize traditional practices and fact fluency, with many observed lessons containing warm-up problems with addition and subtraction and with multiplication tables. While procedural understanding of facts such as these are crucial for the development of mathematical cognition (Rittle-Johnson & Alibali, 1999; Rittle-Johnson & Siegler, 1998), there was little conceptual teaching to bolster the procedural practice. Occasionally the research-based practices used were not clear to both observers, suggesting that they may not be noticeable to students, particularly when procedural practices are much more prevalent in the classroom.

Other times, when research-based practices were evident, they were often brief and overshadowed by the prevalence of traditional practices. For example, during the first observation of Classroom E (which was categorized as "conceptual" and showed some use of research-based practices), the teacher presented two non-traditional problems (a reflexive equation, $140 = 140$, and a right-to-left equation, $36 = 9 \times 4$) in addition to 28 problems presented in the traditional left-to-right format. However, the teacher then stated that one of the equations, $36 = 9 \times 4$, was "backwards" and rewrote it in a traditional format, $9 \times 4 = 36$. This may have further confused the meaning of the equal sign for students (see Capraro, Ding, Matteson, Capraro, and Li, 2007; McNeil, 2008). Instances like these may explain why the overall accuracy on math equivalence problems was so low across classrooms compared to the other problem types tested. Incorporating more research-based practices into instruction, and using fewer traditional practices, may lead to greater student understanding of math equivalence.

Perhaps the easiest change to make would be the incorporation of non-traditional problem formats (e.g., $\_\_ = 3 + 4$, $7 = 3 + \_\_$, $7 = \_\_ + 4$) into the classroom. These types of problems were rarely used by teachers during our observations even though they are relatively easy to integrate into the classroom and have been shown to improve students' understanding of math equivalence (McNeil et al., 2011; McNeil et al., 2015). Incorporation of metacognitive strategies that emphasize conceptual thinking, such as self-explanation and worked examples, may also lead to greater understanding of math equivalence (Barbieri, Miller-Cotto, & Booth, 2019; Carpenter et al., 2003; Rittle-Johnson & Star, 2007). There was some evidence of metacognitive reflection in the four classrooms categorized into the posthoc, conceptual instruction category.

Our original hypothesis was that use of research-based practices for teaching arithmetic would be positively associated with growth in students' understanding of math equivalence across the school year (Johannes, Davenport, Kao, Hornburg, & McNeil, 2017; McNeil et al., 2015; McNeil, Hornburg, Brletic-Shipley, & Matthews, 2019). This hypothesis was based on prior work by McNeil, Hornburg, Brletic-Shipley, and Matthews (2019) demonstrating that an intervention incorporating a conglomerate of research-based practices, including introducing the equal sign outside of an arithmetic context, concreteness fading, non-traditional arithmetic practice, and comparing and explaining a variety of problem formats and problem-solving strategies, improved students' understanding of math equivalence more than well-structured non-traditional arithmetic practice alone. In the end, we could not conceptually replicate this finding in a naturalistic classroom setting because there was so little variability in use of research-based practices in the classroom lessons that we observed.

Results of the present study highlight the disconnect between research-based practices and what actually occurs in classrooms. This disconnect may be mitigated through addressing these research-based practices in professional development seminars. In a study done by Jacobs et al. (2007), students of teachers who participated in a year-long professional development seminar that focused on relational thinking and student understanding of equivalence were better able to solve math equivalence problems than were students of teachers who did not participate in this seminar. These types of seminars may also help teachers recognize misconceptions in their own classrooms, since teachers often are unaware of the prevalence of misconceptions of math equivalence (Stephens, 2006). Additionally, incorporating aspects of teaching relational thinking in professional development seminars may encourage teachers to engage in some of these practices in their classrooms, increasing not only the quantity of research-based nontraditional practices, but potentially the quality of such practices, as well.

One limitation to this study was its size and scope. Although it is striking that the research-based practices were relatively absent in all four schools, including the most affluent one, all eight classrooms were in same city in the Midwestern United States, and each classroom was only observed twice. Furthermore, we were not able to document every instance of instruction that could have aided in students' understanding of math equivalence, as we were limited by our particular checklist of research-based practices. This provides a very narrow picture of what behaviors teachers typically engage in. Additionally, in three observation periods, the math lessons observed did not focus primarily on arithmetic due to miscommunication with the teachers, which left us with only one observation to analyze for these classrooms. Also, observers were also only trained to note each instance of the observed behavior, not the behavior's duration. Instances of practices that were implemented for longer periods of time may have had a larger influence over student understanding. Future research should include more classrooms and more observation sessions, as well as more precise observations with duration of the behaviors recorded, in order to capture a more complete picture of what teachers are doing when teaching arithmetic.

