SCREENING FOR MATHEMATICS DIFFICULTIES IN K–3 STUDENTS

Second Edition

CENTER ON INSTRUCTION

SCREENING FOR MATHEMATICS DIFFICULTIES IN K–3 STUDENTS

Second Edition

Russell Gersten
University of Oregon / Instructional Research Group

Benjamin S. Clarke
Pacific Institutes for Research

Kelly Haymond
Instructional Research Group

Nancy C. Jordan
University of Delaware

CENTER ON INSTRUCTION

This publication was created by the Center on Instruction, which is operated by RMC Research Corporation in partnership with the Florida Center for Reading Research at Florida State University; Instructional Research Group; Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; The Lawrence Hall of Science at University of California, Berkeley; and The Meadows Center for Preventing Educational Risk at The University of Texas at Austin.

The authors acknowledge the editorial and production support provided by Angela Penfold, C. Ralph Adler, and Robert Kozman of RMC Research Corporation.

The development of this document was supported in part by the National Center for Learning Disabilities and in part by the U.S. Department of Education, Office of Elementary and Secondary Education and Office of Special Education Programs, under cooperative agreement S283B050034A. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Preferred citation
Gersten, R., Clarke, B., Haymond, K., & Jordan, N. (2011). Portsmouth, NH: RMC Research Corporation, Center on Instruction.

To download a copy of this document, visit www.centeroninstruction.org.

TABLE OF CONTENTS

PREFACE

Since 2007, when this technical report was originally issued, the assessment field has made considerable progress in developing valid and reliable screening measures for early mathematics difficulties. This update includes new research published since 2007. It focuses on valid and reliable screening measures for students in kindergarten and first grade. However, we also examined data on screening tests for second and third grades because the goal of screening is to identify students who might struggle to learn mathematics during their initial school years.

INTRODUCTION

A major advance in the field of reading over the past 15 years has been the development and validation of screening measures that can detect, with reasonable accuracy, kindergartners and first graders likely to experience difficulty in learning to read. These students now receive additional instructional support during the critical early years of schooling. This is especially important because we know that most students who are weak readers at the end of first grade remain struggling readers throughout the elementary grades (Juel, 1988).

Similarly, studies in early mathematics have shown that students who complete kindergarten with weak knowledge of mathematics tend to experience consistent difficulties in that content area (Duncan et al., 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Morgan, Farkas, & Wu, 2009). In fact, using a nationally representative sample of students, Morgan et al. (2009) found that students who remained in the lowest 10th percentile at both the beginning and end of kindergarten (often considered an indicator of a learning disability in mathematics) had a 70% chance of remaining in the lowest 10th percentile five years later. They also tended to score, on average, two standard deviation units (48 percentile points) below students in the acceptable range of mathematics performance in kindergarten. Jordan et al. (2009) found that kindergarteners' number sense, knowledge of number relationships, and understanding of number concepts predict later mathematics achievement even when controlling statistically for intelligence quotient and socio-economic status.

Designing screening tools

Screening tools that identify students at risk for later mathematics difficulties must address predictive validity and content selection, among other variables. Specifically, the extent to which performance relates to later mathematics performance must be considered in the design of screening tools. For example, a student's score on a kindergarten screening measure should predict difficulty in mathematics at the end of first grade, second grade, and so on. Assessments that show evidence of predictive validity can inform instructional decision-making. Given evidence that predicts later failure, schools and teachers can allocate resources for instructional or intervention services early in regular classroom settings. Early intervention, which might simply entail small-group instruction that provides additional practice, explanation, and/or feedback, might suffice for students who are behind their peers in acquiring critical foundational skills. Instrument design must also be guided by findings from developmental and cognitive psychology on how children develop an emerging understanding of mathematics, and by mathematics educators' expertise. Effective screening tools integrate the knowledge bases of math education and developmental and cognitive psychology.

In this updated report, we describe the aspects of numerical proficiency that emerge consistently as the most important concepts to assess in young students. We also specify areas that seem most fruitful to assess in early screening batteries.

The role of number sense in mathematics development

The concept of number sense permeates the research on early development of numerical proficiency. Kalchman, Moss, and Case (2001) characterized number sense as:

> a) fluency in estimating and judging magnitude, b) ability to recognize unreasonable results, c) flexibility when mentally computing, [and] d) ability to move among different representations and to use the most appropriate representation (p. 2).

However, as Case (1998) noted, "number sense is difficult to define but easy to recognize" (p.1). Precise definitions of number sense remain controversial and elusive. Berch (2005) captured these complexities in his article *Making Sense of Number Sense: Implications for Children with Mathematical Disabilities*:

> Possessing number sense ostensibly permits one to achieve everything from understanding the meaning of numbers to developing strategies for solving complex math problems; from making simple magnitude comparisons to inventing procedures for conducting numerical operations; and from recognizing gross numerical errors to using quantitative methods for communicating, processing, and interpreting information (p. 334).

