Learning and Instruction 59 (2019) 21–33

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Kindergarten children's symbolic number comparison skills relates to 1st grade mathematics achievement: Evidence from a two-minute paper-and-pencil test

Zachary Hawes[a], Nadia Nosworthy[b], Lisa Archibald[a], Daniel Ansari[a,*]

[a] University of Western Ontario, Canada
[b] Andrews University, USA

ARTICLE INFO

Keywords:
Early mathematics achievement
Symbolic number
Non-symbolic number
Number line estimation
Arithmetic

ABSTRACT

Basic numerical skills provide an important foundation for the learning of mathematics. Thus, it is critical that researchers and educators have access to valid and reliable ways of assessing young children's numerical skills. The purpose of this study was to evaluate the concurrent, predictive, and incremental validity of a two-minute paper-and-pencil measure of children's symbolic (Arabic numerals) and non-symbolic (dot arrays) comparison skills. A sample of kindergarten children (M_{age} = 5.86, N = 439) were assessed on the measure along with a number line estimation task, a measure of arithmetic, and several control measures. Results indicated that performance on the symbolic comparison task explained unique variance in children's arithmetic performance in kindergarten. Longitudinal analyses demonstrated that both symbolic comparison and number line estimation in kindergarten were independent predictors of 1st grade mathematics achievement. However, only symbolic comparison remained a unique predictor once language skills and processing speed were taken into account. These results suggest that a two-minute paper-and-pencil measure of children's symbolic number comparison is a reliable predictor of children's early mathematics performance.

1. Introduction

A growing body of research points to basic numerical skills as critical precursors of later mathematics achievement. In general, children who demonstrate a strong proficiency with basic numerical skills, such as being able to quickly and accurately state the larger of two symbolic numbers (7 vs. 3), tend to also demonstrate strong performance in more advanced mathematics tasks, including arithmetic (Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013), word problems (Fuchs et al., 2010), fractions (Mou et al., 2016), and geometry (Lourenco & Bonny, 2017). Presumably, this is due in part to the hierarchical nature of mathematics learning, where earlier learned skills serve as essential building blocks in the construction of more sophisticated mathematics understanding.

The consequences of low numeracy can be substantial, not only influencing one's educational attainment but also one's personal well-being and the associated economic costs. For example, a large study carried out in the UK revealed that low numeracy had more of negative influence on one's life chances than low literacy (Gross, Hudson, & Price, 2009). Low numeracy has been found to coincide with lower earnings, lower spending, poorer health outcomes, and increased trouble with the law (Parsons & Bynner, 2005). Furthermore, Ritchie and Bates (2013) found that numerical knowledge at seven is predictive of one's SES at the age of 42, even after controlling for the individuals' own IQ and the SES of the family which they were born into.

Given the importance of basic numerical skills for later educational and occupational success, it is crucial that educators have access to research-informed and reliable assessments of basic numerical competencies. Such assessment tools are necessary in order for teachers of young children to measure students' initial skills at the beginning of the school year, track growth over the course of the year, and perhaps most importantly, identify children in potential need of early intervention.

Although early numeracy and mathematics assessments do exist, most of these assessments are fairly complicated to administer and require considerable amounts of time. In order for assessments of young children's numerical skills to be of use to the practicing teacher, these assessments should necessarily be easy to administer and require little time. Although much headway has been made in providing teachers

* Corresponding author. Department of Psychology, Brain and Mind Institute, University of Western Ontario, Western Interdisciplinary Research Building, Room 5180, 1151 Richmond Street North, London, ON, N6A 5B7, Canada.
E-mail address: daniel.ansari@uwo.ca (D. Ansari).

https://doi.org/10.1016/j.learninstruc.2018.09.004
Received 16 January 2018; Received in revised form 12 July 2018; Accepted 12 September 2018
Available online 26 September 2018
0959-4752/

with such assessments in literacy (e.g., Diagnostic Reading Assessment (DRA); Beaver, 1968: Dynamic Indicators of Basic Early Literacy Skills (DIBELS); Good & Kaminski, 2002), the same cannot be said of early mathematics assessments (but see some recent advancements by Brankaer, Ghesquière, & De Smedt, 2017 and Purpura & Lonigan, 2015). Moreover, prior efforts to design and measure basic numerical skills have yet to test the ecological validity of such assessments. It remains to be demonstrated whether numerical assessments intended for teacher use are predictive of school mathematics (e.g., teacher assigned math grades). The present study aimed to address this gap.

1.1. Rationale and aims of current study

In this paper, we share the results of implementing a two-minute paper-and-pencil assessment of young children's basic numerical skills. The Numeracy Screener, as it is referred to hereafter, was specifically designed with the educator and researcher in mind, providing both parties with a quick and research-informed method of assessing young children's (K-3) basic numerical skills (for more information see: Nosworthy et al., 2013 and www.numeracyscreener.org). More specifically, the tool was designed to measure children's non-symbolic and symbolic comparison skills. In prior research, it was found that performance on both the symbolic and non-symbolic portions of the assessment were concurrently related to individual differences in arithmetic achievement across 1st to 3rd grade. However, when symbolic and non-symbolic comparison skills were entered in the same model – one that included other control variables – only performance on the symbolic comparison task accounted for unique variance in arithmetic (Nosworthy et al., 2013). Although these findings provide initial promise of the measure, further steps are necessary in order to further test the utility of the Numeracy Screener as a valid and reliable assessment tool. The current study aimed to extend our previous work by (a) increasing the sample size, (b) more narrowly defining a population of interest (kindergarten[1]), (c) testing both the concurrent and predictive validity of the instrument, (d) evaluating the test-retest reliability, (e) examining convergent construct validity by comparing performance to another common measure of magnitude processing (i.e., the number line estimation task), and (f) including school grades as an ecologically valid measure of mathematics achievement. Ultimately, the efforts of this work are directed towards the creation of a valid, reliable, and publicly available assessment tool of young children's basic magnitude processing skills. In working towards this goal, the central aim of the current paper was to examine how basic numerical skills at kindergarten concurrently and longitudinally relate to more formal school mathematics, including arithmetic and teacher-assigned math grades.

