

A Brief, Multiple-Choice Assessment of Mature Number Sense Is Strongly Correlated With More Resource-Intensive Measures

Patrick K. Kirkland¹ , Claire Guang² , Chineme Otuonye³, Nicole M. McNeil³ 

[1] Institute for Educational Initiatives, University of Notre Dame, Notre Dame, IN, USA. [2] Department of Psychology, University of Chicago, Chicago, IL, USA.

[3] Department of Psychology, University of Notre Dame, Notre Dame, IN, USA.

Journal of Numerical Cognition, 2024, Vol. 10, Article e12679, <https://doi.org/10.5964/jnc.12679>

Received: 2023-09-13 • Accepted: 2024-01-25 • Published (VoR): 2024-03-15

Handling Editor: Joonkoo Park, University of Massachusetts Amherst, Amherst, MA, USA

Corresponding Author: Patrick K. Kirkland, 107 Carole Sandner Hall, Notre Dame, IN 46556, USA. E-mail: pkirklan@nd.edu

Supplementary Materials: Materials [see [Index of Supplementary Materials](#)]



Abstract

Students who exhibit mature number sense make sense of numbers and operations, use reasoning to notice patterns, and flexibly choose effective problem-solving strategies (McIntosh et al., 1997, <https://ro.ecu.edu.au/ecuworks/6819>). Due to its dispositional nature, mature number sense is typically measured through in-depth interviews or tests of strategy usage. Yet, the lack of an efficient, rigorously developed measure has made it difficult to collect systematic, replicable evidence on students' mature number sense. To address this, we developed a brief assessment of mature number sense. The present study provides additional convergent evidence of validity for this measure with US students in grades 3–8 (8–14 years old). We compared middle school ($N = 40$) and upper elementary school ($N = 41$) scores from the brief assessment with an established, time-intensive measure (Yang, 2019, <https://doi.org/10.1007/s10649-018-9874-8>) and an in-depth interview of student strategy usage (Markovits & Sowder, 1994, <https://doi.org/10.2307/749290>). We found strong correlations ($r > 0.7$) across all three measures, and this held even when controlling for students' arithmetic scores ($pr > 0.6$). Researchers and educators can now use the brief assessment to investigate students' mathematical thinking and advance knowledge of a key aspect of mathematical cognition.

Keywords

'mature' number sense, assessment, convergent validity, rational numbers, problem-solving strategies, fluency

Highlights

- Upper elementary and middle school students' scores on a brief assessment of number sense were strongly correlated with their performance on a strategy-based test and one-on-one interview.
- Students' performance on the brief assessment of mature number sense was strongly related to, but distinct from, their performance on a timed test of single-digit arithmetic facts.
- The brief assessment can be used to further study mature number sense and the interventions designed to improve students' number sense.

Improving children's "mature number sense" is a central goal of mathematics education reforms in the United States (Common Core State Standards [CCSS], 2010; Kilpatrick et al., 2001; National Council of Teachers of Mathematics [NCTM], 2000, 2014). Students with mature number sense use reasoning to notice patterns, make sense of numbers



and operations, and flexibly select the most effective and efficient problem-solving strategies (McIntosh et al., 1997; R. Reys et al., 1999; Yang, 2005). One goal of mathematics educators is to cultivate these ways of thinking, so students can develop the habit of mind to persistently ask, “Does this make sense?” when encountering mathematical problems (CCSS, 2010; NCTM, 2000; Kilpatrick et al., 2001).

Some have argued that, due to its dispositional nature, mature number sense may be an aspect of mathematical cognition that can only be measured through resource- and time-intensive probes of student thinking, such as in-depth, one-on-one interviews (e.g., Howden, 1989) or lengthy tests of strategy usage (Yang, 2019). However, through work with teachers interested in assessing and supporting their students’ number sense, we identified a need for a practical, yet rigorously developed, measure that could be used in both classroom and research settings. To address this, we developed a brief assessment of mature number sense (Kirkland et al., 2024). Here we examined how upper elementary (3rd–5th grade) and middle school (6th–8th grade) US students’ scores on this assessment compare with scores on an established, time-intensive measure (Yang, 2019) and with strategy use during an in-depth interview of student thinking.

Mature Number Sense

In the United States, “number sense” has emerged as a central goal of mathematics education over the last three decades. In the late 1980’s, government commissioned reports (e.g., National Research Council, 1989) and national curriculum frameworks (e.g., NCTM, 1989) highlighted “number sense” as a core objective of K-12 mathematics education. Since then, various mathematics education researchers have aimed to characterize how “number sense” can be manifested (e.g., Greeno, 1991; Howden, 1989; Markovits & Sowder, 1994; McIntosh et al., 1992; B. J. Reys, 1991). Through various iterations, the focus has remained on whether or not students exhibit the disposition to *make sense of numerical situations*.

Beyond mathematics education, however, researchers across the disciplines of cognitive and developmental psychology, neuroscience, and special education have operationalized “number sense” in a variety of ways (e.g., Berch, 2005; Dehaene, 2001; Gersten et al., 2005; Jordan et al., 2006). In this journal, for example, “number sense” has been used to describe the Approximate Number System or “non-symbolic number sense” (Marinova & Reynvoet, 2020; Reynvoet et al., 2021), irrational number magnitude understanding (Obersteiner & Hofreiter, 2017), and a “sense of numerosity” (dos Santos, 2023). In a systematic review of the number sense literature, Whitacre, Henning, and Atabas (2020) identified a case of polysemy with three distinct constructs all labeled as “number sense”, which they identify as: (a) *approximate number sense* (e.g., Dehaene, 2001) (b) *early number sense* (e.g., Jordan et al., 2006, 2009), and (c) *mature number sense* (e.g., Markovits & Sowder, 1994; McIntosh et al., 1992, 1997). *Approximate number sense*, most commonly studied in numerical cognition literature, examines visual and auditory perception of number and magnitude discrimination, especially the Approximate Number System. Research on *early number sense*, sometimes referred to as “numeracy” (USAID, 2018), focuses on symbolic number recognition, number relations, counting, cardinality, and simple number operations (Jordan et al., 2006). Early number sense is strongly related to students’ early school performance (Jordan et al., 2009). Research on *mature number sense*, most commonly studied by mathematics education researchers, examines flexibility with rational numbers, representations, and operations (Markovits & Sowder, 1994).

Therefore, per Whitacre et al.’s (2020) recommendation, we include the term “mature” to distinguish our construct of interest from approximate number sense (e.g., Dehaene, 2001) and early numeracy (e.g., Jordan et al., 2009). Importantly, this is not to imply that maturity in number sense naturally develops with age or that this construct is just a more mature version of the approximate number system (ANS). It is a theoretically distinct construct from ANS. Rather, we use the term “number sense” because that is what the construct is known as in mathematics education, and we aim for our research to be interpretable across disciplinary lines¹. We define *mature number sense* in line with McIntosh et al.’s (1992) definition of number sense as “a person’s general understanding of number and operations along with the ability and inclination to use this understanding in flexible ways to make mathematical judgments.”

¹ One could argue that to avoid confusion, the construct should instead be called *numerical reasoning* or *mathematical sense-making*. While we agree that these terms align broadly with our research focus, we have chosen to keep *mature number sense* as the name to align our work with well-known mathematics education research on *number sense* (e.g., Markovits & Sowder, 1994).