Berch compiled 30 possible components of number sense based on research from cognitive psychology, developmental psychology and educational research.[1] One recurrent component in all operational definitions of number sense is *magnitude comparison* ability (i.e., the ability to discern quickly the greatest number in a set, and to be able to weigh relative differences in magnitude efficiently—e.g., to know that 11 is a bit bigger than 9, but 18 is a lot bigger than 9). The ability to *decompose numbers in order to solve a problem* has also been cited frequently. For example, students with good number sense can solve 54 + 48 by first decomposing 48 to 4 tens and 8 ones, and then adding the 4 tens to 54 (64, 74, 84, 94), and the 8 ones to 94 to reach 102 (National Research Council, 2001).

Kalchman et al. (2001) more formally, and more forcefully, described number sense as "the presence of powerful organizing schemata that we refer to as *central conceptual structures*" (p. 2). They describe these structures as sets of mental number lines and demonstrate their importance for children's developing proficiency with mathematical procedures and understanding of mathematical concepts. Both Berch (2005) and Griffin, Case, and Siegler (1994) also noted that people who have good number sense seem to develop a mental number line on which they represent and manipulate numerical quantities. The development of a mental number line, therefore, facilitates the solving of a variety of mathematical problems.

Griffin et al. (1994) noted that children develop number sense in large part through formal and informal instruction by parents, siblings, or teachers, although genetic aspects are also clearly involved (Geary, 2004; Petrill, 2006),

Selected components of developing numerical proficiency

Magnitude comparison. As children develop a more sophisticated understanding of number and quantity, they can make more complex judgments about magnitude. For example, one preschooler may know that 9 is bigger than 3, while another will know that 9 is 6 greater than 3. Riley, Greeno, and Heller (1983, cited in National Research Council, 2001) found that, given a picture of five birds and one worm, most preschoolers were able to answer hypothetical questions such as, "Suppose the birds all race over and each one tries to get a worm. Will every bird get a worm?" Their answers demonstrate a gross magnitude judgment that there are more birds than worms. But given

[1] For a full list of possible components of number sense, see Berch (2005).

a specific question about magnitude, for example "How many birds won't get a worm?" (p.169), most preschoolers could not answer correctly. The ability to make more finite types of magnitude comparisons is a critical underpinning of the ability to calculate, as is some ability at mental calculations and an understanding of place value.

Almost all early screening tools use some measure of magnitude comparison. For example, many items in the Number Knowledge Test (Okamoto & Case, 1996) involve magnitude comparison. In Okamoto and Case's view, magnitude comparison is at the heart of number sense.

Using magnitude comparison in screening illustrates that screening instruments by nature are not designed to be comprehensive: a good screening instrument will be related to other critical aspects of performance. While a test may not measure mental calculation and place value directly, measures of magnitude comparison indicate likely performance in those areas. Traditional texts rarely teach magnitude comparison. However, Griffin et al. (1994) found that magnitude comparison is taught, informally but explicitly, in middle-income homes, but is rarely taught in low-income homes. They found that high-SES students entering kindergarten answered the magnitude comparison problems correctly 96% of the time, while low SES children, answered correctly 18% of the time.

Strategic counting. Counting efficiently and counting to solve problems are fundamental skills leading to mathematical understanding and proficiency (Siegler & Robinson, 1982). Geary (2004) noted that young students who use inefficient counting strategies are likely to have difficulty learning mathematics. Researchers typically differentiate between *knowledge of counting principles* and *skill in counting*. An example of a rudimentary counting principle is the realization that "changing the order of counting, or the perceptual appearance of an array, will not affect the quantity, whereas addition and subtraction of an object will affect the quantity" (Dowker, 2005, p.85). A second example is the knowledge that, given a group of 5 objects and a group of 3 objects, you can "count on" from 5 (i.e., count 6, 7, 8) to determine how many objects there are together. Young children often use a much less efficient approach: they count out 3 objects, then 5 objects, and then put them together and begin counting over from 1 to 8.

In most cases, competence in counting relates strongly to knowledge of counting principles (Dowker, 2005). Siegler (1987, 1988) studied the evolution of the min strategy in young children in depth. For example, a child who knows the min strategy, when asked "what is 9 more than 2," will automatically see the efficiency in reversing the problem to 2 more than 9, and simply "count on" from 9. Of course, grasping the min principle demonstrates a grasp of the commutative principle. Students with math difficulties or disabilities (MD) almost invariably use more immature and inefficient counting strategies to solve problems.

Although students should master sequence counting (reciting the counting words without reference to objects) in preschool, strategic counting is the more critical problem-solving math skill. For that reason, most researchers attempt to include a measure of strategic counting in their assessment batteries.

Geary (1990) examined the use of counting strategies by first graders with MD in comparison with their peers. Although both groups used similar strategies to solve problems, students with MD were three to four times more likely to make procedural errors. For example, when they counted on their fingers, they were incorrect half of the time, and when they used verbal counting strategies they were incorrect one third of the time. Some researchers assess counting skill and accuracy, although the ability to count strategically and effectively appears to be more foundational to future success in arithmetic. As students use more effective, efficient counting strategies to solve basic arithmetic combinations, they reinforce their conceptual understanding of important mathematical principles (e.g., commutativity and the associative law).

