Journal for Research in Mathematics Education
2012, Vol. 43, No. 3, xxx–xxx

Measure for Measure: What Combining Diverse Measures Reveals About Children's Understanding of the Equal Sign as an Indicator of Mathematical Equality

Percival Matthews
University of Notre Dame

Bethany Rittle-Johnson and Katherine McEldoon
Vanderbilt University

Roger Taylor
State University of New York at Oswego

Knowledge of the equal sign as an indicator of mathematical equality is foundational to children's mathematical development and serves as a key link between arithmetic and algebra. This study extended prior efforts to use a construct modeling approach to unify diverse measures into a single assessment designed to measure knowledge of the equal sign. Children in Grades 2–6 (N = 224) completed the assessment. The current findings reaffirmed a past finding that diverse items can be integrated onto a single scale, revealed the wide variability in children's knowledge of the equal sign assessed by different types of items, and provided empirical evidence for a link between equal-sign knowledge and success on some basic algebra items.

Key words: Algebra; Arithmetic; Elementary, K–8; Item-response theory; Knowledge; Measurement

---

This research was supported with funding from the National Science Foundation (NSF) grant DRL-0746565. It was also supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B040110 to Vanderbilt University. The opinions expressed are those of the authors and do not represent the views of the NSF or of the U.S. Department of Education. A special thanks to the students and teachers who participated in this research. Thanks to Caroline Cochrane-Braswell, Holly Harris, Robin Hill, Caitlyn Majeika, Laura McClean, Jadie Pinkston, Ann Simonson, Christine Tanner, Kristen Tremblay, and Ross Troseth for help in reviewing the research literature and collecting and coding the data. Thanks to E. Warren Lambert of the Vanderbilt Kennedy Center and Ryan Kettler of the Learning Sciences Institute at Vanderbilt for statistical help. A special thank you to Maria Blanton, Tom Carpenter, Analucia Schliemann, and Jon Star for their expert opinions on the items and construct map.

The equality relationship denoted by the equal sign is a foundational concept that serves as a key link between arithmetic and algebra (Baroody & Ginsburg, 1983; Carpenter, Franke, & Levi, 2003; Kieran, 1981; Knuth, Stephens, McNeil, & Alibali, 2006; MacGregor & Stacey, 1997). Despite decades of research committed to investigating children's understanding of the equal sign and the related concept of mathematical equality, limitations of measurement have posed impediments to maximizing the depths of our knowledge. This study was part of a project using a psychometric approach to develop assessment instruments that enhance our understanding of children's knowledge of the equal sign (see also Rittle-Johnson, Matthews, Taylor, & McEldoon, 2011).

We undertook this study against a backdrop in which Hill and Shih (2009) noted that less than 20% of studies published in the *Journal for Research in Mathematics Education* over the past 10 years had reported on the validity of the measures. The authors concluded that "without conducting and reporting validation work on key independent and dependent variables, we cannot know the extent to which our assessments tap what they claim to." Our assessment instrument is at the vanguard of efforts to address these concerns with respect to measuring children's understanding of the equal sign as an indicator of mathematical equality. More generally, it provides a model for use of a new, psychometrically oriented method for measurement development.

To foreshadow our approach, we used item response theory to integrate item types that are often used separately. Combining these item types allowed us to paint a richer picture of children's understanding than was previously possible from considering the different types of items in isolation. In this article, we map much of the variability in children's knowledge that goes undetected by current measures and provide empirical data that inform the extent to which children's understanding of the equal sign contributes to algebraic thought.

CHILDREN'S KNOWLEDGE OF THE EQUAL SIGN

Introduced by Recorde in the 16th century, the "=" sign has become the universally recognized symbol for indicating mathematical equality (Cajori, 1928). Mathematical equality can loosely be defined as the principle that two sides of an equation have the same value and are thus interchangeable (Kieran, 1981; Wittgenstein, 1961). From a more formal perspective, mathematical equality is an example of an equivalence relation and therefore satisfies the symmetric, reflexive, and transitive properties.

Much of what people know about the equal sign, however, is demonstrated not in their abilities to cite formal mathematical properties but instead lies implicitly in their understandings of the equal sign and its uses (see Zhu & Simon, 1987). Indeed, much of how we come to understand any given symbol—and the concepts that it denotes—is connected to participation in various *language games* (Wittgenstein, 2001; see also de Saussure, 1959). Simply put, it is by mastering the symbol's meaning in different contexts (i.e., becoming familiar with its use in language games) that we come to know its referent. Functional knowledge of mathematical equality is in large part knowledge of the "=" symbol, both in terms of how it is interpreted directly and in terms of the procedures and transformations that it sanctions.

A well-developed conception of the equal sign applicable to elementary and middle school children is characterized by *relational* understanding: realizing that the equal sign symbolizes the sameness of the expressions or quantities represented by each side of an equation (e.g., Baroody & Ginsburg, 1983; Behr, Erlwanger, & Nichols, 1980; Carpenter et al., 2003; McNeil & Alibali, 2005a). There is general agreement that relational understanding of the equal sign supports greater algebraic competence, including equation-solving skills and algebraic reasoning (e.g., Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007; Jacobs, Franke, Carpenter, Levi, & Battey, 2007; Kieran, 1992; Knuth et al., 2006; National Council of Teachers of Mathematics & Mathematical Sciences Education Board, & National Research Council, 1998; Steinberg, Sleeman, & Ktorza, 1991). Because algebra is an important gateway not only into higher mathematics, but also into higher education more generally (Adelman, 2006; Moses & Cobb, 2001), the importance of building high-quality relational understanding of the equal sign is of critical importance.

Unfortunately, numerous studies point to the difficulties that American elementary and middle school children have understanding the equal sign (e.g., Alibali, 1999; Behr et al., 1980; Cobb, 1987; Falkner, Levi, & Carpenter, 1999; Jacobs et al., 2007; Li, Ding, Capraro, & Capraro, 2008; McNeil, 2007; Perry, 1991; Powell & Fuchs, 2010; Rittle-Johnson, 2006; Rittle-Johnson & Alibali, 1999; Weaver, 1973). Children often interpret the equal sign as an operator signal that means "adds up to" or "gets the answer" (e.g., Baroody & Ginsburg, 1983; McNeil & Alibali, 2005b). As a result of this misconception, most children in Grades 1 to 6 solve problems such as $8 + 4 = \square + 5$ incorrectly, writing 12 or 17 in the box (e.g., Falkner et al., 1999). Similarly, most American elementary school children reject closed equations that do not appear in a standard "$a + b = c$" format (e.g., they consider equations such as $3 = 3$ and $7 + 6 = 6 + 6 + 1$ to be false or nonsensical; Baroody & Ginsburg, 1983; Behr et al., 1980; Falkner et al., 1999). These errors are thought to result from children's almost exclusive exposure to problems in standard (indicated)-operations-equals-answer format (e.g., $4 + 5 = 9$) (Falkner et al., 1999; McNeil et al., 2006; Li et al., 2008).

MEASURING KNOWLEDGE OF THE EQUAL SIGN—A UNIFIED APPROACH

Despite the progress that has been made in detailing the sources and signs of children's misconceptions about the equal sign, two important measurement issues remain significant obstacles to advancing our understanding of children's knowledge of the equal sign and the associated concept of mathematical equality. First, researchers have not used assessments with proven reliability and validity when investigating children's understanding of the equal sign (Hill & Shih, 2009; Rittle-Johnson et al., 2011). Second, the assessments of equal-sign knowledge that are currently in use have tended to be incommensurable. That is to say that (a) different researchers have used different types of items to measure what is taken to be the same construct, and (b) this diversity of independently employed items makes it difficult to compare results obtained across studies.

The problem of incommensurable measures is rooted in the fact that research on children's understanding of the equal sign has employed primarily three different types of items: (a) *open equation-solving items*, such as 8 + 4 = □ + 5 (e.g., Alibali, 1999; Behr et al., 1980; Falkner et al., 1999; Jacobs et al., 2007; McNeil, 2007; Matthews & Rittle-Johnson, 2009; Perry, 1991; Powell & Fuchs, 2010; Rittle-Johnson, 2006; Weaver, 1973); (b) *equation-structure items*, such as deciding whether 3 + 5 = 5 + 3 is true or false (e.g., Baroody & Ginsburg, 1983; Behr et al., 1980; Falkner et al., 1999; Molina & Ambrose, 2006; Rittle-Johnson & Alibali, 1999; Seo & Ginsburg, 2003); and (c) *equal-sign-definition* items, such as asking children to provide an explicit verbal definition of what the equal sign means (Behr et al., 1980; Ginsburg, 1977; Knuth et al., 2006; McNeil et al. 2006; Seo & Ginsburg, 2003). Somewhat less frequently, a fourth type of item is used: *advanced relational reasoning* items, such as asking children to solve the equation 37 + 54 = □ + 55 without having to add 37 + 54 (e.g., Blanton, Levi, Crites, & Dougherty, 2011; Carpenter et al., 2003; Jacobs et al., 2007).