[1] In Ontario, kindergarten consists of a two-year play-based program. The first year of kindergarten is known as Junior Kindergarten and begins when children are 3 or 4 years of age; it is the equivalent of what other countries (e.g., the USA) refer to as pre-school. Senior Kindergarten begins when children are 4 or 5 years of age and is more in line with what other countries refer to as kindergarten (e.g., the USA). In the current study, our sample was drawn from Senior Kindergarten students; referred to in this paper as kindergarten students in an effort to ease communication and maintain standards with other countries. In this paper, we also distinguish between informal and formal education by referring to kindergarten as an example of informal education and 1st grade as an example of formal education. This decision is based on the play-based curriculum guidelines of the kindergarten program and the more formal expectations of 1st grade. However, we acknowledge that this distinction is somewhat arbitrary and dependent upon the teacher and does not preclude the possibility of formal learning opportunities that some kindergarten classrooms might afford.

1.2. Overview of children's magnitude processing skills

1.2.1. Attempts to measure children's knowledge and representation of number

To date, efforts to identify early predictors of mathematical success have largely focused on children's numerical magnitude processing skills. Indeed, the study of children's magnitude processing skills has received concerted attention from researchers in cognitive neuroscience, psychology, and education alike (e.g., see De Smedt, Noël, Gilmore, & Ansari, 2013). Presumably, the reason for such convergence has to do with what numerical magnitude tasks are thought to reveal about individuals' underlying representations of number. That is, the accuracy and speed with which an individual can access the numerical magnitude of sets of objects (non-symbolic) or symbolic representations (e.g., 5 or 'five') is typically taken as an indicator of the strength and precision of one's representation of number. Arguably, the three most widely used tasks to measure such a process involve non-symbolic number comparisons, symbolic number comparisons, and number line estimation (e.g., see Schneider et al., 2017). It is for this reason that we selected these tasks in the context of the present study.

Both non-symbolic and symbolic number comparison paradigms are similar in that they involve comparing and identifying the larger of two quantities (be they symbolic, 5 vs. 3, or non-symbolic ∷· vs. ˙.˙) as quickly and accurately as possible. The ability to quickly access numerical magnitudes is fundamental to a range of mathematical tasks, including exact and approximate calculations. For example, to know that combining two sets of objects (●● + ●) results in a total sum that is greater than one set alone, requires attending to the numerosity of the sets and not some other feature, such as physical size or total area (e.g., ●● + ● = ● • • and not ⬤). Similarly, to know that 58 + 45 is either approximately 100 or exactly 103, requires the ability to access the numerical magnitude of the symbolic addends, 58 and 45. In both examples, access to numerical magnitudes and not some other property is the common and critical property involved in the calculation process. Despite the similarities, the two tasks differ in several key regards. The non-symbolic number comparison task is thought to serve as an index of one's Approximate Number System (ANS); an ancient and rudimentary ability to discriminate between non-symbolic numerical magnitudes that is available early in infancy (Xu & Spelke, 2000) and shared with other non-human species (Feigenson, Dehaene, & Spelke, 2004). Symbolic number comparison on the other hand provides a measure of one's understanding of number symbols and the exact quantities that they represent. Performance on this task is mediated through cultural experiences with the symbolic number system and thus takes time to develop and is not immediately available early in life (Núñez, 2017). Taken together, although both tasks are used to measure one's ability to access and make judgments about numerical magnitudes, the two tasks differ with respect to when and how the two systems become available for use. These differences are non-trivial and underscore critical questions and debate regarding the extent to which these two magnitude systems are related and interact with one another over development (e.g., De Smedt et al., 2013; Leibovich, Katzin, Harel, & Henik, 2017). Moreover, questions remain about how individual differences on symbolic and non-symbolic magnitude tasks are related to future mathematics achievement (e.g., see De Smedt et al., 2013).

Number line estimation tasks represent another way in which researchers have attempted to measure individuals' numerical magnitude as well as more general numerical reasoning skills (Schneider et al., 2017). The most common form of this assessment involves presenting participants with horizontal line flanked by two end-points (e.g., 0 and 100) and asking them to estimate the location of a given number (e.g., 73). This task involves the mapping of numerical magnitudes onto continuous space and has been of theoretical and practical interest as performance on this task has been used as an indicator of the accuracy and precision of one's 'mental number line' (Dehaene, 2011).

1.2.2. Relations between non-symbolic number comparison, symbolic number comparison, number line estimation, and mathematics performance

Although non-symbolic number comparison, symbolic number comparison, and number line estimation tasks all are understood to measure one's representation of number, recent research has revealed important differences in how each task relates to mathematics performance (e.g., see Schneider, Thompson, & Rittle-Johnson, in press). Of the three tasks, research on relations between non-symbolic comparison skills and mathematics has revealed the least consistent findings (De Smedt et al., 2013; Schneider et al., 2017). For example, while a host of studies have revealed significant concurrent and longitudinal relations (e.g., Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011), another large body of research has failed to demonstrate such relations (e.g., Holloway & Ansari, 2009; Mundy & Gilmore, 2009; Sasanguie, Van den Bussche, & Reynvoet, 2012). As discussed above, in prior work with the Numeracy Screener, performance on the non-symbolic task was correlated with arithmetic performance but failed to explain unique variance once the influence of other variables was taken into account, such as working memory, spatial skills, vocabulary, and symbolic comparison skills (Nosworthy et al., 2013). The reasons for the inconsistent relations between non-symbolic tasks and mathematics are multifaceted and complex – and beyond the scope this article – but include differences in methodologies (e.g., how the stimuli were created and presented) and the inherent impossibility of controlling for all visual-spatial properties of the stimuli (e.g., Gebuis & Reynvoet, 2012; Leibovich, Katzin, Harel, & Henik, 2016).

Despite the inconsistencies between paradigms and findings, three separate meta-analyses all reach a similar conclusion: Non-symbolic processing skills are statistically significant predictors of individual differences in mathematics performance (r = 0.24, based on 195 effect sizes; Schneider et al., 2017; see also Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014). The evidence for a link between non-symbolic processing and mathematical learning difficulties (aka developmental dyscalculia) is also mixed, with some researchers finding an association (Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010) and others failing to find one (De Smedt & Gilmore, 2011; Iuculano, Tang, Hall, & Butterworth, 2008; Rousselle & Noël, 2007). Taken together, research points to non-symbolic processing as one potential, albeit weak, contributor to children's mathematics learning. However, as has been argued elsewhere, it is possible that non-symbolic skills, especially prior to formal schooling, might play an important role in the grounding of symbolic numbers (Mundy & Gilmore, 2009).