Retrieval of basic arithmetic facts. Early theoretical research on mathematics difficulties focused on *correlates* among students with a mathematics learning disability. Researchers (Goldman, Pellegrino, & Mertz, 1988; Hasselbring et al., 1988) consistently found that struggling elementary students could not retrieve addition and subtraction number combinations automatically. More recently, Geary (2004) found that struggling children typically fail to move from counting on their fingers (or with objects) to solving problems in their heads, without the need for manipulatives.

The research suggests that students with MD retain deficits in their retrieval of basic combinations, even though they often make strides in using algorithms and procedures and solving simple word problems when they receive instruction in these areas (Geary, 2004; Hanich & Jordan, 2001).
These deficiencies suggest underlying problems with what Geary calls *semantic memory* (i.e., the ability to store and retrieve abstract information efficiently),

an ability considered to be essential for succeeding in, and understanding, mathematics.

Word problems. *Adding It Up*, the National Research Panel's (2001) report on mathematics, concluded that, contrary to adults' perceptions, children find solving word problems easier than simple number sentences or simple equations. Jordan, Levine, and Huttenlocher (1994) found that before they begin receiving formal math instruction, young children can solve simple word problems involving addition and subtraction more easily than problems with number combinations—problems that do not refer to objects or provide context. Word problems have only recently been added to early screening batteries.

Numeral recognition: learning to link numerals with names. Numeral recognition is notoriously difficult in English compared to other languages. Some researchers suggest this may be a factor impeding the speed with which Americans learn mathematics.

While numeral recognition is not a mathematics skill *per se*, it serves as a gateway skill to formal mathematics, in the way that letter recognition leads to understanding the written code. Just as letter-naming accuracy and speed predict a child's ability to benefit from typical reading instruction, numeral recognition, measured in early screenings, may identify students with possible difficulties in mathematics. Numeral recognition may not be critical focus in mathematics instruction, but it can reveal potential risk for later failure in mathematics. Children begin to learn about the written symbol system for numerals before they enter school; an assessment of numeral recognition could be a valuable tool to identify at-risk students as they enter kindergarten.

The numbers that children encounter early in life *describe* things, like a home address or telephone number. In stark contrast, formal school settings emphasize the *cardinality* of numbers and their use in abstract computations. For example, figuring out how to solve a simple addition problem depends on a student's ability to recognize the number symbols and use other mathematical concepts such as cardinality, magnitude comparison, and counting.

Assessing number sense for early screening and identification—single and multiple proficiency measures

Single proficiency screening measures. Many researchers (e.g., Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Clarke & Shinn, 2004; Geary, 2004; Jordan, Kaplan, Olah, & Locuniak, 2006) have focused on developing single proficiency measures of discrete aspects of numerical aptitude. In some ways, this approach resembles one used by Kaminski & Good (Good, Gruba, & Kaminski, 2001) for assessing critical beginning reading skills using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), with separate tests for letter-naming fluency, initial sound identification, phoneme segmentation, and the reading of short pseudo-words.

Most of these single proficiency measures are fairly easy to administer and can be completed in a few minutes, usually because they are more focused, faster to administer, and can be used school- or district-wide with large numbers of students. Such measures can be used to quickly identify students whose mathematics achievement is either on track or at risk in one or more critical areas and prompt the provision of additional support. However, as with any screening, these measures merely indicate risk status; they cannot provide a full diagnostic profile. Diagnostic assessments are necessary to determine areas where a student needs additional help.

Multiple proficiency screening measures. In contrast to single proficiency measures, multiple proficiency measures comprise several aspects of number competence, including counting and skip counting, magnitude comparisons, simple arithmetic word problems, simple addition and subtraction, and estimation. Multiple proficiency measures usually provide a composite or total score rather than separate scores on individual skills. Although most of the research in this area is new, multiple proficiency measures appear as promising as single proficiency measures. The pattern of findings in Tables 1 through 5 in Appendix A show that the predictive validities of single proficiency measures are comparable to multiple proficiency measures—somewhat surprising, given that multiple proficiency measures cover a wider range of mathematics proficiencies and skills.

Empirical studies of single proficiency measures

This section summarizes several seminal pieces and samples more contemporary research on single proficiency measures. Tables 2–5 in Appendix A provide key details about each study. This section offers context to these tables and our recommendations in the final section. A thorough review of the literature can be found in Seethaler and Fuchs (2010) and Gersten et al. (2010). For readers interested in the technical information on this research, Appendix A lists the procedures used.

Clarke (2004, 2008). Clarke and Shinn (2004) used individually administered timed measures, each focused on one component of number sense. Fluency measures were designed with the intent to screen all kindergarten and/or first-grade students in a school. Brief fluency measures enable easy identification of the most at-risk kindergarten and first-grade students early in the school year; teachers can then provide interventions to prevent more serious mathematics problems in later grades.

Clarke and Shinn first tested three measures—number identification, quantity discrimination, and missing number—with first-grade students. Each measure was timed for one minute. The number identification measure required students to identify numerals between 1 and 20; the quantity discrimination measure required students to identify the bigger number from a pair of numbers between 1 and 20, and the missing number measure required students to identify a missing number from a sequence of three consecutive numbers in either the first, middle, or last position. The missing number measure functioned as a measure of strategic counting.