Although different in form, each of these types of items has been accepted as tapping children's knowledge of the equal sign as an indicator of the equality relation. A majority of past studies have privileged one or another of these types of items over the others. Researchers who have employed multiple item types have tended to compartmentalize them, analyzing different item types separately (e.g., Jacobs et al., 2007; McNeil & Alibali, 2005b; Powell & Fuchs, 2010; Rittle-Johnson, 2006; Seo & Ginsburg, 2003). Thus, we are left unaware of the relative difficulties of commonly used items or the typical order in which competence is gained.

DETAILING THE CONSTRUCT

To address these limitations, we sought to unify these diverse items into a single assessment. We used Mark Wilson's construct modeling approach to measurement development as a guide for our efforts (Wilson, 2003, 2005). We began by developing a *construct map*, which is a representation of the continuum of knowledge levels for the construct under consideration. We based our construct map for children's knowledge of the equal sign as an indicator of mathematical equality on the general agreement that understanding of the equal sign ranges from operational on the low end to relational on the high end (e.g., Alibali et al., 2007; Carpenter et al., 2003; Kieran, 1981). Our map fleshes out this dichotomy, resolving it into four levels ranging from rigid operational to more advanced comparative relational conceptions of the equal sign. The levels of the construct map are distinguished by the types of contexts in which the equal sign is understood as knowledge progresses (see Table 1). Note that our construct map was not intended to be an expansive concept map covering the universe of thought that the concept of mathematical equality might be argued to span. Instead, it covers a subset of symbolic forms using

Table 1
Construct Map for Knowledge of the Equal Sign as Indicator of Mathematical Equality

Level	Description	Core equation structure(s)
Level 4: Comparative Relational	Successfully solve and evaluate equations by comparing the expressions on the two sides of the equal sign, including using compensatory strategies and recognizing transformations maintain equality. Consistently generate a relational interpretation of the equal sign.	Equations that can be most efficiently solved by applying simplifying transformations: For example, without adding 67 + 86, can you tell if the number sentence "67 + 86 = 68 + 85" is true or false?
Level 3: Basic Relational	Successfully solve, evaluate, and encode equation structures with operations on both sides of the equal sign. Recognize relational definition of the equal sign as correct.	Operations on both sides: $a + b = c + d$ $a + b - c = d + e$
Level 2: Flexible Operational	Successfully solve, evaluate, and encode atypical equation structures that remain compatible with an operational view of the equal sign.	Operations on right: $c = a + b$ No operations: $a = a$
Level 1: Rigid Operational	Only successful with equations with an operations-equals-answer structure, including solving, evaluating, and encoding equations with this structure. Define the equal sign operationally.	Operations on left: $a + b = c$ (including when blank is before the equal sign)

Note. Table adapted from Rittle-Johnson et al. (2011, p. 87).

the equal sign that has been studied previously with school-aged children. Hence, our construct map has mostly been restricted to understanding the equal sign in arithmetical contexts as well as select early extensions to algebra (e.g., letters as variables, addition property of equality).

The construct map we developed in Rittle-Johnson et al. (2011) is presented in Table 1, with less sophisticated knowledge represented at the bottom and more advanced knowledge represented at the top. Students at Level 1, the *rigid operational level*, are expected to have success with equations in the standard operations-equals-answer format, but to fail with equations in other formats. At Level 2, the *flexible operational level*, children maintain an operational view of the equal sign, but become somewhat more flexible with respect to the types of equation formats that they correctly solve and accept as valid. Children at this level specifically become comfortable with equations that are atypical but remain consistent with an operational view of the equal sign, such as equations that are "backwards." These backwards equations remain consistent with an operational view because their mirror images are equations in standard format (e.g., a student in Behr et al., 1980, rewrote □ = 4 + 5 as 5 + 4 = □). At Level 3, children begin to hold a *basic relational* view, although it coexists with an operational view. Children's nascent relational understanding is primarily manifested in their success with equations with operations on both sides (e.g., 4 + 5 + 8 = □ + 8), and they recognize a relational definition of the equal sign as a good definition. Finally, children at Level 4 have a *comparative relational* understanding of the equal sign. The hallmark competence at this level is understanding the symbolic transformations that change the form of an expression without changing the equality relation symbolized by the equal sign. For example, children with a comparative understanding know that performing the same permissible actions (limited to applications of familiar functions, and with certain exceptions, such as division by zero) on each side of an equation maintains the equality of the quantities represented by the two sides and makes it unnecessary to engage in full computation. Similarly, children at this level can recognize the equality between the sides of equations such as 67 + 86 = 68 + 85 without the need to compute. These children realize that the equal sign represents a relation between the two sides of the equation and that the relations among numbers in the two expressions make it unnecessary to carry out the calculations (Carpenter et al., 2003). Because some Level 4 items can be solved using inefficient calculate-and-compare methods, most items assessing Level 4 knowledge require that children show evidence of strategies or explicitly relational reasoning.

Although the construct map is presented as having four levels for purposes of conceptual clarity, our model of the construct is continuous. The continuous nature of the model means that the levels should not be interpreted as discrete stages. Instead, knowledge change is expected to follow a gradual and dynamic progression, with less sophisticated knowledge sometimes coexisting and competing with more advanced knowledge (Siegler, 1996). For example, an operational view of the equal sign can be elicited even from adults in certain circumstances (McNeil & Alibali, 2005a, 2005b).

Evidence from our earlier study strongly supported the validity of our construct map and the reliability of our assessment (Rittle-Johnson et al., 2011). We found that all the items on the original assessments substantially loaded on the same construct, indicating that they could be measured on a unidimensional interval-level scale. We used the assessment from Rittle-Johnson et al. (2011) as a springboard for the development of the refined assessment used in the current study.

THE CURRENT STUDY: GAINING LEVERAGE BY COMBINING MEASURES

In Rittle-Johnson et al. (2011), we focused on the psychometric properties of our original assessment. In this article, we discuss a refinement and replication of our original study with a focus on how the assessment provides insights into the nature of children's knowledge. In addition to replicating our past findings with a new population of public school (rather than parochial school) students, we concentrated on two key contributions that were not addressed in the first article.

Contribution 1: Elaborating the Variability in Children's Knowledge

Our assessment approach adds significant resolution to the wide range of variability that exists among children's knowledge of the equal sign. Currently, we are usually limited to binary partitions that result from characterizing children according to whether they succeed or fail at a particular type of item. Either–or distinctions hide the within-group variability in the binary groups that result. Our integrated assessment approach reveals the variability in children's knowledge that would otherwise go undetected. We did not capitalize on this strength in our previous article.

Contribution 2: Evidence for the Importance of Equal-sign Knowledge

Establishing a unified measurement scale for equal-sign knowledge helps inform our notions of the extent to which children's understanding of the equal sign impacts different kinds of mathematical reasoning. Our scale also can serve as an anchor for evaluating whether new items also load heavily on the construct of equal-sign knowledge. For instance, our earlier work showed that using a single scale could allow us to place a few items typically considered to require algebraic reasoning on the same scale as other items typically used to assess understanding of the equal sign in arithmetic contexts. These included items requiring children to demonstrate advanced relational reasoning about transformations that preserve equality (e.g., if we know that $76 + 45 = 121$, can we tell without adding whether or not $76 + 45 - 9 = 121 - 9$?, inspired by Alibali et al., 2007; Carpenter et al., 2003; Steinberg et al., 1991) and items with equations involving letter variables (e.g., $n + n + 2 = 17$, Jacobs et al., 2007).

This was an important step toward verifying the link between algebraic thought and knowledge of the equal sign. It is often argued that early appreciation of the relational character of the equal sign is key for later success in algebra and other higher mathematics (e.g., Baroody & Ginsburg, 1983; Blanton et al., 2011; Knuth et al., 2006; MacGregor & Stacey, 1997; Steinberg et al., 1991), but the empirical evidence supporting such claims is somewhat indirect. No other researchers have explicitly attempted to put algebraic reasoning items on the same measurement scale as those known to measure knowledge of the equal sign. Although we used a few of these items in Rittle-Johnson et al. (2011), we used more diverse items in this study to expand our earlier findings and to illustrate more explicitly how they related to variability in children's knowledge of the equal sign in arithmetic contexts.

Establishing some subset of algebraic items as internally consistent with other items known to measure knowledge of the equal sign provides empirical evidence for the link between knowledge of the equal sign and algebraic reasoning. Moreover, this feature of our method can be used to evaluate new potential items as they arise, providing insight into whether new candidate items also load heavily on the construct of equal-sign knowledge. In this way, our method potentially may be used as an exploratory tool to inform our conceptions of the construct of equal-sign knowledge and the breadth of its reach.

METHOD

Participants

Data were collected from 13 second-grade through sixth-grade classrooms in two suburban public schools in Tennessee near the end of the school year. There were 224 children who completed the assessment. Of those, 53 were in second grade (23 girls), 46 were in third grade (25 girls), 29 were in fourth grade (14 girls), 59 were in fifth grade (26 girls), and 37 were in sixth grade (16 girls). The mean age was 10.3 years (SD = 1.6; Min = 7.7; Max. = 14.1). Most of the children were Caucasian; approximately 2% of children were from minority groups. The schools served a working- to middle-class population, with approximately 23% of children receiving free or reduced lunch.

Children in Grades 2–5 were taught using the Harcourt Mathematics curriculum (Maletsky et al., 2005), and Grade 6 children were taught using Glencoe's Mathematics: Applications & Concepts Course curriculum (Bailey et al., 2004). Their teachers had not received any professional development focused on supporting understanding of the equal sign or of mathematical equality.