Compared to non-symbolic processing skills, research on the relation between symbolic comparison skills and mathematics presents a much clearer picture. Across different studies and populations, consistent positive correlations have been reported (e.g., see De Smedt et al., 2013). A recent meta-analysis by Schneider et al. (2017) revealed relations between symbolic comparison and mathematical competence to be r = 0.30, averaged across 89 effect sizes. Moreover, several studies have revealed longitudinal relations between symbolic comparison skills and mathematics performance (Bartelet, Vaessen, Blomert, & Ansari, 2014; Matejko & Ansari, 2016; Vanbinst, Ceulemans, Peters, Ghesquière, & De Smedt, 2018; Xenidou-Dervou, Molenaar, Ansari, van der Schoot, & van Lieshout, 2017). For example, a recent study by Xenidou-Dervou et al. (2017) found evidence that symbolic comparison skills (digit range: 1–9) at Kindergarten are a robust and consistent predictor of mathematics achievement (i.e., based on questions targeting number relations, mental arithmetic, applied calculations, and measurement) in both Grade 1 and 2, even after controlling for individual differences in IQ and working memory. Research has also revealed that children with dyscalculia tend to demonstrate weaknesses in symbolic comparison tasks (e.g., see De Smedt & Gilmore, 2011; Landerl & Kölle, 2009; Rousselle & Noël, 2007). Overall, there is evidence to suggest that symbolic number comparison skills play an important role in the learning of mathematics.

Number line estimation skills have also consistently been linked to performance in mathematics (Schneider, Thompson, & Rittle-Johnson, in press; but also see Sasanguie & Reynvoet, 2013). For example, young children's performance on number line tasks has been found to relate to proficiencies in arithmetic (Booth & Siegler, 2008), standardized achievement scores (Sasanguie, De Smedt, Defever, & Reynvoet, 2012) and school grades (Schneider, Grabner, & Paetsch, 2009). A recent review on the topic suggests the correlation falls somewhere between 0.40 and 0.50, indicating a moderate-to-strong relation (Schneider et al., in press).

The number line task was included in the current study not only as a potentially important predictor of children's mathematics performance but also as a way of testing for convergent validity with the Numeracy Screener. Given that both number line and magnitude comparison tasks are thought to measure children's access to numerical magnitudes (Siegler & Booth, 2004), it should be expected that the two tasks are positively correlated. Indeed, prior research has found this to be the case (Laski & Siegler, 2007). Thus, in the current study, we were interested in the extent to which children's performance on the number line task correlated with children's magnitude comparison skills; a finding that would provide evidence of convergent construct validity.

1.3. The present study

In sum, there is a large body of research that suggests non-symbolic comparison, symbolic comparison, and number line estimation tasks are each related to children's mathematics performance. In the current study, we set out to test both the shared and unique associations of each measure with children's arithmetic performance measured in Kindergarten and their school mathematics grades approximately one year later at the end of 1st grade. That is, to what extent are the various numerical measures correlated with one another and to what extent does each measure correlate with arithmetic and mathematics performance once the influence of the other variables is taken into account? The results of such analyses are important in terms of demonstrating the predictive validity of the Numeracy Screener as well as providing further insight into the common and distinct pathways between measures of children's basic number knowledge and their mathematics performance. This study is unique in that it uses children's school mathematics grades as a longitudinal outcome measure of mathematics achievement. To date, the majority of research has looked at relations between basic numerical competencies and standardized measures of arithmetic or mathematics and, thus, relatively little is known about how basic numerical skills relate to teacher- and school-valued mathematics.[2] Using children's overall mathematics grades, which included assessments of children's number sense, geometry, measurement, data management, and early algebra, allowed us to examine the extent to which basic numerical skills potentially relate to more comprehensive measures of mathematics that go beyond assessments that more narrowly focus on numerical reasoning (e.g. arithmetic). Moreover, given that a central goal of ours is to make the Numeracy Screener widely available for classroom use, it is critical to assess how performance on the measure predicts later curriculum-based mathematics performance.

2. Methods

2.1. Participants

Four hundred and thirty-nine 5- to 6-year-old children (212 males, 227 females, M_{age} = 5.86 years, SD = 0.31, range = 5.16–6.42) participated in the study. All children were attending Senior Kindergarten[3]

[2] To view the Ontario mathematics curriculum and standards for grades 1–8, see http://www.edu.gov.on.ca/eng/curriculum/elementary/math18curr.pdf.

[3] Note that not all teachers assigned a letter grade to each one of the strands

at the time of initial testing and were selected from 16 different elementary schools in Southern Ontario, Canada. Ethics approval was granted by the University of Western Ontario as well as the two participating Ontario District School Boards. Written consent was obtained by the principal from each of the 16 schools and all parents/guardians of child participants. No specific criteria were listed for participation, as we were interested in obtaining a representative sample of kindergarten students in Ontario. Details on missing or incomplete data are described further under the descriptions of each measure.

2.2. Testing procedure

All testing was carried out one-to-one by a trained experimenter in a quiet location of the school. The tasks were administered in a semi-random order: The blocked nature of the Numeracy Screener precluded fully random task administration. In version A, the symbolic comparison task was presented first followed by the non-symbolic task. Version B presented the tasks in opposite order. Participants were randomly assigned Version A or B.

Data collection procedures differed somewhat between the two participating school boards and, as outlined below, provided opportunity for different follow-up analyses with each sub-population. In one of the school boards, data were collected at the end of the kindergarten school year (i.e., May–June). In this sample (n = 306), testing lasted about 20 min and the measures were restricted to the Numeracy Screener and the number line estimation task.

Data collected in the other school board (n = 133), occurred earlier in the year (January–March) and was part of larger data collection procedure that included additional measures (i.e., several measures of language and basic processing skills). For the purpose of the current study, two of these measures (sentence recall and rapid color naming) were used as control variables in an attempt to demonstrate incremental validity and rule out potential confounding variables in the relation between performance on the Numeracy Screener and mathematics grades. For example, we wanted to control for processing speed given that the Numeracy Screener is a timed task. Is there something specific about the Numeracy Screener in its relations to children's maths skills that goes beyond speed of processing capacities? Furthermore, it is well established that children's language skills, and more specifically verbal working memory, as captured with the sentence recall task (Alloway & Ledwon, 2014), are related to mathematics performance (e.g., see Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013). Thus, we wanted to rule out the possibility that language skills/working memory might account for the shared relations between performance on the Numeracy Screener and children's mathematics performance. For this sample (n = 133), it was also possible to return for a second wave of data collection towards the end of the school year (May–June). As detailed below, this provided the opportunity to assess the test-retest reliability of the measures. With the addition of the control measures, testing lasted about 40 min per child.