In 2008, Clarke, Baker, Smolkowski, and Chard extended the work to a kindergarten sample, only including numbers between 1 and 10, rather than 1 and 20. Predictive validities were high, ranging from .62 to .64 with a standardized achievement test.

Seethaler and Fuchs (2010). Seethaler and Fuchs (2010) examined the predictive validity of screening measures for risk of math difficulty (MD) in kindergartners. They administered a single proficiency measure, a magnitude comparison (Chard et al., 2005), and a multiple proficiency measure (Number Sense, created by the authors) in September and May to 196 kindergarten students. At the end of first grade, these students' conceptual (e.g., conceptual skills and mental manipulation of whole numbers) and procedural (e.g., the ability to identify and write numerical symbols and perform written calculations) outcomes were measured on The Early Math Diagnostic Assessment (EDMA) and the KeyMath-Revised (KM-R). The authors defined MD as scoring below the 16th percentile on the EDMA at the end of first grade. Comparisons of single and multiple proficiency screening measures, and between conceptual versus procedural outcomes, were conducted.[2] Interestingly, single and multiple proficiency screeners produced similar classification accuracy.

Mazzocco and Thompson (2005). In 2005, Mazzocco and Thompson[3] set out to find the best measure or set of measures to predict kindergartners' degree of risk for mathematics difficulty in third grade. They tracked 226 students from kindergarten through third grade on several measures such as visual-spatial, cognitive, and formal and informal mathematics achievement. Running a set of regression models, the authors found four specific items in the measures that predicted later mathematic difficulty (as evidenced by standard scores of below the 10th percentile on a comprehensive measure of third-grade mathematics). The four items were: reading numerals, number constancy (when observing number sets below 6), magnitude judgments, and mental addition of one-digit numbers. The four-item model successfully classified 84% of third-grade students as at-risk for mathematics difficulties based on their kindergarten performance on the four items.

VanDerHeyden et al. (2001). VanDerHeyden and colleagues[4] created a series of one-minute, group-administered measures to assess kindergarten students' mathematical proficiency. In the first measure, students counted a number of circles and wrote the numeral corresponding to the number of circles they had counted; a modification of this measure had students count the number of circles and then circle the corresponding number from a set of choices. The last measure had students draw the number of circles represented by a numeral they were shown. Predictive validity was examined in terms of how well the measures predicted retention at the end of kindergarten. Scores predicted retention correctly in 71.4% (5/7) of cases and correctly predicted non-retention in 94.4% (17/18) of cases. (It should be noted that predicting retention was based on the three mathematics probes and three reading readiness probes.) Concurrent validity correlations ranged from .44 to .61.

---

2 The researchers used logistic regression and receiver operating characteristics (ROC) analyses.

3 These authors report classification accuracy but do not present predictive validity coefficients, so they do not appear with the other studies reported in Tables 1 through 5 in Appendix A.

4 These authors use math and reading screeners and examine predictive validity in terms of how well the measures predicted retention at the end of kindergarten, not performance on a math outcome measure, so they do not appear with the other studies reported in Tables 1–5 in Appendix A.

This was the first study in the field of school psychology to report the sensitivity and specificity of mathematics screening measures, a step beyond simply using predictive validity. Contemporary screening research uses increasingly complex statistical procedures to evaluate the sensitivity and specificity of a screening measure (e.g. Bryant et al., 2008; Geary, Bailey, & Hoard, 2009; Gersten et al., 2010; Jordan, Glutting, Ramineni, & Watkins, 2010; Seethaler & Fuchs, 2010).

Summary. These studies reveal an emerging picture of critical aspects of measuring early numerical proficiency. First, many measures assess different, discrete skills, with varying degrees of success. The fact that screening for different components of number sense can produce acceptable results further reinforces the multi-faceted nature of numerical proficiency, even at the kindergarten and first-grade levels. Second, strategic counting and magnitude comparison emerged as two key constructs to measure.

Empirical studies of multiple proficiency measures

The Number Knowledge Test (NKT). The Number Knowledge Test (Okamoto & Case, 1996) is an individually administered 10–15 minute measure that assesses students' procedural and conceptual knowledge related to whole numbers. The test examines students' understanding of magnitude, their counting ability, and their competence with basic arithmetic operations.

As the name implies, the NKT focuses exclusively on the domain of number, but unlike single proficiency measures which assess discrete skills and abilities in numerical proficiency, the NKT assesses multiple facets of a student's numerical proficiency, including the application of number to basic arithmetic concepts and operations. The measure has four levels of increasing difficulty and deeper analysis. For example, the NKT includes problems to assess a child's ability to make magnitude comparisons; these problems increase in complexity as the child advances through the levels of difficulty. The magnitude comparison questions explore a child's understanding of magnitude, the word "bigger," and whether a child understands that traditional counting goes from smaller to larger numbers. Figure 1 presents sample items from the Test of Number Knowledge.