Test Development Procedure

We created two comparable forms of the assessment using a step-by-step item-matching procedure to ensure similarity of content and difficulty across forms (see Rittle-Johnson et al., 2011, for details on item matching). Most of the items on our assessment were taken directly from previously published work or created based on items present in previous work (Baroody & Ginsburg, 1983; Behr et al., 1980; Carpenter et al., 2003; Jacobs et al., 2007; Knuth et al., 2006; Matthews & Rittle-Johnson, 2009; Rittle-Johnson, 2006; Seo & Ginsburg, 1983; Warren, 2003; Weaver, 1973). Items were classified as Level 1, 2, 3, or 4 based on their equation structures and required solution strategies, as outlined in Table 1. In total, there were 31 items on each form of the assessment: 13 Level 4 items, 10 Level 3 items, 5 Level 2 items, and 3 Level 1 items. See Appendix A for a complete list of items on one of the forms and Appendix B for the hypothesized position of each item on the four-level construct map.

The items used in Rittle-Johnson et al. (2011) were vetted by a panel of four mathematics education experts, each of whom had over 10 years experience conducting research on elementary school children's algebraic thinking. We made revisions to the original assessment based on empirical evidence of item performance and feedback from the panel. Five items from each form of the assessment used in Rittle-Johnson et al. (2011) were cut due to weak psychometric properties, which were often combined with poor ratings from the experts. Eight items were added to each form based on the advice of the panel.

Two of the eight items added were Level 2 items, asking whether 4 = 4 + 0 was true or false and what the equal sign meant in the context of "1 quarter = 25 pennies."[1] Six of the eight new items on each form were aimed at further assessing more advanced relational thinking (Level 4 items) that is associated with algebraic thought. Four of these items on each form tested children's practical knowledge of the arithmetic properties of equality, which hold that an equality relationship remains true as long as an identical permissible function is applied to both sides of the equation (i.e., to both the expression to the left of and the expression to the right of the equal sign; see Figure 1 for a sample item). These types of problems have been cited both as addressing the concept of equality and as addressing the types of relational thought that underlie formal transformational algebra (Blanton et al., 2011; Carpenter et al., 2003; Jacobs et al., 2007; National Research Council, 2001; Steinberg et al., 1991). Although these items can be solved by computation, we coded performance based upon children's explanations of how the problem can be answered without computation. Answers coded as correct needed to call upon explicit knowledge of the properties of equality (e.g., "minus 7 is on both sides so you don't need it"). The final two new items on each form featured letter variables that are commonly seen in formal algebra. For instance, one asked children to "find the value of n" for the equation $n + n + n + 2 = 17$ (from Jacobs et al., 2007). There was only a single item with letters as variables on each of the original assessments. None of the items using letters as variables featured products symbolized by the juxtaposition of variables or constants (e.g., $3x - 2 = 7$), as expertise with this symbolic form was expected to be confusing for children at the age levels sampled.

7. *Without subtracting* the 9, can you tell if the statement below is true or false?

76 + 45 = 121 is true.

Is 76 + 45 – 9 = 121 – 9 true or false?

True False Can't tell without subtracting

How do you know?

Figure 1. Sample item probing children's knowledge of transformations that preserve equality..

[1] Some researchers caution that using the equal sign in these contexts can lead to confusion due to errors that result from word-matching biases (e.g., Clement, 1982). The current case, however, does not involve constructing an equation, a context in which the error typically arises. Moreover, our panel judged the item as appropriate for measuring knowledge of equal-sign knowledge.

Test Administration

The assessment was administered on a whole-class basis by members of the project team. We used a spiraling technique to distribute the two forms of the assessment in each classroom, alternating between handing out the first and second forms of the assessment. Completion of the assessment required approximately 45 minutes. Test directions were read aloud for each type of item in second-grade classrooms to minimize the possibility that reading level would affect performance. Otherwise, test administration was identical across grade levels. Children received no instruction on the equal sign as part of this study.

Scoring

Each item was scored dichotomously (i.e., 0 for incorrect or 1 for correct). For computation items, children received a point for answers within one of the correct answer to allow for minor calculation errors. This scoring convention is commonly used in the literature (e.g., McNeil, 2008; Rittle-Johnson, 2006). Applying stricter recoding that did not allow for these calculation errors yielded nearly identical results, changing the accuracy classifications for only 3% of all trials. For the nine explanation items, children received a point if they mentioned the equality relation between values on the two sides of the equation (see Table 2). After initial coding, an independent rater coded responses for a randomly selected 20% of the sample, with a mean inter-rater agreement of 0.99 for Form 1 (range .96 to 1.00) and .97 for Form 2 (range .87 to 1.00).

The Rasch Measurement Model

We used a Rasch model along with methods from Classical Test Theory to evaluate the performance of the assessment. Rasch modeling is a one-parameter

Table 2
Coding Scheme for Select Explanation Items

Item	Sample correct responses	Sample incorrect answers
What does the equal sign (=) mean?	"It means the same as" "Is equal to or has the same amount"	"The answer to the question" "The sum"
Without subtracting the 7, can you tell if the statement is true or false? 56 + 85 = 141 is true. Is 56 + 85 – 7 = 141 – 7 true or false?	"Because it's subtracting 7 from both sides and 56 + 85 = 141 then subtract seven and it'll be equal." "Because before you minus seven you have 141 for both."	"I did the math." "In my head I subtracted and got the same answer." "Because I looked at both number sentences and they didn't match"
Without adding 89 + 44, can you tell if the number sentence is true or false? "89 + 44 = 87 + 46"	"Because all you do is take two from 89 and put two on 44." "Because 89 – 87 = 2 and they add 2 to 44 so it is even."	"Because 89 + 44 = 133 and 87 + 46 = 133 too." "You just have to add the numbers up."

member of the item response theory (IRT) family (Bond & Fox, 2007). The Rasch model estimates both respondent ability and item difficulty simultaneously, yielding the probability that a particular respondent will answer a particular item correctly (Rasch, 1960/1993; Wright, 1977). We used Winsteps software version 3.68.0.2 (www.winsteps.com) to perform all IRT estimation procedures. In addition to providing item and respondent parameters, the Rasch model estimation procedure provides information on the goodness of fit between empirical parameter estimates and the measurement model via infit values (see Linacre, 2010). Infit values for an item between 0.5 and 1.5 indicate that the item fits well with the other items on the test.

One intuitive description of a Rasch model is as a probabilistic Guttman scalogram. It integrates the difficulty hierarchy of the Guttman model with a bit more flexibility: The Rasch model is a probabilistic one, which is consistent with a model of human understanding that allows for different types of understandings to coexist at the same time (e.g., Siegler, 1996). One of the advantages of the model is that it uses empirical results to place items on a true continuum. Hence, our construct map is only the conceptual skeleton upon which our model is built. Once the empirical data are used to add substance to the skeleton, those data can be used to seamlessly cover all four levels of the map. These empirical estimates, along with the fit scores, can be used to address our primary concerns about the variability of children's understandings and about whether or not some basic algebra items can properly be grouped with other items that measure understanding of the equal sign.

Because equivalent groups took the two forms of the assessment, we were able to use a single IRT model to estimate item difficulty and respondent ability across the forms (Kolen & Brennan, 2004). Indicators confirmed the baseline equivalence of the knowledge of the equal sign by the children who completed the two forms. The distribution of forms was even within each grade level, and groups were also statistically equivalent in mean age (Form 1 = 10.3 years, Form 2 = 10.2 years) and mean grade (Form 1 = 4.0, Form 2 = 3.9).

RESULTS

In presenting the results, we first briefly discuss evidence for the reliability and validity of the assessment, which supports the feasibility of measuring all the items on the same scale. Next, we describe how our method can illustrate the variability in children's equal-sign knowledge. Finally, we provide evidence for the importance of equal-sign knowledge for some basic algebraic competence.

Evidence for Reliability and Validity—A Single Construct Model

Internal consistency, as assessed by Cronbach's a, was high for both forms of the assessment (Form 1 = .93; Form 2 = .94), providing support for the reliability of the assessment. Multiple measures provided evidence for the validity of our assessment. As discussed previously, four mathematics education experts rated the items, providing evidence for face validity of the test content. The four experts rated nearly all the test items as ranging from *important* (rating of 3 out of 5) to *essential* (rating of 5 out of 5) items for tapping knowledge of equality, with a mean rating of 4.3.

We evaluated whether our construct was reasonably characterized as tapping a single dimension. Within an IRT framework, the unidimensionality of an assessment is often checked by using a principle components analysis (PCA) of the residuals after fitting the data to the Rasch model (Linacre, 2010). This analysis attempts to partition unexplained variance into coherent factors that may indicate other dimensions. The Rasch model accounted for 60% of the variance in our data set. A PCA on the residuals indicated that the largest secondary factor accounted for 2% of the total variance (eigenvalue of 3.2), corresponding to 5% of the unexplained variance. The secondary factor was sufficiently dominated by the Rasch dimension to justify the assumption of unidimensionality (Linacre, 2010). This confirmed that all items on the assessment—including the algebraic ones—substantially loaded on the same construct. Each item tapped children's understanding of the equal sign as an indicator of mathematical equality.