2.3. Tests and materials

2.3.1. Numeracy Screener

Participants were required to compare pairs of magnitudes ranging from 1 to 9. Stimuli were given in both symbolic (56 digit pairs) and non-symbolic (56 pairs of dot arrays) formats. In both formats, numerical magnitude was counterbalanced for the side of presentation (i.e., 2|7, 7|2). In the non-symbolic form, dot stimuli were controlled for area and density (for specific details see Nosworthy et al., 2013).

Table 1
Numerical pairs and ratios for the numerical comparison task.

Number Pair	Ratio
1–9	0.11
1–8	0.13
1–7	0.14
1–6	0.17
1–5	0.20
2–9	0.22
2–8	0.25
2–7	0.29
3–9	0.33
3–8	0.38
2–5	0.40
3–7	0.43
4–9	0.44
3–6	0.50
4–8	0.50
5–9	0.56
4–7	0.57
3–5	0.60
5–8	0.63
2–3	0.67
5–7	0.71
6–8	0.75
7–9	0.78
4–5	0.80
5–6	0.83
6–7	0.86
7–8	0.88

Test items were presented in increasing difficulty levels. That is, the numerical ratio between the numerical magnitudes presented was manipulated so that easier items (with smaller ratios) were presented first and more difficult items gradually followed (increasingly larger ratios). This was done in an effort to maintain student motivation throughout the task. The order of trials in our assessment was similar to the order of ratios presented in Table 1. Order was slightly varied between symbolic and non-symbolic conditions to ensure that the order of presentation of items was not identical between conditions, but both followed a similar pattern where pairs of symbolic and non-symbolic stimuli with relatively smaller ratios were presented before larger ratios. The ratio (small/large) between numerical pairs ranged from 0.11 to 0.89. For example, the ratio between 3 and 5 is 0.60 (see Table 1 for pairs and ratios used).

During the test, participants were told to cross out the larger of the two magnitudes and were given two-minutes to complete the symbolic condition and two-minutes to complete the non-symbolic condition. To ensure that participants understood the task, each child completed three sample items with the examiner and then nine practice items on their own before beginning the assessment (see Fig. 1A and D). This procedure was the same for both symbolic and non-symbolic conditions. During the instructions given for the non-symbolic condition, participants were told not to count the dots. Examiners were again able to emphasize this instruction during the participants' completion of the practice items. (see Fig. 1 for example of the screener; a complete version of the test, available for downloading, can be viewed at www.numeracyscreener.org). Children received individual scores for each section of the Numeracy Screener according to total number of correct responses achieved within the two-minute time limit.

Data were excluded or not available for 44 children on the non-symbolic comparison task and 43 children on the symbolic comparison task. Missing data resulted from not being administered the test (25 non-symbolic; 26 symbolic), failing to complete the task properly (e.g., skipping pages, crossing off all answers; 10 non-symbolic; 5 symbolic) or achieving scores falling outside 2.5 standard deviations from the

(footnote continued)
in Mathematics. This was due to the fact that some teachers had already reported a final grade for a given strand during the winter semester. With that said, all participants received grades for at least three of the five math strands and thus children's grade scores still represent a comprehensive measure of mathematics achievement.

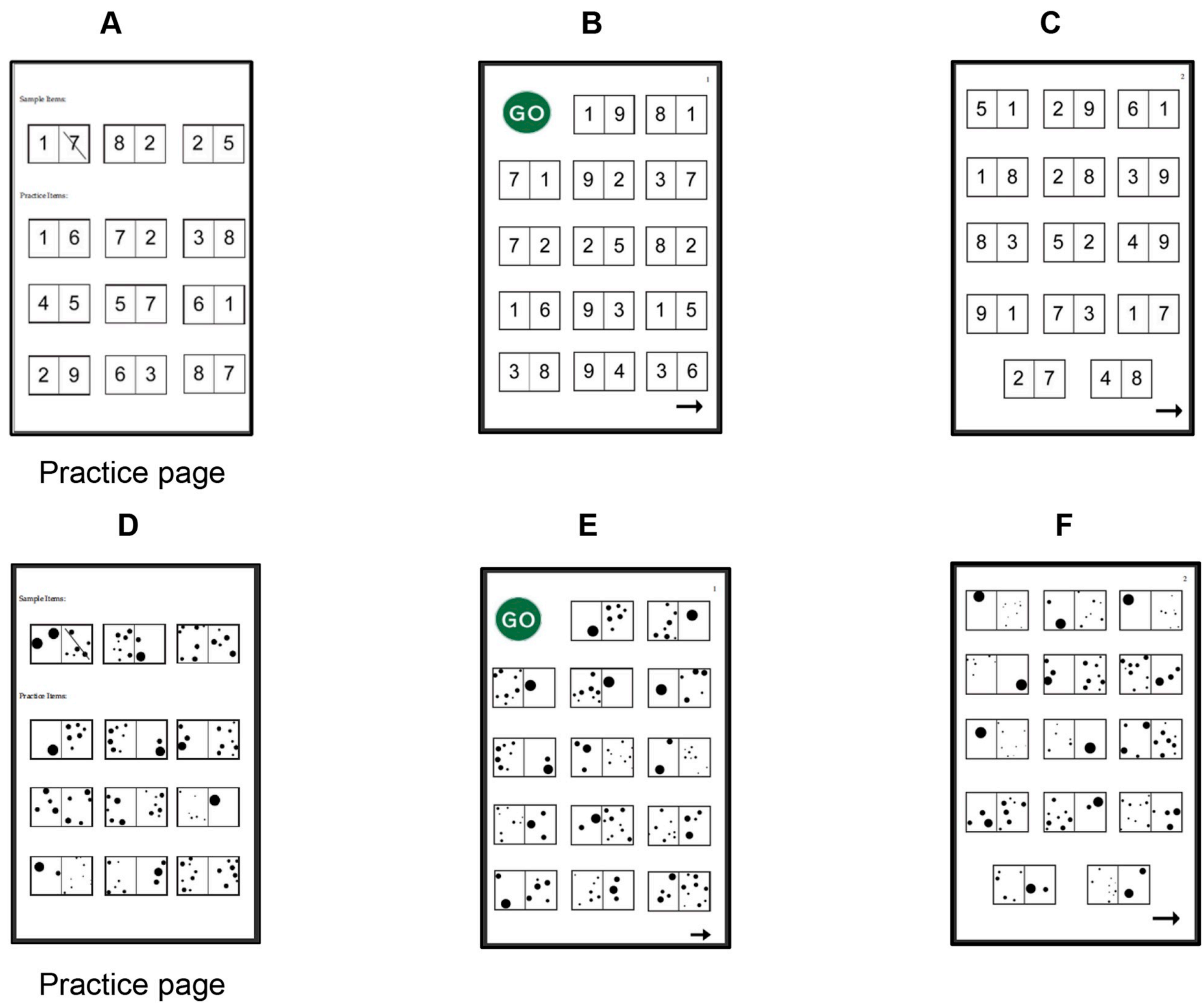

Fig. 1. Example items from the Numeracy Screener. Figures A–C are examples of the symbolic test items. Figures D–F are examples of the non-symbolic test items.

mean and thus were considered outliers (9 symbolic; 12 non-symbolic).