Figure 1

Number Knowledge Test	Example items
Level 0	1. Here are some circles and triangles. Count just the triangles and tell me how many there are.
Level 1	1. If you had 4 chocolates and someone gave you 3 more, how many chocolates would you have? 2. Which is bigger: 5 or 4?
Level 2	1. Which is bigger, 19 or 21? 2. What number comes 4 numbers before 17?
Level 3	1. What number comes 9 numbers after 999? 2. Which difference is smaller, the difference between 48 and 36 or the difference between 84 and 73?

When Baker et al. (2002) and Gersten, Jordan & Flojo (2005) administered the Number Knowledge Test in kindergarten to predict subsequent performance a year after the test was given, it demonstrated significant predictive validity correlations of .73 to the SAT-9 Total Mathematics score administered to students one year later, at the end of first grade. The NKT was a strong predictor of performance on both the Procedures (r=.64) and the Problem Solving (r=.69) subtests.

Jordan et al. (2008). Jordan, Glutting, and Ramineni (2008) developed the Number Sense Brief (NSB), a multi-component number sense battery. The 33-item untimed measure takes approximately 15 minutes to administer. It assesses counting, one-to-one correspondence, number recognition, and nonverbal addition and subtraction. The correlation between student performance on the number sense battery at the beginning of kindergarten with math achievement at the end of third grade was .63.

Jordan's group has consistently studied the link between mathematics and reading disabilities and found that beginning reading skill (as well as overall IQ) strongly predicted later mathematics performance and that the NSB added a significant proportion to the explained variance. That is, early number sense predicts later math achievement, *over and above* reading skill and general cognitive competencies.

CONCLUSION

Research on early mathematics screening was in its infancy when we first wrote this report in 2006. Since then, a wave of early screening studies (e.g., Baglici, Codding, & Tryon, 2010; Bryant et al., 2008; Clarke et al., 2008; Clarke, Gersten, Dimino, & Rolfhus, in press; Jordan, Glutting, & Ramineni, 2008; Lembke & Foegen, 2009; Methe, Hintze, & Floyd, 2008; Seethaler & Fuchs, 2010; VanDerHayden, 2011) has contributed to an emerging knowledge base that permits us to draw conclusions that can guide practice in the field.

Recurring findings from many studies demonstrate that significant mathematical developmental differences exist between students in kindergarten and first grade and, more importantly, those differences can be pinpointed accurately with brief and relatively easy-to-use screening tools. While differences observed in young children may result from exposure to mathematics before formal schooling or from student performance on more formal mathematics in school, screening young children on each component of number sense offers a critical link to instruction and additional instructional services.

The mathematics curriculum changes year to year, and it is possible that certain students may initially learn math at acceptable levels only to experience problems once the content becomes more abstract (e.g., with the introduction of decimals, improper fractions, ratios and proportions, negative numbers). Therefore, as in reading (Scarborough, 2001), we will likely see some students whose mathematics performance may be acceptable in the primary grades but will deteriorate in later grades (Geary, 1993).

The research reviewed in this publication addresses early predictors of mathematics difficulty; it does not necessarily help us understand which students will succeed in math in the early elementary grades but struggle with more intricate and abstract topics such as those involving rational number (i.e., fractions, ratio, proportion) or geometry in fourth and fifth grade. We call for more longitudinal studies to answer these questions and address student learning of more advanced math topics.

In addition, the research we reviewed also supports the importance of working memory (Desoete, Ceulemans, Roeyers, & Huylebroeck, 2009; Geary, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Swanson & Beebe-Frankenberger, 2004) in understanding mathematical proficiency at many different levels. However, few researchers have explored instructional methods for enhancing students' working memory in mathematics. As our understanding of mathematical development advances, so should the design of screening instruments that reflect the complexity of mathematics. As outcome measures become more mathematically sophisticated following the guidelines of the Common Core State Standards (CCSS) and other contemporary state standards, we will likely learn more about longer term predictors of subsequent success.

At present, however, we have sound means for assessing which five- and six-year olds are likely to encounter serious difficulties later in learning mathematics. Each research effort reviewed here assesses some aspect of predictive validity across one or several school years, either by examining correlations over time or, less frequently, by examining student classifications. The strength of predicting later math difficulties varies, but recent research demonstrates that to some extent earlier difficulty in mathematics may underpin struggles with later mathematical achievement.

A small but growing body of research (e.g., Bryant et al., 2008; Fuchs et al., 2005; Fuchs & Karns, 2001; Griffin, Case, & Siegler, 1994) suggests that early intervention in kindergarten and first grade can produce real benefits. As more research focuses on interventions for students identified as at risk, our understanding of the relationship between deficits in foundational skills and later performance will be enriched.

Although progress has been made in terms of understanding what constitutes a multiple proficiency assessment (e.g., Jordan, Glutting, & Ramineni, 2008; Seethaler & Fuchs, 2010), the components of an *efficient* multiple proficiency assessment battery remain unclear. In part, decisions about what works best may be guided by a max-min standard. That is, how can we gain the maximum amount of information in the minimum amount of time? Brief measures of magnitude comparison and strategic counting appear to be important elements. Measures of working memory may well add to a battery's predictive power, but they may be less sensitive to change than the other measures because working memory is less likely to be a focus of instruction.

Future research should attempt to determine the advantages and disadvantages of timed measures. We sense that timed measures may, in many instances, be more potent than untimed screening measures. It may also be true that screening all kindergarten or first-grade students might require timed measures to enhance data collection efficiency.