Measuring Mature Number Sense

In-Depth Interviews of Number Sense Strategy Use

Researchers seeking to characterize students' mature number sense have most commonly assessed students at a single time point, using either structured interviews or other researcher-developed assessments (Alsawaie, 2012; Gay & Aichele, 1997; Nejem & Muhanna, 2013; Purnomo, 2014; R. Reys et al., 1999; Yang, 2005; Yang & Chang, 2023; Yang & Li, 2008). For interview studies, the protocol has most often involved presenting students with non-traditional items designed to assess components of a hypothesized number sense framework (e.g., Alsawaie, 2012; Yang, 2005). Students are asked to solve the item first, and then explain in detail how they came to the specific answer. Experimenters are then tasked with clarifying student strategies using specific probes such as "Please justify your answer." or "Can you do it another way?". Student strategies are coded for whether they are number sense based, algorithmic or rule-based, or fall in a general other category. Researchers then analyze how often students used number sense-based strategies when correctly answering an item. Oftentimes, use of the researchers' hypothesized number sense strategy is the sole determinant of whether or not students are credited with displaying number sense, regardless of solution correctness.

As an example, one problem designed by Yang (2005) to assess students' estimation skills was: "Without calculating an exact answer, circle the best estimate for $72 \div 0.025$." Students were given the following multiple choice answer options: a) A lot less than 72, b) A little less than 72, c) A little more than 72, and d) A lot more than 72. After students shared their answer with the researcher, they were prompted to justify or explain their thinking. Even though 12 of the 21 6th-grade students in this study correctly chose "d) A lot more than 72," only 2 of those 12 students used strategies that were coded as "number sense based." An example of a number sense based strategy is "The result will become larger when it is divided by a number less than 1. Since 0.025 is quite a bit smaller than 1, the result of 72 divided by 0.025 will become very much larger." In contrast, a larger number of students used "rule-based methods" such as, "(0.025) has three decimals, so we must add three zeros to 72; hence, $72 \div 0.025 = 72,000 \div 25$. Since 72,000 is a very large number, the answer should be a lot bigger than 72 when divided by 25." This is taken as evidence of students relying on using the "written method" to solve the problem and lacking the necessary mature number sense to solve it otherwise.

Tests of Mature Number Sense

In hopes of measuring mature number sense in a greater number of students, other teams of researchers developed multiple-choice tests using items similar to those from the interviews (e.g., Kaminski, 2002; McIntosh et al., 1997; R. Reys et al., 1999; Yang & Lin, 2015). These items are often similar to some of the original National Assessment of Educational Progress (NAEP) questions in mathematics (e.g., Carpenter et al., 1981, 1988) such as estimating fraction sums like $\frac{12}{13} + \frac{7}{8}$ or using the properties of operations to rewrite equivalent expressions. The original versions were paper and pencil assessments where students were encouraged to turn the page after a certain amount of time, usually 30-45 seconds (Hsu et al., 2001; Markovits & Sowder, 1994; McIntosh et al., 1997; Menon, 2004). Researchers have frequently suggested a time limit to discourage students from using rule-based strategies. For example, if students had paper and pencil and the time to use a common denominator strategy to add $\frac{12}{13} + \frac{7}{8} = \frac{96}{104} + \frac{91}{104} = \frac{187}{104} = 1\frac{83}{104}$, this would not reflect an understanding of using benchmark fractions as the item intended to assess. To better address this issue, more recent measures have been electronic with an automatic time limit (e.g., Li & Yang, 2010). Overall, the items are often dissimilar from those in traditional grade-level curriculum and are mapped onto the researchers' hypothesized components of mature number sense.

The most frequently used tests in the literature were developed first by an international group of researchers (e.g., McIntosh et al., 1997) and then built upon by Yang and colleagues. Over the last two decades, Yang and colleagues have extensively studied the mature number sense of students in Taiwan using these tests (e.g., Hsu et al., 2001; Li & Yang, 2010; Reys & Yang, 1998). This research group was the first to develop a computer-based assessment of mature number sense with prompts for students to select the reason for their answer (Yang & Li, 2008). The research group hypothesized that doing so would make scores from multiple choice tests more valid, providing evidence about misunderstandings even when students correctly answered items. Similar to the analyses from in-depth interviews, students were given additional credit for selecting the corresponding "number sense based" reason. The reasons provided were eventually tailored to a specific answer selection (Yang & Lin, 2015). That is, for a correct answer option, there would

be a rule-based reason, a number sense based reason, and a common misconception provided. For an incorrect choice, however, only common misconceptions were provided. Throughout this literature, Yang and colleagues consistently report strong psychometric evidence of the scores from the tests (e.g., Li & Yang, 2010).

At the time of the current project's design, the most recent version (Yang, 2019) was a three-tier number sense test used with fifth and sixth grade students. The measure is administered electronically over two 40-minute class periods, with students solving 20 items per session. For each item, students are required to choose an answer, a reason option tailored on their selected answer, and a confidence rating for both their reason and answer. Therefore, in total, students provide answers to 40 items, reasons for their 40 answers, and confidence ratings for these 40 answers and ratings over an 80-minute time period.

A Brief Assessment of Mature Number Sense

After reviewing Yang's test (2019) and others (McIntosh et al., 1997), we chose to develop a new, brief assessment of mature number sense in hopes of providing a practical, efficient assessment that could be used to evaluate the impact of number-sense-focused instructional practices and advance understanding of mature number sense as a construct. Interestingly, despite the growing interest in new practices to improve students' mature number sense in the United States (e.g., Number Talks [Parrish, 2011], Number Sense Routines [Shumway, 2018]), none have been evaluated using any of the existing mature number sense tests. We reasoned that long administration times might be a contributing factor.

Therefore, building from this prior work and in consultation with expert mathematics teachers, teacher educators, and mathematics education researchers, we specified a four-component framework of mature number sense for 3rd-8th grade (8-14 years old) students (Kirkland et al., 2024) as follows:

1. *Understanding basic number concepts and number magnitude:* Strong mature number sense is characterized by a rich conceptual understanding of fractions, decimals, and whole numbers. Students use their understanding of place value and rational number magnitude to efficiently estimate results using concepts such as unit fractions.
2. *Using multiple representations of a number:* Strong mature number sense is characterized by proficiency in translating among multiple representations of rational numbers efficiently and flexibly to solve problems. Students use this understanding to translate between representations such as Arabic numerals and the number line.
3. *Understanding the effect of arithmetic operations on numbers:* Strong mature number sense is characterized by recognizing how the four core arithmetic operations affect whole numbers as well as fractions and decimals. Students demonstrate understanding that patterns observed with operations with whole numbers may not hold true for numbers between 0 and 1. Students use this understanding to efficiently estimate computational results and ensure the results make sense.
4. *Understanding mathematical equivalence:* Strong mature number sense is characterized by understanding the equal sign as a relational symbol, reflecting that the two sides of an equation are equal. Rather than approaching equations with an "operational" approach, students recognize patterns across the equal sign and use this relational thinking to flexibly solve problems (c.f. Jacobs et al., 2007).

We hypothesized, similar to Yang (2019), that mature number sense would be an overarching hierarchical latent construct, with each component theoretically related to mature number sense. That is, a student's understanding of mathematical equivalence reflects their mature number sense as well as a more specific relational understanding of the equal sign. In our initial development of a brief assessment of mature number sense for middle school students (Kirkland et al., 2024) the structure of the response data reflected this hypothesized framework. A bifactor model with each component as a specific factor best fit the response data over theoretically related models. We then created an upper elementary form of the brief assessment, and the response data in the elementary validation study reflected the same structure (Kirkland et al., 2023).

Present Studies

In our initial development (Kirkland et al., 2024), we did not compare the brief assessment with any other existing measures of mature number sense. However, some have argued that mature number sense is an aspect of mathematical

cognition that can only be measured through such resource- and time-intensive probes of student thinking as in-depth one-on-one interviews or lengthy tests of strategy usage (Yang, 2019). For example, Howden (1989) argues that mature number sense's "intuitional" nature makes it impossible to assess with "traditional evaluation techniques." Therefore, it is important to compare student performance across multiple formats to ensure we are measuring the same construct.