Next, as a check on the internal structure of the assessment, we evaluated whether our a priori predictions about the relative difficulty of items were correct (Wilson, 2005). Recall that when creating the assessment, we selected items to tap knowledge at each of the four levels on our construct map. The hypothesized difficulty level for each item (ordinal 1, 2, 3 or 4; see Appendix B) correlated highly with the empirically derived item difficulty, Spearman's $\rho(62) = .91$, $p < .01$. We also used an item-respondent display called a Wright map to help evaluate our difficulty predictions (Wilson, 2005). The Wright map allows for quick visual inspection of whether our construct map correctly predicted relative item difficulties (Figure 2). In brief, a Wright map consists of two columns, one for *respondents* and one for items. On the left column are respondents (i.e., participants). Those with the highest ability scores on the construct are located near the top of the map, whereas those with the lowest scores are located near the bottom. Assessment *items* are located on the right column. The most difficult items are located near the top of the map, and the least difficult ones are near the bottom. The vertical line between these two columns indicates the scale for both the ability and difficulty parameter estimates measured in logits (i.e., log-odds units). Interpreting logits involves knowing that the average of the item distribution was set to 0 logits; negative scores indicate items that were easier than average, and positive scores indicate items that were more difficult than average. The chief advantages of the logit scale are (a) it is an interval-level scale that can be properly used for parametric analysis, and (b) it can be used to calculate the probability that the average participant of a given ability level will be successful on an item of a particular difficulty. The Wright map shown in Figure 2 is condensed to represent the selected items discussed below. The full Wright map is available from the authors upon request.

As can be seen on the Wright map, the items we had categorized as Level 4 items were indeed the most difficult items (i.e., they clustered near the top with difficulty scores greater than 0); the items we had categorized as Levels 1 and 2 items were

51

```
PERSON - MAP - ITEM

PERSONS - LOGITS - ITEMS

             6
             |
           . |
             |
           # 5T
             |
            T|
           . |
             4
       .#### |
             | If 76+45=121,does 76+45-9=121-9?
         .## |
           # 3 Judge "89+44=87+46" as T/F
        .### | If 56+85=141,does 56+85-7=141-7?      n+n+n+2=17. Find the value of n.
           # |S
     ####### S|
       .#### 2
             | c+c+4=16. Find the value of c
     .###### | What does the equals sign mean?
     ####### |
     ####### 1
       .#### |
          .# |
        .### |
        #### MOM
----------------------------------------------------------------------------------------------
         ### | 7+6+4=7+c
        #### | □ +2=6+4          Judge "7+6=6+6+1" as T/F
         .## | 13=n+5. Find the value of n.
          .# -1 Judge "8=8" as T/F      Rating of "The equal sign means two amounts are the same"
          ## |
----------------------------------------------------------------------------------------------
       ##### |
        .### |
         ### -2
       .#### S|
           # |S
    .####### | 8=6+□             Judge "4=4+0" as T/F
----------------------------------------------------------------------------------------------
         .## -3
        #### |
         ### |
           # |
          ## -4
             |
           # |
            T|
           -5T
             | □+5=9
           . |
           . |
            -6

       EACH "#" IS 2 PEOPLE. EACH "." IS 1 PERSON.
```

Figure 2. The Wright map. Each "#" represents two respondents, and each "." represents one respondent. Numbers on the vertical axis represent item difficulties and children's ability estimates in logits. M = mean, S = standard deviation, and T = two standard deviations.

indeed fairly easy items (i.e., clustered near the bottom with difficulty scores less than –1); and Level 3 items fell in between. Overall, the distribution of items on the Wright map supported our hypothesized levels of knowledge, progressing in difficulty from a rigid operational view at Level 1 to a comparative relational view at Level 4. After confirming that items were clustered as expected, we added horizontal lines on the Wright map corresponding to approximate cut points between levels based on empirically observed clustering. We added these lines to aid discussion, but it should be remembered that the construct is at root a continuous measure and that speaking in terms of levels is merely a convention to facilitate discussion.

We also found that children's ability levels behaved as expected. First, the ability estimates of individual children were highly correlated with grade level $r(224) = .72$, $p < .01$. Second, the correlation between the TCAP (Tennessee Comprehensive Assessment Program) mathematics scores for children in Grades 3–6 and their ability estimates was also high $r(170) = .70$, $p < .01$. This positive correlation between our assessment and a general standardized mathematics assessment provides some evidence of convergent validity.

The current results, collected from a new population in a public school district, replicate our original findings of adequate psychometric properties for our assessments (Rittle-Johnson et al., 2011). These findings provide strong evidence for the reliability and validity of our assessment of equal-sign knowledge. In particular, they show that different item types can be measured on a single scale, with a hierarchy of item difficulty that matches our construct map.

Elaborating the Variability in Children's Knowledge

A major goal of this study was to add resolution to our conception of the variability in children's understandings of the equal sign as an indicator of mathematical equality. To illustrate this contribution, we first discuss how the current model augments the discussions put forth by studies that focus on individual types of items. We exploit the fact that individual children's ability scores, as estimated by the Rasch model, can be used to find specific probabilities that a given child will be successful on a given item. Specifically, we can calculate the probability of any participant's success on any given item from log-odds units by using the equation

$$\Pr(\mathit{success}) = \frac{1}{1 + e^{-(\theta - d)}},$$

where θ is a participant's ability estimate, and d is the item difficulty estimate. This is a powerful analytical tool, because it allows us to take a single measure (a child's ability score) and use it to predict the types of items with which a child is likely to struggle, without the usual need for resource-intensive item-by-item error analysis. It is important to note that neither a participant's ability score nor an item's difficulty is interpretable on its own; it is the distance between the two, $\theta - d$, that yields the probability of an individual's success. The contrasts between these probability estimates for different children will be used to help provide a detailed picture of the variability in children's knowledge of the equal sign. To aid interpretation, we selected 6 representative ability estimates—corresponding to high and low scores within Levels 2, 3, and 4—for the purposes of illustration. Table 3 lists the probabilities that children at each of these various ability estimates will generate correct answers for selected items (see Appendix A for a list of all items on Form 2 of the assessment). The interested reader can select given items, using the equation above to further explore the likelihood that children with different ability estimates will correctly solve particular items.

Open-equation-solving items. Several researchers have suggested that children's difficulties with solving open equations should be dependent upon the formats of those open equations (Falkner et al., 1999; Weaver, 1973), and the levels of our construct map are based in part upon this idea. The more nonstandard or unfamiliar the format, the more difficult the item should be. Our earlier study (Rittle-Johnson et al., 2011) was the first to quantify the differences in difficulty among nonstandard equations of different formats. Due to the psychometric focus of that study, however, the practical implications of the findings were rather opaque. Here, we have replicated the findings, and subsequently we discuss what they mean for our picture of children's emerging knowledge of the equal sign.

The simplest nonstandard open-equation item was $\square + 5 = 9$. This was expected to be the case because this item still adhered to standard operations-equals-answer format, and was classified at Level 1. As shown in Table 3, even the lowest performers were able to solve this item correctly over 90% of the time.

The next most difficult items were those with all operations on the right of the equal sign. Although children on the upper half of the ability continuum showed complete mastery for these items, children of lesser ability were substantially less likely to solve these items correctly than items in operations-equals-answer format. For instance, low Level 2 children solved $\square + 5 = 9$ correctly 92% of the time, but solved $8 = 6 + \square$ correctly only 50% of the time.

Finally, items became even more difficult when they involved operations on both sides of the equal sign. For instance, $\square + 2 = 6 + 4$ was considerably more difficult than $\square + 5 = 9$ for all but the most skilled children. Correct performance for low Level 2 children dropped to 10% for this item. Although low Level 3 children had largely mastered solving equations with operations on the right, they were expected to get this item correct only 38% of the time. Interestingly, the somewhat longer item $7 + 6 + 4 = 7 + \square$ was of nearly identical difficulty despite the fact that the item involves an extra addend and an extra addition sign. As predicted by our construct map, the placement of operations relative to the equal sign was a key factor affecting difficulty, whereas the number of addends and whether the unknown was on the left or right side had little impact on difficulty. In summary, equations with all operations on the right were generally more difficult than equations with all operations on the left (the distinction between Level 1 and Level 2), and equations with operations on both sides were generally more difficult than equations with operations only on a single side (the distinction between Level 2 and Level 3).

Equation-structure items. We expected equation format to have a similar impact on the difficulty of rating equations as true or false (see Behr et al., 1980). Indeed, equations with operations limited to the right side, such as 4 = 4 + 0, though nonstandard, were not as difficult for children to grasp as noncanonical equations with no operators (e.g., 8 = 8) or those with operations on both sides, such as 7 + 6 = 6 + 6 + 1 (see Table 3). High Level 2 children largely accepted 4 = 4 + 0 as true (78% correct), but fell to chance for 8 = 8 (46% correct) and rarely accepted 7 + 6 = 6 + 6 + 1 as true (28% correct). High Level 3 performers exhibited clear mastery for 4 = 4 + 0 (94%) and were correct the vast majority of the time for 8 = 8 (80%). Their performance also declined a bit for 7 + 6 = 6 + 6 + 1, the item with operators on both sides of the equal sign (64%).