Finally, while the link between assessment and instruction in early mathematics is neither fully known nor articulated, efforts should continue to develop tools that are compatible with the principles of Response to Intervention (RTI) as defined in the reauthorization of the Individuals with Disabilities Education Act (2004). That such features (Fuchs, Fuchs, & Prentice, 2004) are often present in screening tools does not automatically guarantee their usefulness in the problem-solving and formative assessment phases of RTI. Future research should focus on the role of effective assessment tools within RTI decision-making criteria.

Despite the scarcity of research in early mathematics, strides have been made in recent years to explore critical questions in the assessment and instruction of early mathematics. We hope this publication encourages and energizes researchers to take on the remaining questions in the field. Also, we hope educators will take the findings outlined in this publication to heart, recognize the potential for earlier identification of children with math difficulties, and use the techniques described here to start children on a path to math proficiency as early as possible.

REFERENCES

Baglici, S. P., Codding, R., & Tryon, G. (2010). Extending the research on the tests of early numeracy: Longitudinal analyses over two school years. *Assessment for Effective Intervention, 35*(2), 89–102.

Baker, S., Gersten, R., Flojo, J., Katz, R., Chard, D., & Clarke, B. (2002). *Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays* (Technical Rep. No. 0305). Eugene, OR: Pacific Institutes for Research.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38,* 333–339.

Bramlett, R. K., Rowell, R. K., & Mandenberg, K. (2000). Predicting first grade achievement from kindergarten screening measures: A comparison of child and family predictors. *Research in the Schools, 7,* 1–9.

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. (2008). Mathematics intervention for first and second grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*(1), 20–32.

Case, R. (1998). A psychological model of number sense and its development. Annual meeting of the American Educational Research Association, San Diego, CA.

Chard, D., Clarke, B., Baker, S. K., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment Issues in Special Education, 30,* 3–14.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29,* 46–57.

Clarke, B., Gersten, R., Dimino, J., & Rolfhus, E. (in press). Assessing student proficiency in early number sense (working title). Longmont, CO: Cambium Learning Sopris.

Clarke, B., & Shinn, M. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33,* 234–248.

Desoete, A., Ceulemans, A., Roeyers, H., & Huylebroeck, A. (2009). Subitizing or counting as possible screening variables for learning disabilities in mathematics education or learning? *Educational Research Review, 4*(1), 55–66.

Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience, and education.* New York: Psychology Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Fuchs, L. Compton, D., Fuchs, D., Paulsen, K., Bryant, J., & Hamlett, C. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73,* 311–330.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37,* 292–306.

Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal, 101,* 495–510.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 33,* 386–404.

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345–362.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment, 27,* 265–279.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78,* 1343–1359.

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2010). *Universal screening in mathematics for students in the primary grades.* Manuscript submitted for publication.

Gersten, R., Jordan, N.C., & Flojo, J. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38,* 293–304.

Goldman, S.R., Pellegrino, J.W., & Mertz, D.L. (1988). Extended practice of basic addition facts: Strategy changes in learning disabled students. *Cognition and Instruction, 5,* 223–265.

Good, R. H., Gruba, J., & Kaminski (2001). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 679–700). Washington, DC: National Association of School Psychologist.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning or arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.

Hanich, L., & Jordan, N. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology, 93,* 615–626.

Hasselbring, T., Sherwood, R., Bransford, J., Fleenor, K., Griffith, D., & Goin, L. (1988). Evaluation of a level-one instructional videodisc program. *Journal of Educational Technology Systems, 16,* 151–169.

Jordan, N. C., Glutting, J., & Ramineni, C. (2008). A number sense assessment tool for identifying children at risk for mathematical difficulties. In A. Dowker (Ed.), *Mathematical difficulties: Psychology and intervention* (pp.45–58). San Diego, CA: Academic Press.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review, 39,* 181–195.

Jordan, N., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77,* 153–175.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45,* 850–867.

Jordan, N., Levine, S., & Huttenlocher, J. (1994). Development of calculation abilities in middle and low income children after formal instruction in school. *Journal of Applied Developmental Psychology, 15,* 223–240.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for the development of mathematical understanding: Rational numbers and functions. In S. Carver & D. Klahr (Eds.), *Cognition and Instruction* (pp.1–38). Mahwah, NJ: Lawrence Erlbaum Associates.

Lembke, E. S., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and grade 1 students. *Learning Disabilities Research & Practice, 24,* 12–20.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41,* 451–459.

Mazzocco, M. & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research and Practice, 20,* 142–155.

Methe, S. A., Hintze, J. M., & Floyd, R. G. (2008). Development and validation of early numeracy skill indicators. *School Psychology Review, 37,* 359–373.

Morgan, P., L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42,* 306–321.

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Panel. (2001). *Eager to learn: Educating our preschoolers.* M.S. Donovan & M.S. Burns (Eds.). Washington, DC: National Academy Press.