As discussed above, the tests by Yang and colleagues are the most widely used in the literature and have the strongest evidence of psychometric properties. Yang and colleagues have collected a wide range of evidence for the validity of the scores from their measure through expert teacher interviews, student interviews during problem solving, and other classical test theory analyses (e.g., Li & Yang, 2010). In addition, scores reflect both correct answers as well as identified "number-sense strategies" (Yang, 2019). Given its wide use and inclusion of strategy responses, it is important to show that the brief assessment measures the same overall construct as the Yang test (Kirkland, 2022).

Several prior studies have combined the use of standardized items with in-depth interviews (Gay & Aichele, 1997; Markovits & Sowder, 1994; Pike & Forrester, 1997; Yang, 2003; Yang et al., 2004; Yang & Wu, 2010). However, they often did so with the same items (e.g., Gay & Aichele, 1997) or with interview items focusing on separate components of mature number sense than test items (e.g., Markovits & Sowder, 1994). Yang's (2003) analysis of a classroom number sense intervention is the one example in the literature where the interview items and test items were different, and both separately designed to measure the overall construct of mature number sense. However, in that case, there was no comparison of individual performance across interview and paper-and-pencil test scores. It remains unclear if scores from standardized multiple-choice assessments and in-depth interview protocols are consistent.

The brief assessment of mature number sense is a timed, multiple-choice measure. One might argue that it is measuring "speed of retrieval" of mathematical knowledge and not the level of reasoning reflected in mature number sense. Therefore, we felt it was important to distinguish mature number sense from performance on a timed arithmetic test (Kirkland, 2022). High performance on a timed arithmetic test does not require flexibility of strategy use or conceptual understanding of rational numbers or operations. However, students' performance on different tests of mathematical performance are often highly correlated and interdependent (Kilpatrick et al., 2001), and there is evidence that speed in solving arithmetic facts would be strongly related to conceptual mathematics knowledge. Fuchs et al. (2016) found fact automaticity at 2nd grade predictive of fourth-grade algebra readiness and McGlaughlin et al. (2005) found college students' fact automaticity significantly predicted their success in college algebra. Theoretically, researchers have argued that higher-order problem solving is directly related to lower-level knowledge and the speed of accessing that knowledge (e.g., Anderson, 2002; Haverty et al., 2000). We have some initial evidence that scores on the brief assessment are related to, but distinct from middle school students' addition fact test scores ($r = 0.59$, $p < .001$). It is important to show a similar pattern is observed for student scores on more time-intensive measures. Furthermore, fluently recalling multiplication facts is a key standard in 3rd grade in the United States ("By the end of Grade 3, know from memory all products of two one-digit numbers." CCSS.MATH.CONTENT.3.OA.C.7). However, we are not aware of any study examining students' mature number sense and performance on a multiplication fact test. Therefore, it is unclear how much overlap there may be between students' proficiency at solving multiplication facts and their mature number sense.

To address these concerns, we aimed to compare students' scores on the brief assessment (Kirkland et al., 2024) with their performance during an in-depth one-on-one interview (e.g., Markovits & Sowder, 1994), a time-intensive test of strategy usage (Yang, 2019), and an arithmetic fact test. Our research question and associated hypotheses were the following:

Question: Does comparing students' scores on the brief assessment of mature number sense (Kirkland et al., 2024) with those on an existing measure of mature number sense (Yang, 2019) and a rubric from an in-depth interview of student thinking (e.g., Markovits & Sowder, 1994) provide additional evidence of convergent validity for the brief assessment?

Hypothesis 1: Assuming the three measures evaluate the same underlying construct of mature number sense, we predict high correlations among the brief assessment, interview rubric, and time-intensive test across both 6th-8th grade students and 3rd-5th grade students.

Hypothesis 2: We hypothesize a moderate correlation between scores on these measures and scores on an arithmetic fact test.

Hypothesis 3: We hypothesize that the high correlations among the measures will persist even after controlling for students' performance on an arithmetic fact test.

Method

Participants

We recruited 6th-8th grade students (11-14 years old) and 3rd-5th grade (8-11 years old) from schools surrounding a university in the midwestern United States. The goal was not to conduct a formal comparison of the grade levels, but rather to obtain evidence of validity in two different samples with the two forms of the brief assessment. The samples were recruited in August 2021 (middle schoolers) and August 2022 (elementary). To recruit, we sent invitations home through both school and community partners. In addition, we advertised the study on our research lab's social media, through direct contact to our research lab's email list, and through the University's weekly listserv. Based on our initial power analysis, we aimed to recruit 40 students each for the elementary and middle school studies. Due to strong responses to recruitment, we ended up recruiting 41 middle school and 43 elementary students. One middle school student and two elementary students were not able to return for session two, leaving 40 middle school and 41 elementary school participants with complete data. Students' self-identified demographics are summarized in Table 1 below.

Table 1

Student Demographics

| | Middle School | Elementary |
|-------------------------------|---------------|------------|
| Variable | N (%) | N (%) |
| Grade Level | | |
| 3 rd Grade | – | 14 (34%) |
| 4 th Grade | – | 7 (17%) |
| 5 th Grade | – | 20 (49%) |
| 6 th Grade | 15 (38%) | – |
| 7 th Grade | 11 (28%) | – |
| 8 th Grade | 14 (35%) | – |
| Gender | | |
| Male | 17 (43%) | 21 (51%) |
| Female | 21 (53%) | 19 (46%) |
| Prefer Not to Say | 2 (5%) | 1 (2%) |
| Race/Ethnicity | | |
| Asian | 1 (3%) | – |
| Black or African American | 1 (3%) | 2 (5%) |
| Hispanic or Latino | 1 (3%) | 4 (10%) |
| White | 30 (75%) | 32 (78%) |
| Two or More Races/Ethnicities | 6 (15%) | 3 (7%) |
| Other Races/Ethnicities | 1 (3%) | – |
| Total | 40 | 41 |

Measures

Participants completed the following measures across two sessions in the study:

Brief Assessment of Mature Number Sense

The Brief Assessment of Mature Number Sense (Kirkland et al., 2024) is an electronic, 24 item multiple-choice test of students' mature number sense, aligned with our theoretical framework detailed above. Items differ from a traditional curriculum and are designed to specially assess students' number sense. To discourage the use of traditional algorithms, each item has a 60-second time limit and students are not allowed to use paper and pencil. The total sum score is used in the analyses. There are separate forms for middle school and upper elementary to account for content differences across grade levels. However, the forms share eight common items to allow for vertically scaled scoring across elementary and middle school when needed. Figure 1 includes an example common item for each of the four components in our framework. In the validation analysis (Kirkland et al., 2023, 2024), student scores were reliable over time ($r = 0.83$ for MS, $r = 0.84$ for ELEM) and had evidence from expert reviews, factor analyses, student retrospective think-alouds, and item response theory analyses to support our validity argument.