Overall, the data suggest that it is important to consider *which* nonstandard formats children accept. Some nonstandard formats are much more difficult than others. Moreover, it appears that equation format affects difficulty similarly for both open-equation and equation-structure item types. For instance, as shown in Table 3, items with all operations on the right were of similar difficulty for both open-equation and equation-structure item types (e.g., 8 = 6 + □ and 4 = 4 + 0). Problems with operations on both sides were also of similar difficulty across item types and proved to be more difficult than problems with operations on the right (e.g., 7 + 6 + 4 = 7 + □ and 7 + 6 = 6 + 6 + 1).

Equal-sign items. We also were interested in how children's abilities to give a relational definition of the equal sign was related to their success on other items designed to tap their explicit knowledge of the equal sign, as well as on open-equation and equation-structure items. First, we should note that requiring children to provide a relational definition of the equal sign was more difficult than other equal-sign items. Recognizing a relational definition of the equal sign from a list—as opposed to generating one—was much easier. For instance, asking children to rate the phrase "The equal sign means two amounts are the same" as a good or not good definition was much easier than generating a relational definition of the equal sign (see Table 3).

Providing a relational definition of the equal sign proved to be very difficult for many children. For example, both high Level 3 children (at the beginning relational level) and high Level 2 children (at the flexible operational level) were unlikely to offer a relational definition of the equal sign (20% and 1% probabilities, respectively; see Table 3). Similar performance on this item, however, stands in stark contrast to the differences in performance these same children demonstrate on other items. For instance, when asked to evaluate the equation 4 = 4 + 0 as true or false, the high Level 3 child is successful 93% of the time, whereas the low Level 2 child is expected to be successful only 47% of the time. Similarly, when asked to solve □ + 2 = 6 + 4, high Level 3 children are successful 64% of the time, despite their general failure to generate a relational definition of the equal sign. By contrast, the low Level 2 children are expected to succeed on this item only 10% of the time. These examples show how including diverse measures in one hierarchy allows our assessment to map variability that might otherwise remain unnoticed.

Table 3
Coding Scheme for Select Explanation Items

Item	Item difficulty estimate	Child ability estimate (θ = Rasch ability estimate)					
		Low Level 2 $\theta = -2.7$	High Level 2 $\theta = -1.35$	Low Level 3 $\theta = -0.95$	High Level 3 $\theta = 0.10$	Low Level 4 $\theta = 0.61$	High Level 4 $\theta = 3.75$
Open-equation items							
□ + 5 = 9	-5.21	0.92	0.98	0.99	1.00	1.00	1.00
8 = 6 + □	−2.71	0.50	0.80	0.85	0.94	0.97	1.00
□ + 2 = 6 + 4	−0.47	0.10	0.29	0.38	0.64	0.75	0.99
7 + 6 + 4 = 7 + □	−0.31	0.08	0.26	0.35	0.60	0.72	0.98
Equation-structure items							
4 = 4 + 0 T or F	−2.56	0.47	0.77	0.83	0.93	0.96	1.00
8 = 8 T or F	−1.18	0.18	0.46	0.56	0.78	0.86	0.99
7 + 6 = 6 + 6 + 1 T or F	−0.7	0.12	0.34	0.44	0.69	0.79	0.99
Equal-sign items							
Rate "The equal sign means two amounts are the same" as good or bad	−.96	0.15	0.40	0.50	0.74	0.83	0.99

Table 3 (*Continued*)

What does the equal sign (=) mean?	1.51	0.01	0.05	0.08	0.20	0.29	0.90
Letters as variable							
$13 = n + 5$	−.75	0.12	0.35	0.45	0.70	0.80	0.99
$c + c + 4 = 16$	1.76	0.01	0.04	0.06	0.16	0.24	0.88
$m + m + m = m + 12$	2.67	0.00	0.02	0.03	0.07	0.11	0.75
Advanced relational thinking items							
Explain if "89 + 44 = 87 + 46" is true or false without computing.	2.15	0.01	0.03a	0.04	0.11	0.18	0.83
Explain why, without subtracting the 9, if 76 + 45 = 121 is true, 76 + 45 – 9 = 121 – 9 is true, too.	3.48	0.00	0.01	0.01	0.03	0.05	0.57

Note. See Table 2 for coding criteria for explanation items. Each of the six rightmost columns corresponds to a child of a given ability estimate, as denoted by the Rasch ability estimate in the column heading. Entries below the column headings represent the probabilities that the average child of a given ability estimate will answer an item correctly.

Our data also helped chart the variability within the group of children who succeeded at offering a relational definition of the equal sign. In our sample, 29 children (13% of the sample) showed mastery for generating a relational definition of the equal sign (difficulty = 2.51). We defined *mastery* as receiving ability estimates that were at least one logit above the item difficulty (i.e., the point at which children are expected to succeed on the item at least 73% of the time). Of these master definers, 13 children scored more than two logits higher than the difficulty of the item (i.e., they were expected to get the item correct at least 88% of the time). These 13 most skilled children were more likely than the other master definers to succeed on more difficult items asking them to provide an explanation using explicit arithmetic properties of equality. Take, for instance, performance on the item that asks, "76 + 45 = 121 is true. Is 76 + 45 – 9 = 121 – 9 true or false? How do you know?" (difficulty = 3.41). The more highly skilled subgroup of master definers was successful on this item roughly twice as often as the less skilled subgroup of master definers (29% vs. 52%). Thus, there was even detectible variability within the mastery-level children.

Finally, providing a relational definition of the equal sign did not necessarily mean that children had abandoned an operational view of the equal sign. Whereas 26% of our sample could offer a relational definition, only 10% of our sample—or roughly 2 out of every 5 relational definers—provided only a relational definition. The remaining children offered alternative operational interpretations alongside their relational definitions. This was consistent with our construct map, which posits that a relational view of the equal sign can coexist with other less sophisticated views. We could not include both codings (i.e., strictly relational vs. relational alongside operational) in our Rasch model due to the model's requirements for item independence, so we do not have a difficulty estimate for the stricter coding of definitions.

Evidence for the Importance of Equal-sign Knowledge

Equations with letters as variables. We evaluated whether the items involving letters as variables loaded on the construct of equal-sign knowledge, as expected. Each of the three items featuring letters as variables (see Table 3) was well aligned with the construct, both according to indicators from classical test theory and measures specific to the Rasch model. Item total correlations were above .45, and infit statistics of model fit were within acceptable ranges of 0.5 to 1.5 for each of the items. This is an important point: the items using letters as variables fit the Rasch model as well as other items typically used to study knowledge of the equal sign. This suggests that this subset of algebraic items relies, in part, on knowledge of the equal sign.

The three items were not of equal difficulty. Instead, difficulty varied as the format of the item changed. The item $13 = n + 5$ received a difficulty rating of –.75, making it a mid Level 3 difficulty item. Use of a letter as a variable appeared to make this item more difficult than other open-equation items with a similar format and no letter variable. For example, $8 = 6 + \square$ was significantly easier, with a difficulty of −2.71 (Table 3). Even though use of the letter n is logically similar to the use of the blank for representing the unknown, it appears that the use of a letter renders the item more difficult.

The item $c + c + 4 = 16$, on the other hand, was a low Level 4 item. Even though all operations were on the left, this item proved more difficult than the most difficult open-equation-solving items on the assessment (i.e., $43 + \square = 48 + 76$, difficulty = 1.08). This is presumably because the item involved multiple instances of the unknown, demanding more novel application of the concept of equality (Carpenter et al., 2003).

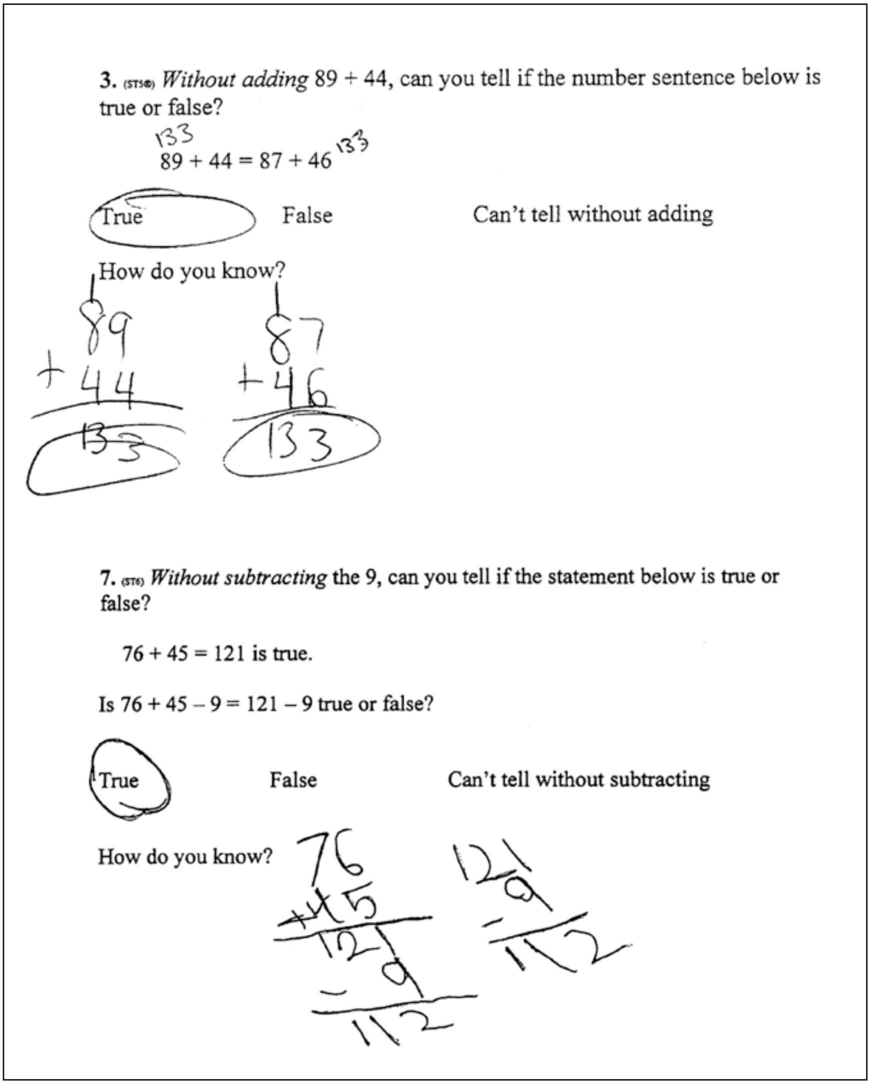
Finally, $m + m + m = m + 12$ emerged as a high difficulty Level 4 item (difficulty = 2.67). This item involved both multiple instances of a letter as variable and operations on both sides of the equal sign. In fact, it was the most difficult of all assessment items not requiring an explanation. Only 11% of low Level 4 children were expected to solve this item properly, whereas 72% of them are expected to solve $7 + 6 + 4 = 7 + \square$ correctly (see Table 3). It was of even higher difficulty than the item asking for a verbal definition of the equal sign.