Okamoto, Y. and Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61,* 27–58. doi: 10.1111/j.1540–5834.1996.tb00536.x

Petrill, S. (February, 2006). Mathematics disability: Definition, etiology, and comorbidity. Paper presented at the 14th Annual PCRC (Pacific Coast Research Conference), Coronado, CA.

Scarborough, H. (2001). Connecting early language and literacy to later reading (dis) abilities. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford.

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*(1), 37–60.

Siegler, R. (1987). The perils of averaging data over strategies: The example of children's addition. *Journal of Experimental Psychology: General, 116,* 260–264.

Siegler, R. (1988). Individual differences in strategy choices: good students, not-so-good students, and perfectionists. *Child Development, 59,* 833–851.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in Child Development and Behavior* (pp. 241–311). New York: Academic Press.

Swanson, H. L., & Beebe-Frankenberger, M. E. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96,* 471–491.

VanDerHeyden, A. M. (2011). Technical adequacy of Response to Intervention decisions. *Exceptional Children, 77*(3), 335–350.

VanDerHeyden, A. M., Witt, J. C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30*(3), 363–82.

APPENDIX A

Summary of the Evidence Base on Early Screening Measures as of December 2010

Table 1

Measures of magnitude comparison

Study	Screening measure[a]	Grade	n[b]	Outcome measure	Predictive validity[c]
Baglici et al. (2010)	Name the larger of two items: number sets 0 to 20	K	61	Timed mathematics computation	.02 (ns)
Chard et al. (2005)	Name the larger of two items: number sets 0 to 20	K 1st	436 483	Number Knowledge Test	.50 .53
Clarke et al. (2008)	Name the larger of two items: number sets 0 to 10	K	254	Stanford Early School Achievement Test	.62
Clarke & Shinn (2004)	Name the larger of two items: number sets 0 to 20	1st	52 348	Woodcock-Johnson Applied Problems Timed computation	.79 .70
Clarke et al. (in press)	Name the larger of two items: number sets 0 to 20 for K and 0 to 99 for 1st	K 1st	323 348	Terra Nova	.49 .62
Lembke & Foegen (2009)	Name the larger of two items: number sets 0 to 10 and 0 to 20 (i.e., 13:8)	K 1st	44 28	Test of Early Mathematics Ability–3	.35 .43
Seethaler & Fuchs (2010)	Name the larger of two items: number sets 0 to 10	K	196	Early Math Diagnostic Assessment: Math Reasoning Numerical Operations Key Math–Revised: Numeration Estimation	 .53 .75 .34 .65

Note: All coefficients p < .05 unless noted otherwise.

[a] All measures were timed.

[b] All study samples were from a single district except for Lembke & Foegen (2009), which sampled three districts in two states, and Clarke et al. (in press), which sampled four districts in two states.

[c] All predictive validity measured screeners administered in the fall and mathematics outcomes administered in the spring of that same year. Although Seethaler & Fuchs (2010) calculated two predictive validity coefficients, only the coefficients from fall and spring of kindergarten were used in this table.

Table 2

Measures of strategic counting

Study	Screening measure[a]	Grade	n[b]	Outcome measure	Predictive validity[c]
Baglici et al. (2010)	Name the missing number in a string of numbers between 0 and 20	K	61	Timed mathematics computation	.47
Clarke et al. (2008)	Name the missing number in a string of numbers between 0 and 10	K	254	Stanford Early School Achievement Test	.64
Clarke & Shinn (2004)	Name the missing number in a string of numbers between 0 and 20	1st	52	Woodcock-Johnson Applied Problems Math computation probes	.72 .67
Clarke et al. (in press)	Name the missing number in a string of numbers between 0 and 20 for K and 0 and 99 for 1st	K 1st	323 348	Terra Nova	.48 .55
Lembke & Foegen (2009)	Name the missing numbers in a pattern: counting by ones to 20, by fives to 50, and by tens to 100 (i.e., 6 _ 8 9). Items are the same for K and 1st grade	K 1st	44 28	Test of Early Mathematics Ability–3	.37 .68
Methe et al. (2008)	Students “count on” four numbers from a given number between 1 and 20 (e.g., experimenter says 8 and student says 9, 10, 11)	K	64	Test of Early Mathematics Ability-3	.46

Note: All coefficients p < .05 unless noted otherwise.

[a] All measures were timed.

[b] All study samples were from a single district except for Lembke & Foegen (2009), which sampled three districts in two states, and Clarke et al. (in press), which sampled four districts in two states.

[c] All predictive validity measured screeners administered in the fall and mathematics outcomes administered in the spring of that same year.

Table 3

Fact retrieval

Study	Screening measure[a]	Grade	n[b]	Outcome measure	Predictive validity[c]
Bryant et al. (2008)	TEMI: addition/ subtraction (sums or minuends range from 0 to 18)	1st	126	Stanford Achievement Test–10	.55
Clarke et al. (in press)	Basic facts: Students are presented 40 problems that can be composed and decomposed in base-10 system	1st	329	Terra Nova	.50

Notes: All coefficients $p < .05$ unless noted otherwise.

[a] All measures were timed.

[b] All study samples were from a single district except for Clarke et al. (in press), which sampled four districts in two states.

[c] All predictive validity measured screeners administered in the fall and mathematics outcomes administered in the spring of that same year.