Figure 1

Sample Items Assessing Mature Number Sense Organized by Component

| Understanding the Effect of Operations on Numbers | Number Concepts and Magnitude |
|--|--|
| <p>What is the best estimate for $26 \times \frac{3}{4}$?</p> <p> <input type="radio"/> Less than 26 <input type="radio"/> Equal to 26 <input type="radio"/> Greater than 26 <input type="radio"/> Can't know </p> | <p>Which fraction represents the largest amount?</p> <p> <input type="radio"/> $\frac{5}{6}$ <input type="radio"/> $\frac{5}{7}$ <input type="radio"/> $\frac{5}{8}$ <input type="radio"/> $\frac{5}{9}$ </p> |
| Multiple Representations of a Number | Mathematical Equivalence |
| <p>What is true about the number $\frac{2}{5}$ (two-fifths)?</p> <p> <input type="radio"/> It is greater than $\frac{1}{2}$ (one half). <input type="radio"/> It is the same as 2.5. <input type="radio"/> It is equivalent to 0.4 <input type="radio"/> It is less than $\frac{1}{6}$ (one sixth). </p> | <p>$47 + 16 + \underline{\quad} = 16 + 8 + 47$</p> <p> <input type="radio"/> 8 <input type="radio"/> 63 <input type="radio"/> 71 <input type="radio"/> 4716 </p> |

Three-Tier Number Sense Test

The Three-Tier Number Sense Test (TTNST, Yang, 2019) was the latest version of the tests developed by Yang and colleagues at the time of this project's design. The three tiers are: number sense item, reason for answer choice, and confidence rating. The reasons provided differ based on the students' selected answer to the item. The TTNST includes 40 multiple choice items balanced across Yang's (2019) hypothesized five components of mature number sense: understanding the basic meaning of numbers and operations, recognizing the number size, understanding multiple representations of numbers and operations, recognizing the relative effects of operations on numbers, and judging the reasonableness of computational results. The test is administered electronically over two sessions. For each item, students have 40 seconds to choose an answer, 60 seconds to choose a reason, and 20 seconds to choose a confidence rating for their answer and reason. Students' scores are calculated by factoring in both their answers and reasons (Yang,

2019). For each correct answer, students earn 4 points. For the reason options, the “number-sense based” reason is worth 4 points, a rule-based reason is worth 2 points, a misconception reason is worth 1 point, and a guessing reason is worth 0. Students who choose the wrong answer do not earn any points for their reason choice. Therefore, students can earn 0, 4, 5, 6, or 8 points on each of the 40 items and total scores can range from 0 to 320. Yang provides evidence for the reliability and validity of the TTNST scores from internal consistency analysis, expert item review, student interviews, and factor analyses.

Scripted Mature Number Sense Interview

Participants completed a structured interview where they solved 14 items designed to assess mature number sense, balanced across our four components. The items were not the same as those included in either of the other two measures. For middle school students, items were chosen from [Markovits and Sowder's intervention study \(1994\)](#). For elementary students, we adapted nine of the middle school items to be grade-level appropriate. For example, the middle school version “Are there decimals between 0.74 and 0.75?” was adapted to “Are there any numbers between 2 and 3?” We intentionally kept 5 items common across middle school and elementary in the interview to mimic the common linking items in the brief assessment. Students were prompted to first solve the item mentally within 60 seconds, and then to “Explain in your own words how you got that answer.” After students explained their first strategy, interviewers always asked students “Is there any other way you could have solved the problem?” before moving on to the next item. There was no time limit for student explanations. The interview sessions were audio recorded and transcribed for coding.

Coders used a structured rubric to evaluate students' mature number sense for each item based on both their answers and reasoning. The rubric had a 4-point scale: 1) Beginning, 2) Approaching, 3) Meeting, and 4) Exceeding. In the coding manual, example strategies were provided for each category and item. These strategies were based on established mature number sense interview protocols (e.g., [Markovits & Sowder, 1994](#); [Yang, 2005](#)) and rewarded student flexibility in strategy choice. For example, when asked about decimals between 0.74 and 0.75, a student answering “there are no numbers between because 0.75 is the next one up” would be coded as “beginning.” A student answering “yes there are, but we can't count how many. There are infinite of them, like 0.741” would be coded as “exceeding.” Students received a score for the number sense displayed while answering each item. The average score across all items was used as the variable of interest.

Interview coders were not aware of student performance on the other assessments. For all participants, a second coder independently scored all responses. Any disagreements were re-examined by both coders until agreement was reached.

For the middle school interview, there was initial exact agreement on 85% (478 of the 560 possible codes) of responses, and after discussion between coders, there were only 2 responses (0.4%) that required a third-party mediation from the primary author. For elementary, the initial agreement was 94% (541 of the 574 possible) and there were 0 responses that required third-party mediation. To further consider the reliability of the scale ([Koo & Li, 2016](#); [Syed & Nelson, 2015](#)), we used a one-way random effects model analyzing absolute agreement for multiple raters with a single value. The mean ICC across the 14 items reflected moderate to strong reliability of the rubric scores for middle school ($M = 0.93$, range: 0.69 – 0.98) and strong reliability for elementary ($M = 0.95$, range: 0.83 – 1.00). In addition, we calculated the ICC for total scores across raters, and this also reflected strong reliability of the interview scores for both middle school (ICC = 0.988, 95% CI [0.977, 0.993]) and elementary (ICC = 0.997, 95% CI [0.994, 0.998]).

Single-Digit Addition Fact Test

Used with middle school students, this measure ([Geary et al., 1996](#)) includes all possible combinations of randomly presented single-digit addition facts with the numbers 1-9. Students were tasked with solving as many as they could in 1 minute. The order was predetermined randomly and then kept standard for all participants. Students' total number of correct answers was the score of interest.

Multiplication Fact Test

After receiving feedback from our advisory board, we replaced the addition fact test with a multiplication fact test (Burns et al., 2015; Nelson et al., 2013) for our work with upper elementary students due to multiplication's primacy in US elementary mathematics curricular standards. The task followed the same procedure as the addition task, except with multiplication as the operation.

Procedure

Students participated after school hours in a university research lab over two sessions scheduled about a week apart. Days between sessions ranged from 1 to 26, with a median of 7 days for both samples. Each session lasted approximately 45-60 minutes. In session 1, students completed the brief assessment of mature number sense, the first half of the TTNST (Yang, 2019), and the addition or multiplication fact test, in that order. In session 2, students completed a math beliefs survey (outside of the scope of this manuscript), the second half of the TTNST, and the structured mature number sense interview, in that order. The order was kept as fixed so that students were not exposed to measures requesting the reasoning for their answer (e.g., TTNST or interview) before taking the brief assessment. The interview was chosen to be last so that any potential carryover effects from experimenter probes did not impact scores on the other measures. All measures were administered electronically on a Chromebook with an experimenter present. The interview was audio recorded and transcribed later for analysis.

Analysis

To examine how the measures used in this study relate, we ran a series of correlational analyses. We first calculated the zero-order correlations across the four measures. We predicted that the brief assessment would correlate highly ($r > 0.50$) with the Yang (2019) TTNST, and this association would be stronger than their respective correlations with addition or multiplication fact test scores. In addition, we expected students' interview scores to correlate highly with student performance on both the brief assessment and the Yang (2019) measure. Furthermore, we expected these associations to be stronger than the correlation between interview scores and addition or multiplication fact test scores.

We also ran a correlation analysis of students' total scores on the brief assessment with students' total scores from the TTNST and the scripted interview after partialling out students' addition or multiplication fact test scores. We tested the resulting Pearson correlation (denoted as pr for partial correlation) for significance using the standard t distribution and confidence interval. In both cases, we predicted the two number sense measures would be significantly related, even after controlling for addition or multiplication fact test scores.

Results

Middle School Students' Mature Number Sense

Middle school students solved on average 16.8 (70%) items correctly on the brief assessment of mature number sense ($SD = 3.82$). These scores were higher than those observed in the original validation study ($M = 56\%$ correct) and a recent longitudinal study ($M = 60\%$ correct in Kirkland et al., 2022). Across the two sessions, students had an average score of 168 ($SD = 46.2$) on the Three-Tier Number Sense Test (Yang, 2019). Similar to the brief assessment, students in this sample outperformed the larger validation sample (Yang), where students solved on average 45.4% of the items correctly, compared to 61.1% correctly in this study.