Finally, even though children's understanding of math equivalence was low and far below that of children in higher-achieving countries (e.g., Capraro et al., 2010), children did grow in their understanding of math equivalence over the time period studied. This growth does not appear to merely be explained by an increase in general math understanding. Future work should explore what factors account for the growth seen across classrooms. If few research-based practices are being used, then how does children's understanding of math equivalence grow? Identifying such factors could help expand the focus of current interventions.

Teachers do not seem to be using research-based practices for improving students' understanding of math equivalence that align with recent policy changes. Thus, it could be useful for future studies to examine how to improve teachers' uptake of teaching practices that are effective. Doing so could help teachers align their practices with the recent standards changes and may bolster student understanding in mathematics.

Acknowledgments

The research reported here is based on a senior thesis conducted by Silla under the direction of McNeil and Kloser and assistance of Hornburg. It was funded through the ISLA Undergraduate Research Opportunity Program at the University of Notre Dame. Thanks to members of the Cognition, Learning, and Development Lab at the University of Notre Dame for assistance with data collection.

References

Barbieri, C. A., Miller-Cotto, D., & Booth, J. L. (2019). Lessening the load of misconceptions: Design-based principles for algebra learning. *Journal of the Learning Sciences, 28,* 381-417.

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.

Blanton, M. L. & Kaput, J. J. (2003). Developing elementary teachers' "algebra eyes and ears". *Teaching Children Mathematics, 10,* 70-77.

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.

California Department of Education (2009). *CST released test questions: Released questions for the California Standards Test (CSTs)*. Retrieved from http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp

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.

Capraro, M. M., Ding, M., Matteson, S., Capraro, R. M., & Li, X. (2007). Representational implications for understanding equivalence. *Research in Brief, 107*, 86-88.

Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education, 19*, 385-401.

Carpenter, T. P., Franke, M. K., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in early 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.

Cobb, P. (1987). An investigation of young children's academic arithmetic contexts. *Educational Studies in Mathematics, 18*, 109-124.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*, 1863-1871.

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., McNeil, N.M., & Borjas, S. (2015). Benefits of "concreteness fading" for children's mathematics understanding. *Learning and Instruction, 35*, 104-120.

Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2010). *U.S. math performance in global perspective: How well does each state do at producing high-achieving students?* Cambridge, MA: Program on Education Policy and Governance & Education, Harvard University Kennedy School.

Hiebert, J. & Lefevre, P. (2013). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-23). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hornburg, C. B., Devlin, B. L., & McNeil, N. M. (2020). *Earlier understanding of mathematical equivalence in elementary school predicts greater 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.
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.
Johannes, K., Davenport, J., Kao, Y., Hornburg, C. B., & McNeil, N. M. (2017). Promoting children's relational understanding of equivalence. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 600-605). Austin, TX: Cognitive Science Society.
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.
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*, e14-e28.
McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds in equivalence problems. *Developmental Psychology, 43,* 687-695.
McNeil, N. M. (2008). Limitations to teaching children 2 + 2 = 4: Typical arithmetic problems can hinder learning of math equivalence. *Child Development, 79,* 1524-1537.
McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, *76*, 883-899.
McNeil, N. M., Fyfe, E. R., & Dunwiddie, A. E. (2015). Arithmetic practice can be modified to promote understanding of math equivalence. *Journal of Educational Psychology, 107*, 423-436.
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*, 1620-1633.
McNeil, N. M., Hornburg, C. B., Brletic-Shipley, H., & Matthews, J. M. (2019). Improving children's understanding of math equivalence via an intervention that goes beyond nontraditional arithmetic practice. *Journal of Educational Psychology*, *111*, 1023-1044.
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 math equivalence. *Child Development*, *90*, 940-956.
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.
NGA Center & CCSSO. (2010). *Common core state standards for mathematics*. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf
Organisation for Educational Co-operation and Development (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris.
Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal, 112,* 627-648.
Powell, S. R., & Fuchs, L. S. (2010). Contribution of equal-sign instruction beyond word-problem tutoring for third-grade students with mathematics difficulty. *Journal of Educational Psychology, 102,* 381-394.
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.
Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In E. Donlan (Ed.), *The development of mathematical skills* (pp. 75-110). London, England: Psychology Press.
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.
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: Lawrence Erlbaum Associates.
Stephens, A. C. (2006). Equivalence and relational thinking: Preservice elementary teachers' awareness of opportunities and misconceptions. *Journal of Mathematics Teacher Education*, *9*, 249-278.