2.3.2. Number line estimation

For this task, children were asked to estimate the spatial position of an Arabic digit on a physical number line. Participants were presented with a number line 25 cm in length with the Arabic digit 0 at one end and the Arabic digit 10 at the other end and a target number in a large font printed above the line (see Fig. 2). The children were presented with target numbers 1–9, one at a time, in random order, and were asked to indicate where the number would go on the line. Each item was presented on its own sheet of paper. Instructions were given as follows, "*This number line goes from 0 at this end to 10 at this end. If this is 0 and this is 10, where do you put N?*" (with *N* being the number specified on the particular trial). To ensure that participants understood the instructions given, each child completed one sample item before beginning the test. The number five was used for this purpose and was thus excluded in any analyses of the task. For the sample item only, participants were provided with feedback if it was clear that they did not understand the task. Individual performance on the task was measured as the average percentage of estimation error (i.e., percent absolute error [PAE]) across all eight trials. To calculate PAE, the following equation was used (see Siegler & Booth, 2004):

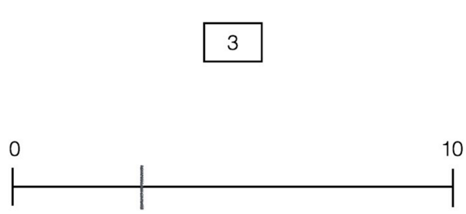

Fig. 2. Example of a test item on the number line estimation task. Here, the hatch mark represents where an individual may place his or her response to the question "*If this is 0 and this is 10, where do you put 3?*"

$$\text{PAE} = \left| \frac{\text{Estimate} - \text{Estimated Quantity}}{\text{Scale of Estimates}} \right| \times 100$$

For example, if a child was asked to estimate the location of 3 on the number line and placed his/her response at the location that corresponded to 5, percent absolute error would be 20%:

[(5–3)/10] x 100

Data were excluded or not available for a total of 73 children due to the following reasons: incomplete or mishandled booklets (7), not completing the task properly (e.g., placing each answer in the exact same spot), (35), incorrect test administration (28), or achieving scores falling outside 2.5 standard deviations from the mean (11).

Table 2
Description of letter grades according to provincial curriculum expectations.

Teacher-assigned letter grade	Achievement of Provincial Curriculum Expectations
A- to A+	The student has demonstrated the required knowledge and skills. Achievement exceeds the provincial standard.
B- to B+	The student has demonstrated most of the required knowledge and skills. Achievement meets the provincial standard.
C- to C+	The student has demonstrated some of the required knowledge and skills. Achievement approaches the provincial standard
D- to D+	The student has demonstrated some of the required knowledge and skills in limited ways. Achievement falls much below the provincial standard.
R	The student has not demonstrated required knowledge and skills

2.3.3. Arithmetic task

A simple non-standardized, paper-and-pencil arithmetic measure was administered to measure participants' arithmetic skills. Participants were given 5 single-digit addition (1 + 2, 1 + 3, 4 + 1, 3 + 2, 5 + 1) and 5 single-digit subtraction problems (3-1, 2-1, 4-3, 3-2, 4 -2). Children received one point for each correctly answered problem for a maximum score of ten. No time constraints were given to complete this task.

2.3.4. Mathematics achievement

Report card grades in mathematics were collected in the late spring of first grade (12–16 months from the first testing session). A single grade score for each child was derived by converting teacher-assigned letter grades (e.g., A+; see Table 2) into discrete numerical values as follows: R = 0, D- = 1, D = 2, D+ = 3 … A- = 10, A = 11, A+ = 12. Because the Ontario Mathematics curriculum (Grades 1–8) is divided into five separate math strands, children were given a grade for each strand.[2] The strands include: 1) *Number Sense and Numeration* (e.g., adding and subtracting numbers to 20, representing and ordering whole numbers to 50, counting by 1's, 2's, 5's, and 10's, etc.), 2) *Measurement* (e.g., measuring using non-standard units, developing a sense of area, reasoning about size and number of units, etc.), 3) *Geometry and Spatial Sense* (classifying and sorting 2- and 3D shapes, relating shapes to other shapes, describing location using positional language etc.), 4) *Patterning and Algebra* (e.g., creating and extending repeating patterns, concepts of equality using concreate models, etc.) and 5) *Data Management and Probability* (e.g., collecting and organizing categorical data, reading and displaying data using graphs, describing the likelihood that an event will occur, etc.). For the purposes of this paper an overall grade was calculated by averaging performance across the five strands and, thus, represents a comprehensive measure of teacher-rated mathematics achievement. Note that a detailed list of the concepts and content children are expected to master for each strand can be found on page 32 of the Ontario Grade 1 mathematics curriculum: http://www.edu.gov.on.ca/eng/curriculum/elementary/math18curr.pdf. Five children were considered outliers and not included in the analyses as their scores fell more than 2.5 standard deviations below the mean.

2.3.5. Sentence recall

Sentence recall was included as a measure of children's language skills. This task has also been found to be highly related to children's verbal working memory capacities (Alloway & Ledwon, 2014). Thus, the inclusion of this measure allowed us to determine whether any relations between numerical skills and mathematics remain once children's language/working memory skills are taken into account. The sentences for this task were from Redmond (2005) and consisted of 16 sentences of 10 words each and 10–14 syllables. The sentences were presented to children through headphones of an audio recording of an adult female speaker. The sentences were presented in a fixed order. For each sentence children received either a score of 2 (perfect), 1 (three or fewer errors), or 0 (more than four errors or no response). There were no outliers on this task.

2.3.6. Rapid color naming (RCN)

This task was included as control for children's cognitive processing speed. Given that the Numeracy Screener is a timed task, it is important to rule out the possibility that any observed relations between the screener and math were not purely a result of children's processing speed. Children were presented with 50 items (individual patches of color) and given a score based on the total time (seconds) taken to complete the task. Thus, faster response times (lower scores) indicate better performance. Four children were considered outliers and not included in the analyses as their scores were more than 2.5 standard deviations above the mean.