Students had an average score of 2.26 ($SD = 0.57$) on the mature number sense interview (sum score $M = 31.6$, $SD = 7.95$). Thus, students on average were rated as just above "Approaching" and not quite at "Meeting" on the rubric. While this is lower than expected based on the reported performance on the two tests above, the individual items chosen for the interview may have been more difficult than those used in either other measure, $M = 7.28$ correct (52%), $SD = 2.59$. We examine this more in the discussion below. For the addition fact test, students on average solved 24.1 ($SD = 6.7$) single-digit addition problems correctly in one minute. The descriptive statistics on the measures for middle school students are summarized in Table 2.

Table 2*Correlation Between Measures in Middle School Study*

| Measure | <i>M</i> | <i>SD</i> | 1 | 2 | 3 | 4 |
|-----------------------|-------------|-----------|------|------|------|---|
| 1. Brief Assessment | 16.82 (70%) | 3.82 | – | | | |
| 2. Yang's TTNST | 168 (61%) | 46.20 | 0.71 | – | | |
| 3. Scripted Interview | 2.26 (52%) | 0.57 | 0.79 | 0.73 | – | |
| 4. Addition Fact Test | 24.10 | 6.66 | 0.53 | 0.53 | 0.58 | – |

To examine the association between the measures of mature number sense, we first calculated the zero-order correlations of all measures (Table 2). As predicted, students' mature number sense scores on the brief assessment correlated strongly with their scores from both Yang's TTNST, $r(38) = 0.71$, $p < .001$, and the scripted interview, $r(38) = 0.79$, $p < .001$. Student scores on the scripted interview also correlated highly with their performance on the TTNST, $r(38) = 0.73$, $p < .001$. As predicted, the correlation between students' brief assessment and addition fact test scores, $r(38) = 0.53$, $p < .001$, was weaker than its correlations with the other measures of mature number sense. This was true for the TTNST, $r(38) = 0.53$, $p < .001$, and the scripted interview as well, $r(38) = 0.58$, $p < .001$.

As predicted, even after controlling for students' addition fact test scores, students' scores on the brief assessment of mature number sense significantly correlated with their scores on the TTNST, $pr = 0.60$, $t(37) = 4.59$, $p < .001$, and the scripted interview, $pr = 0.70$, $t(37) = 5.91$, $p < .001$. The same pattern was true when looking at interview scores and scores on the Yang measure, $pr = 0.61$, $t(37) = 4.72$, $p < .001$. We see here consistent evidence that scores from a brief assessment correlate strongly with those from Yang's TTNST as well as those from a scripted interview, even after controlling for scores on another measure of mathematics performance. In addition, a similar pattern is observed between the other two measures of mature number sense.

Upper Elementary School Students' Mature Number Sense

The 3rd–5th grade students solved on average 14.8 (62%) items correctly on the brief assessment ($SD = 5.2$). These scores were higher than those observed in the original validation study ($M = 46\%$ correct). Students had an average score of 110.8 ($SD = 57.4$) on Yang's Three-Tier Number Sense Test (2019), which was significantly lower than the middle school students, $t(79) = 4.95$, $p < .001$. Students had an average sum score of 30.4 ($SD = 9.5$) and mean score of 2.17 ($SD = 0.68$) on the mature number sense interview. Similar to the middle school sample, students performed worse on the interview ($M = 48\%$ correct). For the multiplication fact test, students on average solved 13.2 ($SD = 8.7$) single-digit multiplication problems correctly in one minute. The descriptive statistics on the measures for elementary students are summarized in Table 3.

Table 3*Correlation Between Measures in Elementary Study*

| Measure | <i>M</i> | <i>SD</i> | 1 | 2 | 3 | 4 |
|-----------------------------|-------------|-----------|------|------|------|---|
| 1. Brief Assessment | 14.83 (62%) | 5.20 | – | | | |
| 2. Yang's TTNST | 110.8 (43%) | 57.35 | 0.88 | – | | |
| 3. Scripted Interview | 2.17 (48%) | 0.68 | 0.89 | 0.85 | – | |
| 4. Multiplication Fact Test | 13.24 | 8.65 | 0.77 | 0.74 | 0.70 | – |

We then calculated the zero-order correlations of all measures in the study (Table 3). Once again, students' scores on the brief assessment correlated strongly with their scores from both Yang's TTNST, $r(39) = 0.88$, $p < .001$, and the scripted interview, $r(39) = 0.89$, $p < .001$. Student interview scores also correlated highly with their TTNST performance, $r(39) = 0.85$, $p < .001$. As predicted, the correlation between students' brief assessment and multiplication fact scores, $r(39)$

$= 0.77$, $p < .001$, was weaker than its association with the other measures of mature number sense. This was true for the TTNST, $r(39) = 0.74$, $p < .001$, and the scripted interview as well, $r(39) = 0.70$, $p < .001$. Interestingly, the association between upper elementary students' multiplication fact test scores and mature number sense was descriptively higher across all measures (r 's = 0.77, 0.74, 0.70) than middle school students' addition fact test scores and mature number sense (r 's = 0.53, 0.53, 0.58). We examine this more in the discussion below.

As predicted, even after controlling for students' multiplication fact test scores, students' scores on the brief assessment highly correlated with their TTNST scores, $pr = 0.71$, $t(38) = 6.25$, $p < .001$, as well as their interview scores, $pr = 0.77$, $t(38) = 7.51$, $p < .001$. The same pattern was true when looking at interview scores and scores on Yang's TTNST, $pr = 0.69$, $t(38) = 5.83$, $p < .001$. For elementary students', we see a consistent pattern of scores on the brief assessment strongly correlating with performance on an in-depth interview of student thinking and a time-intensive test of strategy usage.

Student Use of Number Sense Strategies

A key objection to using brief assessments is their inability to adequately capture the depth of students' conceptual understanding of mathematics. This criticism is frequently directed at multiple-choice formats, which fail to directly assess the strategies used to solve problems. To examine the validity of such an argument, we calculated students' total correct on both the TTNST and the interview items and analyzed the new correlations. Here we are interested to see if there is a significant difference in the correlations with or without student reasoning included in scoring. In middle school, students' brief assessment scores and interview total correct scores correlated strongly, $r(38) = 0.83$, $p < .001$, which was only slightly higher than the $r = 0.79$ observed when using rubric scores based on student reasoning. The difference between these two correlations was statistically negligible, $r_1 - r_2 = 0.04$, $p = .04$ (Counsell & Cribbie, 2015)². Brief assessment scores also still correlated strongly with TTNST total correct scores, $r(38) = 0.73$, $p < .001$. Compared to $r = 0.71$ with reasoning included, this difference was statistically negligible, $r_1 - r_2 = 0.01$, $p < .001$.

For elementary students, the same pattern of results held. Students' brief assessment scores and interview total correct scores correlated strongly, $r(39) = 0.84$, $p < .001$, which was a negligible difference compare to the $r = 0.89$ observed when including student reasoning, $r_1 - r_2 = 0.05$, $p = .007$. The same was true for TTNST total correct scores. There was still a strong correlation, $r(39) = 0.88$, $p < .001$, that was negligibly different when compared to $r = 0.88$ with reasoning included, $r_1 - r_2 = 0$, $p < .001$.

Therefore, while there is inevitably some information lost when moving from an in-depth interview or a lengthier test of strategy usage to a brief multiple-choice measure, the difference here is not substantial in scoring. Overall, the strong correlation between student scores on the three measures provide substantial evidence that they measure the same construct.

Discussion

The purpose of this study was to gather further validity evidence for a brief assessment of mature number sense. More specifically, we aimed to address the argument that a brief multiple-choice assessment cannot effectively measure students' mature number sense. Based on the results summarized above, we contend that there is evidence to suggest that all three assessments measure the same construct. The brief assessment's scores correlated strongly with those from the more time-intensive Three-Tier Number Sense Test (Yang, 2019) and a scripted interview (Markovits & Sowder, 1994). This correlation is especially encouraging because the Three-Tier Test and the interview assess students' reasoning in addition to their answers. A common critique of brief, multiple-choice assessments is their perceived inability to adequately capture students' rich mathematical thinking and conceptual understanding. Although we acknowledge that a brief multiple-choice measure can never capture the same depth of student thinking as a one-on-one interview, this study demonstrates a strong relation between the scores from both types of assessments.