Three main points summarize our findings about items involving letters as variables. First, items with letters as variables fit our model well, suggesting that they tapped the construct of equal-sign knowledge. Second, items with letters as variables were more difficult than similarly formatted items without letters as variables. Third, children with higher overall knowledge of the equal sign tended to perform better on these less familiar types of problems.

Items testing advanced relational thinking. The advanced relational thinking items were intended to go beyond testing procedural proficiency in order to illustrate the link between equal-sign knowledge and whether or not children could express generalizations about properties of equality and the transformations they allow in natural language. Our six advanced relational thinking items were well aligned with the construct. All but one had infit statistics between 0.5 and 1.5 and all but one had an item total correlation above .2. Items testing for explicit verbal knowledge of properties of equality fit the Rasch model nearly as well as items more typically used to study knowledge of the equal sign.

As can be seen on the Wright map (Figure 2), these items were the most difficult items on the assessment. The two items in Figure 3 illustrate this point. Even high Level 4 children, who were expected to define the equal sign relationally 90% of the time, were expected to answer these items correctly only 83% and 57% of the time, respectively (see comparative difficulty levels in Table 3). As predicted by our construct map, it appears that articulating knowledge about the transformations that preserve equality represents an advance in equal-sign knowledge over interpreting the equal sign in a relational manner. These items help to distinguish among children who all successfully provided relational definitions of the equal sign.

Consider the explanations that children gave for these items. Some children could evaluate the truth of these equations via procedural routes, but nonetheless could not explain the logical shortcuts that proceed from relational thinking (and thus were scored as 0s). For instance, the two children highlighted in Figure 3 correctly



3. (ST5®) *Without adding* 89 + 44, can you tell if the number sentence below is true or false?

133 133

89 + 44 = 87 + 46

True False Can't tell without adding

How do you know?

89 + 44 = 133 87 + 46 = 133

7. (ST6) *Without subtracting* the 9, can you tell if the statement below is true or false?

76 + 45 = 121 is true.

Is 76 + 45 – 9 = 121 – 9 true or false?

True False Can't tell without subtracting

How do you know?

76 + 45 = 121 121 – 9 = 112

121 – 9 = 112

Figure 3. Children's use of the inefficient solve-and-compare strategy.

recognized that the expressions are equal but used resource-intensive solve-and-compare strategies to justify their answers. Although the solve-and-compare strategy does require that children realize—on some level—that the equal sign expresses the interchangeability of each side of an equation, this strategy is not very efficient and does not capture comparative relational thinking (Level 4).

Another example of such advanced knowledge of the equal sign is reflected by answers that demonstrate use of compensation strategies, as in Figure 5. In this example, the child has used the fact that 87 is two less than 89, which means that the addend coupled with the 87 must be two more than the one coupled with 89 in order to preserve equality. Carpenter et al. (2003) have argued that such use of compensation strategies instead of full calculation is an indicator of relational thinking: These strategies are reflective of the realization that the equal sign represents a relation between the two sides of the equation and that a relation among the numbers in the two expressions makes it unnecessary to carry out the calculations.

In summary, items that required explaining transformations that preserve equality fit our model of equal-sign knowledge as laid out in our construct map. The fit of these items, despite the fact that they are much less frequently used to measure knowledge of the equal sign, helps to illustrate an important feature of our method: It can be used to explore new potential items that measure knowledge of the equal sign. Kieran (1981) has previously suggested that failures on some more advanced mathematical items may stem from incorrect conceptions of the equal sign. As more candidate items are generated, our method can assess whether or not these new items rely, at least in part, on knowledge of the equal sign, and if so, where they fall in the hierarchy of difficulty.

DISCUSSION

Several decades of research have catalogued the difficulties that U.S. elementary school children have with understanding the equal sign as an indicator of mathematical equality (e.g., Alibali, 1999; Baroody & Ginsburg, 1983; Behr et al., 1980; Falkner et al., 1999; Jacobs et al., 2007; Li et al., 2008; Lindvall & Ibarra, 1980; Matthews & Rittle-Johnson, 2009; McNeil, 2007; Powell & Fuchs, 2010; Weaver, 1973). The current study sought to add resolution to that picture by placing previously incommensurable measurement items onto a single scale. Our findings reaffirmed our past findings that diverse items can indeed be integrated onto a single scale and provided further support for the validity of our assessment and of our construct map for knowledge of the equal sign. Moreover, because we used a new sample, our findings provide early evidence for the generalizability of our construct map to other children who have not received specialized instruction on the equal sign. Below, we discuss how our assessment helps expand our abilities to reveal children's knowledge of the equal sign, both in terms of describing variability and in terms of the construct's link to some basic algebra items. We then discuss some of the study's limitations and ways that the assessment might be improved.

Elaborating the Variability in Children's Knowledge

Our assessment augments the information gleaned from any single item type measuring equal-sign knowledge. First, our data clearly demonstrate that not all nonstandard equation formats are equally challenging. As detailed in our construct map, the more that an equation varies from the standard $a + b = c$ format, the more difficult it is likely to be. Difficulty increases as form changes from all operations

7. (ST6) *Without subtracting* the 9, can you tell if the statement below is true or false?

76 + 45 = 121 is true.

Is 76 + 45 – 9 = 121 – 9 true or false?

True | False | Can't tell without subtracting

How do you know?

again → 76 + 45 - 9 = 121 - 9

121 - 9 = 121 - 9

Figure 4. Recognition that performing the same operation on both sides preserves equality.

3. (ST5a) *Without adding* 89 + 44, can you tell if the number sentence below is true or false?

89 + 44 = 87 + 46

True | False | Can't tell without adding

How do you know?

because 89-87=2 + they add 2 to 44 so it is even

Figure 5. Use of compensation strategy demonstrates the ability to use equality-preserving transformations to reduce the need for computation.

on the left side to all operations on the right side, becoming most difficult when operations are included on both sides of the equal sign. Although the comparative difficulties of these problems have been suggested in several studies (e.g., Baroody & Ginsburg, 1983; Carpenter et al., 2003; Weaver, 1973), differences in performance according to item format have rarely been quantified. Our findings confirm the importance of equation format by assessing a wide range of item formats and revealing that the effects of item format were similar for open-equation-solving items and equation-structure items. They also indicate that differences in item difficulty can be quite substantial.

Second, we were able to compare the difficulties of different types of items all thought to tap understanding of the equal sign, because our assessment measures difficulty on an interval scale. Focusing across rows in a given column in Table 3 allows us to compare the differences in the probabilities of success on different items for children of a given ability level. The table helps to highlight some interesting differences. For instance, consider low Level 4 children. Table 3 clearly reveals that providing a relational definition of the equal sign is typically much more difficult than working with equations with operations on both sides, be it solving an open equation or accepting nonstandard formats as valid. The ability to accept and solve equations in nonstandard formats and the ability to give a relational definition of the equal sign lie substantially far apart on the scale of increasing knowledge of the equal sign.

This point illustrates a more general feature of our model in that it helps make clear exactly how wide the variability is among children who might otherwise be grouped together based on similar performance on a particular item. For example, even though 75% of our sample failed to define the equal sign relationally, we could easily detect differences within this group because of the nature of our assessment and measurement scale.