2.4. Analytical approach

Analyses were carried out using JASP (V. 0.8.1). This software package allowed us to conduct a combination of both frequentist and Bayesian statistics. The Bayesian analyses are useful in that the associated Bayes factors provide an easy-to-interpret degree of strength of evidence for or against any given association. Said differently, Bayes factors can be interpreted as likelihood ratios in support for the alternative hypothesis over the null (BF_{10}; evidence for associations between variables) or, conversely, support for the null hypothesis over the alternative (BF_{01}; evidence against associations between variables). For example, a Bayes factor of 5 ($BF_{10} = 5$) suggests that the alternative hypothesis is five times more likely than the null (i.e., it is five more times likely that a relationship exist than not). Alternatively, in terms of proportions, a Bayes factor of 5 can be interpreted as 5 parts in favor of a relation and 1 part in favor of no relation, which when calculated as a percent (i.e., $5/6 = 0.83 \times 100$) suggests an 83% chance of there being a relation compared to a 17% chance of there not being one. These statistics provide a similar function and practicality as effect sizes but are arguably even more directly interpretable. In the current study, we used the default options in JASP to carry out Bayesian Pearson correlational analyses and Bayesian linear regression analyses. The default prior width is set to 1 for Pearson correlations and 0.354 (prior for r scale covariates) for linear regression analyses. Although Bayes factors provide a degree of evidence that is somewhat open to interpretation (e.g., Should an association that is three times more likely than the null be considered sufficient evidence?), the following recommendations have been outlined as a general guideline for the interpretation of Bayes factors: Bayes factors of 0–3 offer weak support for H_1, 3–20 positive support for the H_1, 20–150 strong support for H_1, and values greater than 150 as very strong evidence in favor of the H_1 (Andraszewicz et al., 2015; Jarosz & Wiley, 2014; Raftery, 1995). We adhere to these guidelines in interpreting the results of the present paper.

3. Results

3.1. Descriptive statistics

Descriptive statistics are presented in Table 3. The column labeled *N* refers to the number of participants after the outliers were removed according the criteria described above. Histograms and scatterplots demonstrated relatively normal distributions of data. Moreover, as can

Table 3
Descriptive statistics of measures.

	N	Mean Score (SD)	Range (min - max)	Kurtosis	Skewness	Reliability
Kindergarten Measures						
Age (years)	439	5.86 (.31)	5.2–6.4	−0.664	−0.026	N/A
Symbolic Comparison	399	37.56 (10.94)	8–56	−0.343	−0.369	.72
Non-symbolic Comparison	392	39.03 (8.10)	14–55	−0.074	−0.324	.61
Arithmetic	397	5.13 (3.07)	0–10	−1.222	0.079	.60
Number Line (PAE)	355	13.51 (6.36)	1–32	−0.249	0.654	.32
Sentence Recall	133	16.10 (8.13)	0–31	−0.288	0.823	.86
Rapid Color Naming	129	65.80 (19.23)	39–125.1	1.595	1.300	.73
1st Grade Measures						
Overall Mathematics Grade	434	8.29 (1.77)	3–12	−0.212	−0.400	N/A

Note: Reliability estimates are based on test-retest correlations. For the symbolic and non-symbolic comparison tasks the highest possible score is 56.

Table 4
Correlations between age, gender, and the various measures.

Variable	1	2	3	4	5	6	7	8	9
1. Age	–								
BF_{10}									
2. Gender	.03	–							
BF_{10}	.07								
3. Sentence Recall	.27*	.19	–						
BF_{10}	14.85	1.06							
4. Rapid Color Naming	-.26*	-.03	-.41***	–					
BF_{10}	9.99	.12	14275.60						
5. Symbolic Comparison	.16*	.09	.41***	-.59***	–				
BF_{10}	11.37	.33	11928.13	1.644E+10					
6. Non-Symbolic Comparison	.18**	.15*	.35***	-.41***	.61***	–			
BF_{10}	37.95	4.90	423.02	7844.86	2.300E+49				
7. Number Line (PAE)	-.19**	.03	-.18	.24*	-.19**	-.15	–		
BF_{10}	40.30	.08	.85	4.23	34.34	2.44			
8. Arithmetic	.34***	-.01	.43***	-.26*	.33***	.23***	-.21***	–	
BF_{10}	3.780E +9	.07	31366.32	7.4	6.307e+7	1115.22	202.01		
9. Overall Maths Grade	.24***	.03	.48***	-.40***	.31***	.23***	-.25***	.31***	–
BF_{10}	13032.44	.07	1.39e+06	4845.36	2.976E+7	2112.13	3838.85	3.578E+7	

* $p < .05$, ** $p < .01$, *** $p < .001$. BF_{10} = Bayes Factors in support of alternative hypothesis over the null. BF_{10} between 0 and 3 = weak support for an association; BF_{10} between 3 and 20 = positive support for an association; BF_{10} between 20 and 150 = strong support for an association; $BF_{10} > 150$ = very strong evidence in favor of an association. Note gender was dummy coded where 0 = males and 1 = females.

be seen in Table 3, kurtosis and skewness values were all within the acceptable limits of ± 2 (Gravetter & Wallnau, 2014; Field, 2009).

3.2. Correlational analyses

Pearson bivariate coefficients were calculated to determine the strength of associations between measures (see Table 4). As shown in Table 4, there was strong evidence ($BFs_{10} > 20$) for relations between the mathematics measures; the only exception being the relation between non-symbolic comparison and number line estimation ($BF_{10} = 2.44$). Importantly, there was very strong evidence ($BFs_{10} > 2000$) of relations between all numerical measures in kindergarten and children's mathematics grades one year or more later (see Fig. 3). The associated Bayes factors were all over 2000 indicating, at a minimum, there was 2000 times more evidence in favor of there being a relation than not.

There was also some evidence of a gender difference on the non-symbolic comparison task; females were 4.9 times more likely to outperform males on this measure. In light of this finding, as well as the correlation between age and performance on several measures, all subsequent analyses include age and gender as covariates.