Table 4

Exploratory measures: word problems as a reasonable long-term predictor

Study	Screening measure[a]	Grade	n[b]	Outcome measure	Grade Outcome	Predictive validity[c]
Locuniak & Jordan (2008)	Eight-item story problems with four addition and four subtraction story problems	K	198	Calculation fluency	Middle of 2nd	.51

Note: All coefficients $p < .05$ unless noted otherwise.

a Untimed measure.

b Study samples were from a single district.

c Correlated the fall of kindergarten screening measure with criterion measures administered in the winter of 2nd grade.

Table 5

Multiple number proficiency tests

Study	Screening measure[a]	Grade	n	Outcome measure	Grade Outcome	Predictive validity[b] (r)
Baker et al. (2002)	Number Knowledge Test:	K	64	Stanford Achievement Test–9	1st	.73
Jordan et al. (2008)	Number Sense Brief: 33 items assessing counting, one-to-one correspondence, number recognition, nonverbal addition and subtraction	K	200	Woodcock-Johnson–III	3rd	.63
Seethaler & Fuchs (2010)	Number Sense: 30 items	K	196	Early Math Diagnostic Assessment: – Math Reasoning – Numerical Operations Key Math–Revised: – Numeration – Estimation	K	 .56 .62 .40 .74

Notes: All coefficients $p < .05$.

[a] All measures were timed except Number Sense and Number Knowledge Test.

[b] Although Seethaler & Fuchs (2010) calculated two predictive validity coefficients, only the fall and spring of kindergarten were used in this table.

APPENDIX B

Procedure for Reviewing the Literature on Early Screening in Mathematics

The authors conducted a literature search using the ERIC and PSYCHINFO databases with the descriptors screening and mathematics, and limiting the search to empirical studies published between 1996 and 2011 and to studies involving children ranging in age from birth to 12 years old. Dissertations were excluded. We also conducted a manual search of major journals in special, remedial, and elementary education (*Journal of Special Education, Exceptional Children, Journal of Educational Psychology*, and *Journal of Learning Disabilities*) to locate relevant studies.

This search resulted in the identification of 47 studies. Of this total, 19 studies were selected for further review based on analysis of the title, keywords, and abstracts. Of these 19 studies, 13 met our criteria for inclusion. Of the 13 studies identified, 10 focused on single proficiency measures and 3 studies on multiple proficiency measures.[5]

Our criteria for inclusion limited our review to studies that targeted kindergarten and first-grade students, included screening measures and outcome variables specific to mathematics performance, and reported predictive validity. Additionally, we focused on single proficiency studies that provided correlations between screeners administered in the fall and mathematics outcomes administered in the spring of that same year. We excluded several studies (Bramlett, Rowell, & Mandenberg, 2000; Fuchs et al., 2007; Jordan, Kaplan, Locuniak, & Ramineni, 2007) in which more than 12 months passed before outcomes were assessed. For the data on single proficiency measures, presented in Tables 1–4, we thought it best to compare measures across a similar time frame.

Similarly, studies that included winter-to-spring predictive validity coefficients, were omitted from Table 1 in order to allow for meaningful comparisons across measures. We did not, however, apply this criterion to studies of composite or multiple proficiency measures, which used longer, varying time frames (listed in Table 5 along with the actual time frame). We also excluded studies that used one or more norm-referenced standardized measures as a screener because we were interested in an efficient screener or screening batteries. Many of the standardized measures are much longer than we would recommend for a screener, often taking between one and three hours.

[5]One study, Seethaler and Fuchs (2010) used both a single proficiency and a multiple proficiency measure

Description of the data presented in the tables

We limited the data presented in Tables 1–4 to predictions from fall to spring so that the reader can make meaningful comparisons between measures. We have found that earlier compilations of the literature mixed studies looking at concurrent and predictive validity together, and merged studies looking at prediction over three months with those examining predictive validity over a three-year period.

To eliminate this problem, we only present fall-spring predictive validity for single proficiency measures. We do present longer-term predictive validity coefficients for the longer, multiple competency measures because most of the studies reported data over a longer time frame. Therefore, the evidence in Table 5 is not easily or quickly compared with the evidence in the other tables. It is always more difficult to predict over longer periods because more uncontrolled events transpire. All things being equal, we would expect the correlations in Table 5 to be lower than those in other tables. As will be seen, this is usually not the case.

Tables 1–4 list the key proficiencies and results from selected single proficiency measures. In order to allow the reader to home in on salient features of the evidence base, we organized the following sections and tables around five key constructs that recur in the literature—magnitude comparison, strategic counting, retrieval of basic arithmetic facts, word problems, and numeral recognition. Each proficiency/number sense component has its own table so the reader can obtain a sense of how robust the proficiency component is as a screener, identify the grades and number of children covered by the screener in each study, and select the measure used to assess predictive validity. With the exception of Clarke et al. (in press) and Lembke and Foegen (2009), many studies included in the tables suffer from a limitation in that the research was conducted in a single school district. All measures were timed with the exception of the Story Problems measure (See Table 4) used by Locuniak and Jordan (2008).

CENTER ON
INSTRUCTION