Link Between Knowledge of the Equal Sign and Algebra

Our assessment approach also allowed us to explicitly map some basic algebraic items onto items commonly accepted as measuring knowledge of the equal sign. Simple algebraic items involving letters as variables (e.g., $c + c + 4 = 16$) and advanced relational thinking items requiring explanation of equality-preserving transformations loaded heavily on our equal-sign knowledge construct (i.e., the psychometric properties of the items indicated that they fit well with the other items). Children with more advanced knowledge of the equal sign were more likely to solve both types of algebraic items correctly. Importantly, this effect for equal-sign knowledge prevailed even though these children presumably had no more experience with letters as variables or advanced relational thinking items than their peers in the same classrooms. These findings provide important empirical evidence to support claims that young children's knowledge of the equal sign supports algebraic thinking (Knuth et al., 2006; Lindvall & Ibarra, 1980; MacGregor & Stacey, 1997; Steinberg et al., 1991).

Focusing on equations involving letters as variables, analysis of performance suggests two trends. First, the use of letters as variables may have added difficulty when compared to items using blanks, even though they were logically similar. Second, it appears that equation format seems to have influenced difficulty with equations involving letters as variables in much the same way that it influenced difficulty for equations that involved no letters as variables, namely, equations with operations on both sides were more difficult for children to solve than those that involved operators on a single side only. The use of multiple instances of a given unknown, which is only possible with letters as variables, also increased item difficulty.

Performance data for advanced relational reasoning items added even more nuance to the picture. Providing verbal explanations for the equality-preserving transformations on equations using only numbers proved to be more demanding—sometimes much more so—than was solving equations using relatively unfamiliar letters as variables. This highlights the fact that procedural competence with mathematical equality can sometimes precede ability to articulate rules governing the domain (Rittle-Johnson, Siegler, & Alibali, 2001).

Limitations

Our assessment and construct map have been developed with two different samples of elementary school children. In both samples, schools were using traditional mathematics curricula that did not focus explicitly on the equal sign or on mathematical equality. Children's operational view of the equal sign is largely thought to result from repeated exposure to equations in a standard operations-equals-answer format (Alibali et al., 2007; Falkner et al., 1999; Li et al., 2008; McNeil & Alibali, 2005a; Seo & Ginsburg, 2003). Children exposed to multiple equation formats from early on infrequently develop an operational view of the equal sign, demonstrating more facility with the concept of equality earlier (Li et al., 2008). Thus, our construct map may apply only to children from similar educational backgrounds. For example, we might expect the construct map to look different for a population of Chinese elementary school students or for students using a curriculum focused on exposing children to varied equation formats (e.g., Wynroth's 1975 curriculum studied by Baroody & Ginsburg, 1983). With wider scale testing, we can investigate the extent to which our construct map generalizes across different student populations.

A second limitation of the current study is that it relies on children's written responses on a paper-and-pencil assessment. Children likely have knowledge that is not revealed on paper-and-pencil tests. For example, Jacobs et al. (2007) used a student interview to reveal whether children used relational thinking to solve some of the problems on the written test, including a prompt to encourage use of relational thinking. We incorporated these prompts for relational thinking in our written assessment, but children often will say more than they will write. Integrating items from a structured interview with the written assessment would help to reveal the impact of response format on performance.

Finally, our assessment does not systematically measure other constructs that may be required for the solution of any given item. Certainly, each individual item requires multiple types of knowledge or skills (e.g. addition skills, knowledge of commutativity). Our model, however, was not concerned with whether or not any individual item loaded on several constructs. It was instead designed to investigate whether the items in toto loaded heavily on the construct of equal-sign knowledge. Thus, our assessment and construct map have little to say about other constructs that may be involved for individual items on the assessment. Although these issues were beyond the scope of our project, future research could identify additional mathematics skills and how they relate to success on our measure.

The Power of the Method

Our final point is about method. Our intent was to build upon prior research in order to gain more leverage from the items typically used to measure knowledge of the equal sign. Our contribution, therefore, is first and foremost one of method. Science has historically been constrained by the limits of method (Kuhn, 1996). We have presented evidence that suggests that our current science regarding understanding of the equal sign—and by extension, of mathematical equality—has similarly been limited by method. The use of diverse items without a unifying metric does not allow us to take advantage of the full leverage that an integrated assessment affords.

Our use of IRT in the context of Wilson's construct-modeling approach provides a model for the application of a powerful psychometric method for designing measurements of mathematical constructs. Research in mathematics education is full of potential for practical uses of this approach. In our case, we chose to use the method due to the lack of an integrated assessment of equal-sign knowledge. In another case, Clements, Sarama, and Liu (2008) used the approach to construct an assessment for measuring mathematics ability in children aged 3–5. The authors found that existing measures such as the Woodcock-Johnson III had not been validated for children in this young age range. They consulted experts, developed a construct map, and designed items to measure the construct. In the final analysis, the construct-based approach holds potential to help fill gaps in the field wherever measures are wanting or nonexistent.

In summary, we demonstrated how our new assessment, built using a construct modeling approach, stands to enrich our knowledge of children's understanding of the equal sign—and by extension, the concept of mathematical equality. This approach allows us (a) to compare difficulty across item types, (b) to improve the resolution of our pictures of both item difficulty and children's proficiencies, and (c) to expand our assessment flexibly to include more advanced items that demand equal-sign knowledge. Moreover, it may eventually allow us to expand the difficulty level upward to find places at which more advanced students and adults falter in their understanding or activation of their knowledge of equal sign as an indicator of mathematical equality (see Kieran, 1981; MacGregor & Stacey, 1997; McNeil & Alibali, 2005a).

REFERENCES

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education.

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

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

Bailey, R., Day, R., Frey, P., Howard, A. C., Hutchens, D. T., McClain, K., . . . Willard, T. (2004). *Mathematics: Applications and concepts: Course 2.* New York, NY: Glencoe/McGraw-Hill.

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

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

Blanton, M., Levi, L., Crites, T., & Dougherty, B. J. (2011). *Developing essential understandings of algebraic thinking for teaching mathematics in grades 3–5* (R. M. Zbiek, Series Ed., & B. J. Dougherty, Vol. Ed.). Reston, VA: National Council of Teachers of Mathematics.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahway, NJ: Erlbaum.

Cajori, F. (1928). *A history of mathematical notations*. LaSalle, IL: Open Court.

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

Clement, J. (1982). Algebra word-problems solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education, 13*, 16–30. doi:10.2307/748434

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology, 28,* 457–482. doi:10.1080/01443410701777272

Cobb, P. (1987). An investigation of young children's academic arithmetic contexts. *Educational Studies in Mathematics, 18,* 109–124. doi:10.1007/BF00314722

De Corte, E., & Verschaffel, L. (1981). Children's solution processes in elementary arithmetic problems: Analysis and improvement. *Journal of Educational Psychology, 73,* 765–779. doi:10.1037/0022-0663.73.6.765

Saussure, F. de (1959). *Course in general linguistics* (C. Bally & A. Sechehaye with A. Riedlinger, Eds.; W. Baskin, Trans.). New York, NY: Mc-Graw-Hill.

Education Consumers Foundation. (2010). *Correcting TCAP grade inflation: What to expect in 2010.* Retrieved from http://www.education-consumers.org/TCAP_NAEP.htm

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

Ginsburg, H. (1977). *Children's arithmetic: The learning process*. New York, NY: D. Van Nostrand.

Hill, H. C., & Shih, J. C. (2009). Examining the quality of statistical mathematics education research. *Journal for Research in Mathematics Education, 40,* 241–250.

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.

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

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

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*, 297–312. doi:10.1207/s1532690xci2403_3

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer.

Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago, IL: University of Chicago 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 the United States. *Cognition and Instruction, 26*, 195–217. doi:10.1080/07370000801980845

Linacre, J. M. (2010). A user's guide to *Winsteps ministep Rasch-model computer programs*. Retrieved from http://www.winsteps.com/winman/index.htm?copyright.htm

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

MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation: 11–15. *Educational Studies in Mathematics, 33*, 1–19.

Maletsky, E. M., Andrews, A. G., Bennett, J. M., Burton, G. M., Luckie, L. A., McLeod, J. C., . . . Scheer, J. K. (2005). *Harcourt Math*. Orlando, FL: Harcourt.

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of experimental child psychology, 104,* 1–21. doi:10.1016/j.jecp.2008.08.004

McNeil, N. M. (2007). U-shaped development in math: 7-year-olds outperform 9-year-olds on equivalence problems. *Developmental Psychology, 43*, 687–694. doi: 10.1037/0012-1649.43.3.687

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

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

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

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing 4 = 2 + 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development, 82*, 1620–1633. doi:10.1111/j.1467-8624.2011.01622.x

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

Molina & Ambrose (2006) Reference to come. <This ref. is missing- see p 5.>

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

National Council of Teachers of Mathematics & Mathematical Sciences Education Board, & National Research Council. (1998). *The nature and role of algebra in the K–14 curriculum*: *Proceedings of a national symposium.* Washington, DC: National Academy Press.

National Research Council. (1998). *The nature and role of algebra in the K-14 curriculum*. Washington, DC: National Academy Press.

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

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development, 6*, 449–468. doi:10.1016/0885-2014(91)90049-J

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. doi:10.1037/a0018447

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: MESA Press. (Original work published 1960)

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

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189. doi:10.1037/0022-0663.91.1.175

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

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346–362. doi:10.1037/0022-0663.93.2.346

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. 161–188). Mahwah, NJ: Erlbaum.

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

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: NY: Oxford University Press.