3.3. Main analyses

3.3.1. Relationship between numerical measures and arithmetic

Our first analysis was an attempt to extend a previous finding that involved using the very same magnitude comparison task but with an older population of first to third grade children (e.g., see Nosworthy et al., 2013). Findings from the Nosworthy et al. (2013) study indicated that symbolic comparison performance was a unique predictor of 6- to 9-year-olds' arithmetic scores while non-symbolic performance was not. To extend this finding with the current data, linear regression analyses were conducted with arithmetic as the outcome variable and children's age, gender, number line estimation (PAE), non-symbolic comparison scores, and symbolic comparison scores entered as simultaneous predictors. Note that listwise deletion was used to treat missing data for this and all other regression analyses; a decision that allowed us to compute Bayes factors on the same dataset using the statistical software JASP (version 0.8.1). The overall regression model was statistically significant, $F(5, 306) = 16.00$, $p < .001$, $R^2 = 0.21$. Table 5 shows the contribution of each variable along with their associated Bayes Factor inclusion values. Age, symbolic comparison, and number line estimation (PAE) all explained statistically significant unique variance on the arithmetic ($ps < .05$). However, the associated Bayes factor between

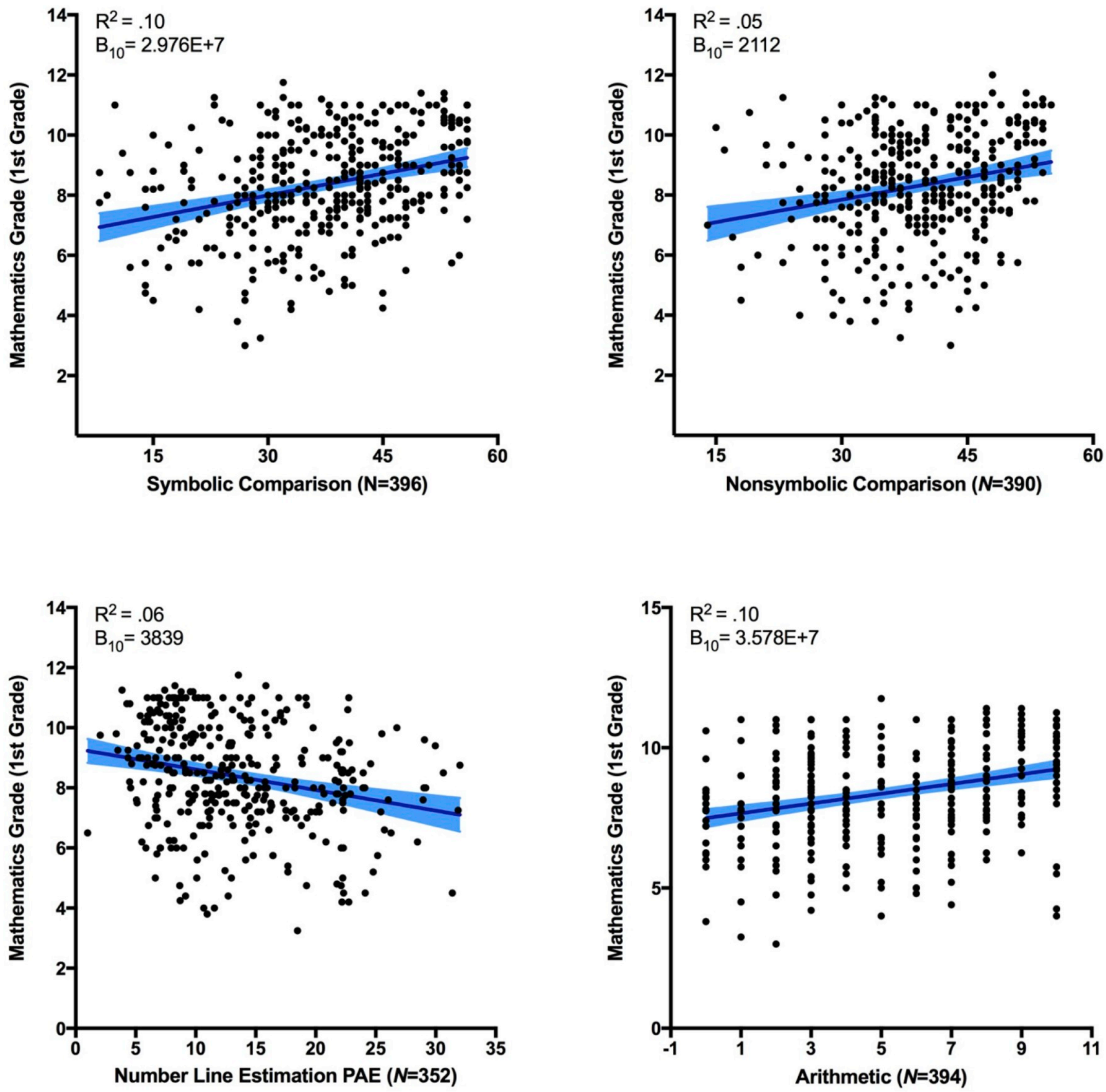

Fig. 3. Scatterplots showing relationship between performance on numerical tasks and children's overall grade in mathematics one year later. The solid dark blue line represents the linear regression line for each relationship. The light blue bands represent 95% confidence bands around the line of best fit. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

number line and arithmetic failed to reveal convincing evidence for a unique relation (BF_{10} = 2.09). There was strong evidence for unique relations between symbolic comparison and arithmetic (BF_{10} = 13.79) and extremely strong evidence for unique relations between age and arithmetic (BF_{10} = 88539.49). Taken together, our findings align with the original findings of Nosworthy et al. (2013).

3.3.2. Longitudinal relations between numerical skills and school-based mathematics

Our next set of analyses addressed the question of how basic number skills in Kindergarten longitudinally relate to school-based mathematics one or more year later. Linear regression analyses were conducted with 1st grade mathematics grades as the outcome variable and children's age, gender, arithmetic, number line estimation, non-symbolic comparison, and symbolic comparison scores as predictor variables. Note that arithmetic, the outcome variable in the previous analysis, was entered as covariate for this and the following analysis in an attempt to control for the effects of mathematical ability at time 1. All variables were entered in one step. The overall model was significant, *F*(6,

Table 5
Regression analyses and Bayes factors explaining variance in kindergarten children's arithmetic performance (N = 312).

Outcome: Arithmetic					
Predictors	β	*t*	*p*	ΔR^2	Bayes factor
Age	.28	5.18	< .001***	.13	88539.49
Gender	-.05	−1.02	.309	.00	.27
Number Line (PAE)	-.12	−2.25	.025*	.01	2.09
Non-symbolic Comparison	.01	1.41	.158	.04	.50
Symbolic Comparison	.17	2.51	.013*	.02	13.79

* $p < .05$, ** $p < .01$, *** $p < .001$. BF_{10} = Bayes Factors in support of alternative hypothesis over the null. BF_{10} between 0 and 3 = weak support for an association; BF_{10} between 3 and 20 = positive support for an association; BF_{10} between 20 and 150 = strong support for an association; BF_{10} > 150 = very strong evidence in favor of an association.