2) In negligible effect testing, the null hypothesis states that the difference is non-negligible, whereas the alternative hypothesis is that the difference is practically negligible (Counsell & Cribbie, 2015).

Importantly, the brief assessment captures very similar information about student thinking but in a substantially shorter amount of time, with a median duration of under 10 minutes. This median duration aligns with previous findings when administered in a whole classroom setting (e.g., [Kirkland et al., 2024](#)). In contrast, the median completion time for Yang's TTNST in the present study was 28 minutes, but this assessment requires an 80-minute allotment when administered in a classroom setting. Unlike an in-depth interview, the brief assessment does not require one-on-one administration. We do not wish to imply that one-on-one student interviews or probing student thinking through follow-up questions lack value for educators and researchers. However, these methods can be impractical at scale due to time and resource constraints. The brief assessment, therefore, offers a more efficient yet effective measure of mature number sense, demonstrating strong correlations with more resource-intensive measures.

Student Performance Across the Three Measures

From the results presented here, we see that scores of students' mature number sense from standardized multiple-choice assessments and in-depth interview protocols are consistent. However, there were some overall performance differences across the three measures. Student scores on the scripted interview items ($M = 52\%$ correct for MS, 48% for ELEM) were lower than their brief assessment scores ($M = 70\%$ correct for MS, 62% for ELEM). To analyze this difference, we looked at item-level performance on the interview. There were two Effect of Operations items (#6 and #10) that students struggled with significantly (5% for MS, 17% for ELEM and 8% for MS, 22% for ELEM correct respectively). These items as well as the other Effect of Operations items are shown below in [Table 4](#).

Table 4

Effect of Operations Items in Scripted Interview

| Item # | Item | % Correct |
|--------------------------------|---|-------------------------|
| Middle School Interview | | |
| 2 | Please choose the best estimate for the answer below. $217 \div 0.35$ Correct Answer: Greater than 217 Answers Provided: <ul style="list-style-type: none"> • Less than 217 • Equal to 217 • Greater than 217 • Can't Tell | 38% (15 of 40 students) |
| 6 | Which of the below is the closest estimate of 18×86 ? Correct Answer: 18×90 Answers Provided: <ul style="list-style-type: none"> • 20×90 • 20×86 • 18×90 • Can't answer without calculating | 5% (2 of 40 students) |
| 10 | 50×30 is an estimate of 53×27 . Is the exact answer for 53×27 greater than, less than, or equal to 1,500? Correct Answer: Less than 1500 Answers Provided: <ul style="list-style-type: none"> • Less than 1500 • Equal to 1500 • Greater than 1500 • Can't Tell | 8% (3 of 40 students) |
| 14 | Which of the following is the closest estimate of 42×34 ? Correct Answer: 40×35 Answers Provided: <ul style="list-style-type: none"> • 40×34 • 40×35 • 42×30 • Can't Tell | 45% (18 of 40 students) |

| Item # | Item | % Correct |
|------------------------------------|---|------------------------|
| Elementary School Interview | | |
| 2 | What is true about the answer for $38 \div 6$? Correct Answer: Between 6 and 7 Answers Provided: <ul style="list-style-type: none"> • Between 5 and 6 • Between 6 and 7 • Between 7 and 8 • Between 8 and 9 | 61% correct (25 of 41) |
| 6 | Which of the below is the closest estimate of 28×9 ? Correct Answer: 30×9 Answers Provided: <ul style="list-style-type: none"> • 30×10 • 28×10 • 30×9 • Can't answer without calculating | 17% correct (7 of 41) |
| 10 | 30×20 is an estimate of 33×17 . Is the exact answer $>$, $<$, or $=$ to 600? Correct Answer: Less than 600 Answers Provided: <ul style="list-style-type: none"> • Less than 600 • Equal to 600 • Greater than 600 • Can't Tell | 22% correct (9 of 41) |
| 14 | Which of the following is equivalent to $63 - 28$? Correct Answer: $65 - 30$ Answers Provided: <ul style="list-style-type: none"> • $(60 - 20) + (8 - 3)$ • $65 - 30$ • $63 - 30 - 2$ • None of the above | 27% correct (11 of 41) |

If we remove items #6 and #10 from the analysis, then student performance on the interview ($M = 60\%$ correct for MS, 52% for ELEM) was similar to the other two measures. These items may have been too difficult for students to solve without the use of paper and pencil. In the original study (Markovits & Sowder, 1994), it is not clear if students were able to use paper and pencil on these estimation items. In addition, before the number sense intervention in that study, middle school students also struggled significantly on these two items (e.g., 0% correct on #6) before improving after intervention (e.g., 70% correct on #6). Interestingly, middle school students did significantly better on Item #14 despite it being a very similar estimation of two-digit multiplication problem. However, their reasoning revealed students were often guessing here based on the options provided ($M = 1.65$ on the rubric). If we were to use this interview protocol in the future, the Effects of Operations items should be altered to be less difficult mentally.

Multiplication Facts and Mature Number Sense

As predicted, mature number sense was moderately correlated with tests of basic arithmetic facts. Interestingly, the correlation appeared descriptively stronger with upper elementary students' performance on multiplication facts than with middle school students' performance on addition facts. However, it is important to note that our study was not designed to compare these associations directly. Nonetheless, the findings are significant. To our knowledge, this is the first study to examine mature number sense together with scores from a multiplication test. Notably, the association between these two constructs was relatively constant across the elementary grades ($r = 0.69$ in 3rd, 0.68 in 4th, 0.67 in 5th), despite students being first introduced to multiplication in 3rd grade.

The current study cannot establish the direction or cause of the association between mature number sense and multiplication fact performance; therefore, future research is necessary to clarify the mechanisms involved. Understanding the temporal and causal dynamics of this association is crucial, given the ongoing debates about instructional focus

in mathematics education. Some have argued that we should prioritize mature number sense in instruction, positing that it facilitates developing fluency in multiplication (Boaler, 2015). Thus, mature number sense should be the focus of instruction first. According to this perspective, instruction should initially focus on nurturing mature number sense and de-emphasize timed practice with multiplication facts. On the other hand, findings from others (e.g., Nelson et al., 2016) suggest that proficiency with multiplication facts, particularly their automatic retrieval, predicts later math success, extending beyond even the grade levels where these facts are formally taught. The rationale is that quick retrieval or derivation of single-digit facts reduces cognitive load, which might be essential for the reasoning processes inherent in mature number sense (Berrett & Carter, 2018; Fuchs et al., 2005). Additionally, well-structured practice with multiplication facts might build a robust, interconnected memory network of number knowledge in long-term memory that facilitates number sense through the spreading activation among the nodes. Exploring these mechanisms and determining the time course and causal factors behind the associations found in the present study will deepen our understanding of mathematical cognition and may help guide the development of more effective educational practices.