Steinberg, R. M., Sleeman, D. H., & Ktorza, D. (1991). Algebra students' knowledge of equivalence of equations. *Journal for Research in Mathematics Education, 22*, 112–121.

U.S. Chamber of Commerce. (2011). *Leaders and laggards: A state-by-state report card on educational effectiveness*. Retrieved from http://www.uschamber.com/reportcard/2007

Warren, E. (2003). Young children's understanding of equals: A longitudinal study. In N. Pateman, G. Dougherty, & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 379–387). Honolulu, HI: PME.

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

Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research, 8*, 1–22.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus* (D. F. Pears & B. F. McGuiness, Trans.). London: Routledge.

Wittgenstein, L. (2001). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Malden, MA: Blackwell.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14*, 97–116. doi:10.1111/j.1745-3984.1977.tb00039.x

Wynroth, L. (1975). *Wynroth math program—the natural numbers sequence. Ithaca*, NY: Author.

Zhu, X., & Simon, H. A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166. doi:10.1207/s1532690xci0403_1

Authors

Percival Matthews, Department of Psychology, 208 Haggar Hall, University of Notre Dame, Notre Dame, IN 46556; pmatthew@nd.edu

Bethany Rittle-Johnson, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203; bethany.rittle-johnson@vanderbilt.edu

Katherine McEldoon, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203; k.mceldoon@vanderbilt.edu

Roger Taylor, 464 Mahar Hall, SUNY Oswego Oswego, NY 13126; roger.taylor@oswego.edu

Accepted November 15, 2011

APPENDIX A

Full[2] Assessment Form 2
Equation Structure Items

1. For each example, decide if the number sentence is true. In other words, does it make sense?

After each problem, circle True, False, or Don't Know.

Samples:

$3 + 4 = 7$	True	False	Don't Know
$3 + 4 = 12$	True	False	Don't Know
a) $5 + 3 = 8$	True	False	Don't Know
b) $3 = 3$	True	False	Don't Know
c) $5 + 5 = 5 + 6$	True	False	Don't Know
d) $31 + 16 = 16 + 31$	True	False	Don't Know
e) $7 + 6 = 6 + 6 + 1$	True	False	Don't Know
f) $6 = 6 + 0$	True	False	Don't Know

2. For each example, decide if the number sentence is true. Then, explain how you know.

a) $7 = 3 + 4$	True	False	Don't Know
b) $6 + 4 = 5 + 5$	True	False	Don't Know

3. *Without adding* 67 + 86, can you tell if the number sentence below is true or false? $67 + 86 = 68 + 85$. How do you know?

4. Find a number that can go in each box.
   $8 + 2 + = 10 + \square$
   b) Could another number go in the boxes? YES NO
   Explain why or why not.

5. $17 + 12 = 29$ is true.
   Is $17 + 12 + 8 = 29 + 8$ true or false? How do you know?

6. $2 \times 3 = 6$ is true.
   Is $2 \times 3 \times 4 = 6 \times 4$ true or false? How do you know?

---

[2] Some items, such as 1a, served as fillers, and were not included in the Rasch model. All scored items are listed in Appendix B.

7. *Without subtracting* the 7, can you tell if the number sentence below is true or false?
   56 + 85 = 141 is true.
   Is 56 + 85 – 7 = 141 – 7 true or false? How do you know?

8. Is the number that goes in the box the same number in the following two number sentences?
   $2 \times \square = 58$ $8 \times 2 \times \square = 8 \times 58$
   How do you know?

Equal Sign Items

9. What does the equal sign (=) mean? Can it mean anything else?

10. Which of these pairs of numbers is equal to 6 + 4? Circle your answer.
    (a) 5 + 5
    (b) 4 + 10
    (c) 1 + 2
    (d) none of the above

11. Which answer choice below would you put in the empty box to show that five cents is the same amount of money as one nickel? Circle your answer.
    5 cents □ One nickel
    a) 5¢
    b) =
    c) +
    d) don't know

12. Is this a good definition of the equal sign? Circle *good* or *not good*.

a. The equal sign means the same as.	Good	Not good
b. The equal sign means add.	Good	Not good
c. The equal sign means the answer to the problem.	Good	Not good

13. Which of the definitions above is the *best* definition of the equal sign? Write a, b, or c in the box below.

14. a) Is this statement true or false?

    1 dollar = 100 pennies

    b) What does this equal sign mean?

Equation Solving Items

DIRECTIONS: *Find the number that goes in each box.*

15. $3 + 4 = \square$

16. $4 + = \square$

17. $8 = 6 + \square$

18. $3 + 4 = \square$

19. $\square + 2 = 6 + 4$

DIRECTIONS: *On these problems, we really need you to show your math. Find the number that goes in each box.*

20. $7 + 6 + 4 = 7 + \square$

21. $8 + \square = 8 + 6 + 4$

22. $6 - 4 + 3 = \square + 3$

DIRECTIONS: *Find the number that goes in each box. You can try to find a shortcut so you don't have to do all the adding. Show your work and write your answer in the box.*

23. $898 + 13 = 896 + \square$

24. $43 + \square = 48 + 76$

25. Find the value of z. In other words, what value of z will make the following number sentence true? Circle your answer.
    $10 = z + 6$

26. Find the value of n.
    $n + n + n + 2 = 17$

27. Find the value of m.
    $m + m + m = m + 12$

APPENDIX B

Item Statistics for Mathematical Equivalence Assessment Form 1
(by Hypothesized Level, in Order of Increasing Difficulty)

Hypothesized level	Corresponding # from Appendix A	Item type	Item summary[a]	Expert rating	Item difficulty
1	16	Open equation	4 + □ = 8	2.75	−6.11
	1c	Structure	Judge "5 + 5 = 5 + 6" as true / false.	4.50	−3.81
	10	Definition	Identify pair that is equal to 6 + 4.	2.75	−2.71
2	1b	Structure	Judge "3 = 3" as true or false.	4.25	−0.55
	2a	Structure	Explain the judgment of "7 = 3 + 4" as true or false.	4.75	−1.27
	1f	Structure	Judge "6 = 6 + 0" as true or false.	5.00	−2.13
	11	Definition	5 cents □ 1 nickel. Select choice that shows they are the same.	2.50	−2.27
	17	Open equation	8 = 6 + □	3.75	−2.71
3	12a	Definition	"The equal sign means the same as." Is this a good or not good definition?	3.75	−0.12
	25	Open equation	10 = z + 6. Find the value of z.	4.33	−0.88
	1d	Structure	Judge "31 + 16 = 16 + 31" as true or false.	3.50	−0.80

[a] Wording of summaries does not duplicate actual wording verbatim. See Appendix A for exact wording of items from Form 1 of the assessment.

APPENDIX B (*Continued*)

Item Statistics for Mathematical Equivalence Assessment Form 1
(by Hypothesized Level, in Order of Increasing Difficulty)

Hypothesized level	Corresponding # from Appendix A	Item type	Item summary	Expert rating	Item difficulty
3	19	Open equation	□ + 2 = 6 + 4	5.00	−0.47
	1e	Structure	Judge "7 + 6 = 6 + 6 + 1" as true or false.	4.50	−0.47
	20	Open equation	7 + 6 + 4 = 7 + □	5.00	−0.31
	18	Open equation	3 + 4 = □ + 5	4.33	−0.23
	22	Open equation	6 – 4 + 3 = □ + 3	4.75	0.18
	2b	Structure	Explain the judgment of "6 + 4 = 5 + 5" as true or false.	5.00	0.18
	21	Open equation	8 + □ = 8 + 6 + 4	4.50	0.26
4	14	Definition	What does the equal sign mean in this statement? 1 dollar = 100 pennies	3.33	−0.15
	26	Open equation	$n + n + n + 2 = 17$. Find the value of n.	4.25	0.35
	13	Definition	Which definition of the equal sign is the best?	4.25	0.76
	23	Open equation	898 + 13 = 896 + □. Try to find a shortcut.	4.75	1.00
	24	Open equation	43 + □ = 48 + 76 Try to find a shortcut.	4.50	1.00
	9	Definition	What does the equal sign mean?	4.25	1.51

APPENDIX B (*Continued*)

Item Statistics for Mathematical Equivalence Assessment Form 1 (by Hypothesized Level, in Order of Increasing Difficulty)

Hypothesized level	Corresponding # from Appendix A	Item type	Item summary	Expert rating	Item difficulty
	27	Open equation	$m + m + m = m + 12$. Find the value of m.	4.00	2.67
	7	Structure	Explain without subtracting the 7. If $56 + 85 = 141$, does $56 + 85 - 7 = 141 - 7$?	5.00	2.76
	3	Structure	Explain the judgment of "67 + 86 = 68 + 85" as true or false without computing.	4.75	3.12
4	5	Structure	If $17 + 12 = 29$, does $17 + 12 + 8 = 29 + 8$? Explain.	5.00	3.12
	4	Structure	Find a number that can go in each box. $8 + 2 + \square = 10 + \square$ Could another number go in the boxes? Explain.	5.00	3.56
	6	Structure	If $2 \times 3 = 6$, does $2 \times 3 \times 4 = 6 \times 4$? Explain.	5.00	3.73
	8	Structure	Is the number that goes in the box the same number in the following two number sentences? Explain. $2 \times \square = 58$ $8 \times 2 \times \square = 8 \times 58$	5.00	6.55