Limitations

While we feel confident in the results presented above, there are several limitations to the studies presented. Our sample size was intentionally smaller due to the intensive nature of one-on-one interviews and the expected high correlations between the three measures of mature number sense in our power analysis. However, the modest sample sizes led to more uncertainty in our estimates, especially with middle school students, of the association between the brief assessment and both the Yang TTNST, MS: $r(38) = 0.71$, 95% CI [0.52, 0.84]; ELEM: $r(39) = 0.88$, 95% CI [0.78, 0.93], and the scripted interview, MS: $r(38) = 0.79$, 95% CI [0.63, 0.88]; ELEM: $r(39) = 0.89$, 95% CI [0.80, 0.94]. A larger sample size may have allowed for a more precise estimate of the correlations between the three measures, especially with middle school students. The samples for both studies are not representative of the racial and ethnic diversity typically found in the US public school population. Moreover, since participation required coming to campus, the sample might be biased toward students who identify more with school mathematics. Students who do not enjoy math or who struggle with it may have been less inclined to participate. Conversely, it is also plausible that some parents encouraged participation for children who need additional support in math. Future studies should conduct research on-site at after-school programs and community centers to ensure broader representation. Although this approach was not feasible for the current studies due to the COVID-19 pandemic, it should be a priority moving forward.

Conclusion

Our goal was to demonstrate that mature number sense can be measured through a brief assessment, enabling teachers to efficiently gather information on their students' number sense and allowing researchers to contribute to fundamental knowledge of this construct. Building on initial work (Kirkland et al., 2024) to rigorously develop a practical measure of mature number sense with valid and reliable scores (AERA, APA, NCME, 2014), the studies presented here offer significant evidence of the brief assessment's ability to accurately and efficiently measure a core component of mathematical proficiency (Kilpatrick et al., 2001). The brief assessment of mature number sense can be used beyond these studies to further investigate the properties of mature number sense and, critically, examine if certain instructional practices can improve students' mature number sense. It can also help characterize the relationship between mature number sense and other constructs in mathematical cognition, especially students' performance on a multiplication fact test. By continuing to depict mature number sense's nomological network, researchers can more fully understand how students develop key mathematical concepts and provide educators with rigorous evidence on effective strategies to help ensure all students have the disposition to *make sense of numerical situations*.

Funding: This research was supported by the National Science Foundation under grant DRL EHR 2100214.

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

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

Ethics Statement: All research procedures were consistent with APA ethical guidelines. The study procedures were approved under the University of Notre Dame Institutional Review Board (IRB) protocol 19-04-5346.

Related Versions: Portions of the middle school study were previously reported in the first author's dissertation (Kirkland, 2022). This work includes an extension of the analysis of the middle school students' data as well as the inclusion of a second study with 3rd-5th grade students.

Data Availability: Data are available upon request to the first author.

Supplementary Materials

The Supplementary Materials contain the following items (for access see Kirkland et al., 2024):

- Table S1 shows the items used in the number sense interview with middle school students.
- Table S2 shows the items used in the number sense interview with elementary school students.
- Table S3 shows the items used in the brief assessment of mature number sense with middle school students.
- Table S4 shows the items used in the brief assessment of mature number sense with elementary school students.

Index of Supplementary Materials

Kirkland, P. K., Guang, C., Otuonye, C., & McNeil, N. M. (2024). *Supplementary materials to "A brief, multiple-choice assessment of mature number sense is strongly correlated with more resource-intensive measures"* [Research materials]. OSF. <https://osf.io/4aywh/>

References

- Alsawaie, O. N. (2012). Number sense-based strategies used by high-achieving sixth grade students who experienced reform textbooks. *International Journal of Science and Mathematics Education*, 10(5), 1071–1097. <https://doi.org/10.1007/s10763-011-9315-y>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, 26(1), 85–112. https://doi.org/10.1207/s15516709cog2601_3
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, 38(4), 333–339. <https://doi.org/10.1177/00222194050380040901>
- Berrett, A. N., & Carter, N. J. (2018). Imagine math facts improves multiplication fact fluency in third-grade students. *Journal of Behavioral Education*, 27(2), 223–239. <https://doi.org/10.1007/s10864-017-9288-1>
- Boaler, J. (2015, January 28). Fluency without fear: Research evidence on the best ways to learn math facts. *YouCubed*. <https://www.youcubed.org/wp-content/uploads/2017/09/Fluency-Without-Fear-1.28.15.pdf>
- Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2015). Number of repetitions required to retain single-digit multiplication math facts for elementary students. *School Psychology Quarterly*, 30(3), 398–405. <https://doi.org/10.1037/spq0000097>
- Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Jr., Lindquist, M. M., & Reys, R. (1981). *Results from the second mathematics assessment of the National Assessment of Educational Progress*. National Council of Teachers of Mathematics.
- Carpenter, T. P., Lindquist, M. M., Brown, C. A., Kouba, V. L., Silver, E. A., & Swafford, J. O. (1988). Results of the fourth NAEP assessment of mathematics: Trends and conclusions. *The Arithmetic Teacher*, 36(4), 38–41. <https://doi.org/10.5951/AT.36.4.0038>
- CCSS. (2010). *Common Core State Standards for Mathematics*. National Governors Association Center for Best Practices and the Council of Chief State School Officers.

- Counsell, A., & Cribbie, R. A. (2015). Equivalence tests for comparing correlation and regression coefficients. *British Journal of Mathematical & Statistical Psychology*, 68(2), 292–309. <https://doi.org/10.1111/bmsp.12045>
- Dehaene, S. (2001). Précis of the number sense. *Mind & Language*, 16(1), 16–36. <https://doi.org/10.1111/1468-0017.00154>
- dos Santos, C. F. (2023). Non-numerical methods of assessing numerosity and the existence of the number sense. *Journal of Numerical Cognition*, 9(2), 363–379. <https://doi.org/10.5964/jnc.10215>
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, 97(3), 493–513. <https://doi.org/10.1037/0022-0663.97.3.493>
- Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., Seethaler, P. M., & Tolar, T. D. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebraic knowledge. *Developmental Psychology*, 52(12), 2085–2098. <https://doi.org/10.1037/dev0000227>
- Gay, A. S., & Aichele, D. B. (1997). Middle school students' understanding of number sense related to percent. *School Science and Mathematics*, 97(1), 27–36. <https://doi.org/10.1111/j.1949-8594.1997.tb17373.x>
- Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development*, 67(5), 2022–2044. <https://doi.org/10.2307/1131607>
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, 38(4), 293–304. <https://doi.org/10.1177/00222194050380040301>
- Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education*, 22(3), 170–218. <https://doi.org/10.2307/749074>
- Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W. (2000). Solving inductive reasoning problems in mathematics: Not-so-trivial pursuit. *Cognitive Science*, 24(2), 249–298. https://doi.org/10.1207/s15516709cog2402_3
- Howden, H. (1989). Teaching number sense. *The Arithmetic Teacher*, 36(6), 6–11. <https://doi.org/10.5951/AT.36.6.0006>
- Hsu, C.-Y., Yang, D.-C., & Li, F. M. (2001). The design of the fifth and sixth grade number sense rating scale. *Chinese Journal of Science Education*, 9(4), 351–374.
- Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education*, 38(3), 258–288.
- Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77(1), 153–175. <https://doi.org/10.1111/j.1467-8624.2006.00862.x>
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, 45(3), 850–867. <https://doi.org/10.1037/a0014939>
- Kaminski, E. (2002). Promoting mathematical understanding: Number sense in action. *Mathematics Education Research Journal*, 14(2), 133–149. <https://doi.org/10.1007/BF03217358>
- Kirkland, P. K. (2022). *Characterizing mature number sense and its association to other constructs in middle school students* [Doctoral dissertation, University of Notre Dame]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/characterizing-mature-number-sense-association/docview/2661059562/se-2>
- Kirkland, P. K., Cheng, Y., & McNeil, N. M. (2024). A validity argument for a brief assessment of mature number sense. *Journal for Research in Mathematics Education*, 55(1), 51–67. <https://doi.org/10.5951/jresmetheduc-2022-0071>
- Kirkland, P. K., Guang, C., Cheng, Y., Trinter, C., Kumar, S., Nakfoor, S., Sullivan, T., & McNeil, N. M. (2022). Middle school students' mature number sense is uniquely associated with grade-level mathematics achievement. In A. E. Lischka, E. B. Dyer, R. S. Jones, J. N. Lovett, J. Strayer, & S. Drown (Eds.), *Proceedings of the Forty-Fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 911–919). Middle Tennessee State University. <https://par.nsf.gov/biblio/10402434-middle-school-students-mature-number-sense-uniquely-associated-grade-level-mathematics-achievement>
- Kirkland, P. K., Guang, C., & McNeil, N. M. (2023). Exploring the association between upper elementary students' mature number sense and grade-level mathematics achievement. In T. Lamberg & D. Moss (Eds.), *Proceedings of the Forty-Fifth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 408–417). <https://par.nsf.gov/biblio/10487476-exploring-association-between-upper-elementary-students-mature-number-sense-grade-level-mathematics-achievement>
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.) (2001). *Adding it up: Helping children learn mathematics*. Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council / National Academy Press.

- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Li, M.-N. F., & Yang, D.-C. (2010). Development and validation of a computer-administered number sense scale for fifth-grade children in Taiwan. *School Science and Mathematics*, 110(4), 220–230. <https://doi.org/10.1111/j.1949-8594.2010.00024.x>
- Marinova, M., & Reynvoet, B. (2020). Can you trust your number sense: Distinct processing of numbers and quantities in elementary school children. *Journal of Numerical Cognition*, 6(3), 304–321. <https://doi.org/10.5964/jnc.v6i3.292>
- Markovits, Z., & Sowder, J. (1994). Developing number sense: An intervention study in Grade 7. *Journal for Research in Mathematics Education*, 25(1), 4–29. <https://doi.org/10.2307/749290>
- McLaughlin, S. M., Knoop, A. J., & Holliday, G. A. (2005). Differentiating students with mathematics difficulty in college: Mathematics disabilities vs. no diagnosis. *Learning Disability Quarterly*, 28(3), 223–232. <https://doi.org/10.2307/1593660>
- McIntosh, A., Reys, B., & Reys, R. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics*, 12(3), 2–8.
- McIntosh, A., Reys, B., Reys, R., Bana, J., & Farrell, B. (1997). *Number sense in school mathematics: Student performance in four countries*. MASTEC: Mathematics, Science & Technology Education Centre. <https://ro.ecu.edu.au/ecuworks/6819>
- Menon, R. (2004). Preservice teachers' number sense. *Focus on Learning Problems in Mathematics*, 26(2), 49–61.
- National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and Evaluation Standards for School Mathematics*. NCTM.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and Standards for School Mathematics*. NCTM.
- National Council of Teachers of Mathematics (NCTM). (2014). *Principles to Actions: Ensuring Mathematical Success for All*. NCTM.
- National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. National Academy Press.
- Nejem, K. M., & Muhanna, W. (2013). The effect of using computer games in teaching mathematics on developing the number sense of fourth grade students. *Educational Research Review*, 8(16), 1477–1482.
- Nelson, P. M., Burns, M. K., Kanive, R., & Ysseldyke, J. E. (2013). Comparison of a math fact rehearsal and a mnemonic strategy approach for improving math fact fluency. *Journal of School Psychology*, 51(6), 659–667. <https://doi.org/10.1016/j.jsp.2013.08.003>
- Nelson, P. M., Parker, D. C., & Zaslofsky, A. F. (2016). The relative value of growth in math fact skills across late elementary and middle school. *Assessment for Effective Intervention*, 41(3), 184–192. <https://doi.org/10.1177/1534508416634613>
- Obersteiner, A., & Hofreiter, V. (2017). Do we have a sense for irrational numbers? *Journal of Numerical Cognition*, 2(3), 170–189. <https://doi.org/10.5964/jnc.v2i3.43>
- Parrish, S. D. (2011). Number talks build numerical reasoning. *Teaching Children Mathematics*, 18(3), 198–206. <https://doi.org/10.5951/teachmath.18.3.0198>
- Pike, C. D., & Forrester, M. A. (1997). The influence of number-sense on children's ability to estimate measures. *Educational Psychology*, 17(4), 483–500. <https://doi.org/10.1080/0144341970170408>
- Purnomo, Y. W. (2014). Assessing number sense performance of Indonesian elementary school students. *International Education Studies*, 7(8), 74–84. <https://doi.org/10.5539/ies.v7n8p74>
- Reynvoet, B., Ribner, A. D., Elliott, L., Van Steenkiste, M., Sasanguie, D., & Libertus, M. E. (2021). Making sense of the relation between number sense and math. *Journal of Numerical Cognition*, 7(3), 308–327. <https://doi.org/10.5964/jnc.6059>
- Reys, B. J. (1991). *Developing number sense: Curriculum and evaluation standards for school mathematics* (Addenda Series, Grades 5-8). ERIC.
- Reys, R., Reys, B., Emanuelsson, G., Johansson, B., McIntosh, A., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics*, 99(2), 61–70. <https://doi.org/10.1111/j.1949-8594.1999.tb17449.x>
- Reys, R., & Yang, D.-C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education*, 29(2), 225–237. <https://doi.org/10.2307/749900>
- Shumway, J. F. (2018). *Number sense routines: Building mathematical understanding every day in Grades 3-5*. Stenhouse Publishers.
- Syed, M., & Nelson, S. C. (2015). Guidelines for establishing reliability when coding narrative data. *Emerging Adulthood*, 3(6), 375–387. <https://doi.org/10.1177/2167696815587648>
- USAID. (2018). *USAID Education Policy: November 2018*. https://www.usaid.gov/sites/default/files/documents/1865/2018_Education_Policy_FINAL_WEB.pdf

- Whitacre, I., Henning, B., & Atabas, S. (2020). Disentangling the research literature on number sense: Three constructs, one name. *Review of Educational Research*, 90(1), 95–134. <https://doi.org/10.3102/0034654319899706>
- Yang, D.-C. (2003). Teaching and learning number sense – An intervention study of fifth grade students in Taiwan. *International Journal of Science and Mathematics Education*, 1(1), 115–134. <https://doi.org/10.1023/A:1026164808929>
- Yang, D.-C. (2005). Number sense strategies used by 6th-grade students in Taiwan. *Educational Studies*, 31(3), 317–333. <https://doi.org/10.1080/03055690500236845>
- Yang, D.-C. (2019). Development of a three-tier number sense test for fifth-grade students. *Educational Studies in Mathematics*, 101(3), 405–424. <https://doi.org/10.1007/s10649-018-9874-8>
- Yang, D.-C., & Chang, T.-M. (2023). Number sense performance of gifted and general fourth graders in Taiwan. In D. Ortega-Sánchez (Ed.), *Education Annual Volume 2023* (pp. 1-15). <https://doi.org/10.5772/intechopen.111752>
- Yang, D.-C., Hsu, C.-J., & Huang, M.-C. (2004). A study of teaching and learning number sense for sixth grade students in Taiwan. *International Journal of Science and Mathematics Education*, 2(3), 407–430. <https://doi.org/10.1007/s10763-004-6486-9>
- Yang, D.-C., & Li, M. F. (2008). An investigation of 3rd-grade Taiwanese students' performance in number sense. *Educational Studies*, 34(5), 443–455. <https://doi.org/10.1080/03055690802288494>
- Yang, D.-C., & Lin, Y.-C. (2015). Assessing 10- to 11-year-old children's performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research*, 57(4), 368–388. <https://doi.org/10.1080/00131881.2015.1085235>
- Yang, D.-C., & Wu, W.-R. (2010). The study of number sense: Realistic activities integrated into third-grade math classes in Taiwan. *The Journal of Educational Research*, 103(6), 379–392. <https://doi.org/10.1080/00220670903383010>



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



leibniz-psychology.org

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