Table 6
Regression analyses and Bayes factors showing relationship between age, gender, and numerical skills at Kindergarten and 1st Grade mathematics grades (N = 310).

Outcome: Math grade					
Predictors	β	t	p	ΔR^2	Bayes Factor
Age	.10	1.88	.061	.07	1.03
Gender	-.03	-.52	.607	.00	.22
Arithmetic	.23	3.94	< .001***	.08	654.22
Number Line (PAE)	-.13	−2.44	.015*	.02	3.96
Non-symbolic Comparison	-.02	-.34	.735	.01	.20
Symbolic Comparison	.23	3.28	.001**	.03	139.07

* $p < .05$, ** $p < .01$, *** $p < .001$. BF_{10} = Bayes Factors in support of alternative hypothesis over the null. BF_{10} between 0 and 3 = weak support for an association; BF_{10} between 3 and 20 = positive support for an association; BF_{10} between 20 and 150 = strong support for an association; $BF_{10} > 150$ = very strong evidence in favor of an association.

303) = 13.34, $p < .001$, $R^2 = 0.21$. As shown in Table 6, arithmetic, number line, and symbolic comparison were all statistically significant unique predictors of 1st grade mathematics grades. The associated Bayes factors revealed positive support for unique relations between number line estimation and mathematics grades ($BF_{10} = 3.96$), strong support for unique relations between symbolic comparison and mathematics ($BF_{10} = 139.07$), and very strong support for unique relations between arithmetic and mathematics ($BF_{10} = 654.22$). These findings indicate that arithmetic, number line estimation, and symbolic comparison skills at kindergarten are unique predictors of children's school grades in mathematics 12–16 months later.

To further determine the strength of the relationship between kindergarten numerical skills and first grade mathematics, we carried out a similar analysis to the above but with the addition of two control variables: sentence recall and rapid color naming. Recall that these data were only available for a subsample ($n = 133$). Including the two control variables resulted in the following model fit: $F(8, 114) = 9.88$, $p < .001$, $R^2 = 0.41$. With the inclusion of these control variables, the unique relations between arithmetic and mathematics and number line and mathematics were no longer statistically significant. However, there was strong support for unique relations between sentence recall and mathematics ($BF_{10} = 18.53$) and very strong support for unique relations between symbolic comparison and mathematics Bayes factor ($BF_{10} = 649.43$). These findings indicate that symbolic comparison skills at kindergarten remains a strong predictor of 1st grade mathematics even when controlling for domain-general (sentence recall) and task-specific cognitive demands, such as speed of processing (rapid color naming). Moreover, symbolic comparison skills appear to explain unique variance in children's mathematics grades to a greater extent than the other numerical tasks, including arithmetic, number line, and non-symbolic comparison.

Table 7
Regression analyses and Bayes factors showing relationship between age, gender, and numerical skills at Kindergarten and 1st Grade mathematics grades controlling for sentence recall and rapid color naming (N = 123).

Outcome: Math grade					
Predictors	β	t	p	ΔR^2	Bayes Factor
Age	.11	1.32	.190	.10	.49
Gender	-.07	-.87	.386	.00	.31
Sentence Recall	.23	2.69	.008**	.15	18.53
Rapid Color Naming	-.02	-.18	.857	.05	.24
Arithmetic	.05	.58	.562	.01	.31
Number Line (PAE)	-.13	−1.69	.095	.02	.76
Non-symbolic Comparison	.01	.08	.936	.03	.21
Symbolic Comparison	.38	3.34	.001**	.06	649.43

* $p < .05$, ** $p < .01$, *** $p < .001$. BF_{10} = Bayes Factors in support of alternative hypothesis over the null. BF_{10} between 0 and 3 = weak support for an association; BF_{10} between 3 and 20 = positive support for an association; BF_{10} between 20 and 150 = strong support for an association; $BF_{10} > 150$ = very strong evidence in favor of an association.

3.3.3. Test-retest analyses

Test-retest correlations were carried out on a subsample of children ($n = 133$) for whom data was collected at two time points, approximately three months apart from one another (Mean days between testing = 89.55 (13.90)). As shown in Table 3, the test-retest correlations and 95% confidence intervals (CIs) for symbolic comparison, non-symbolic comparison, arithmetic, and number line estimation were, 0.72, [0.63, 0.80] 0.61 [0.49, 0.71] 0.60 [0.48, 0.70] and 0.32 [0.15, 0.47], respectively. These results indicate that only the symbolic comparison task achieved the recommended cut-off for satisfactory test-retest reliability (Cohen & Swerdlik, 2009).

4. Discussion

The present study investigated the concurrent, predictive, and incremental validity of a two-minute paper-and-pencil Numeracy Screener. The tool was developed primarily for teachers – but also researchers – as a relatively quick and easy-to-administer assessment of young children's basic magnitude processing. In prior research with the measure, both symbolic and non-symbolic comparison skills were correlated with children's (Grades 1–3) standardized arithmetic scores. However, only performance on the symbolic comparison task was uniquely related to arithmetic once additional control variables, such as working memory and vocabulary, were taken into account (Nosworthy et al., 2013). The current study sought to build on and extend these findings in several key regards. First, we were interested in testing the extent to which performance on the measure in kindergarten (prior to formal schooling) concurrently relates to arithmetic. In other words, could we replicate our initial findings with a group of children who had not yet received formal mathematics instruction? Second, we aimed to test the predictive and incremental validity of the measure by examining how performance in kindergarten predicts children's school grades in mathematics 12–16 months later, even after controlling for potentially confounding variables (e.g., number line estimation, language skills, processing speed). Third, and finally, we evaluated the test-retest reliability of the measure.

4.1. Links between magnitude processing skills and arithmetic

With respect to our first goal, we found evidence for a unique relation between performance on the symbolic comparison task and a simple paper-and-pencil measure of arithmetic. Performance on both the non-symbolic comparison task and number line estimation were not uniquely related to children's arithmetic performance. This result adds further support for the concurrent validity of the symbolic comparison task as it loosely replicates our original findings with the measure (Nosworthy et al., 2013). This finding also suggests that the relation between symbolic comparison and arithmetic is present prior to school entry. An important line of future inquiry will be studying what gives rise to these individual differences in children's symbolic mathematics understanding prior to formal schooling. For example, an emerging body of research points to home numeracy practices as a key contributor to young children's symbolic math knowledge before and during the pre-school years (Gunderson & Levine, 2011; LeFevre et al., 2009). A better understanding of how home and informal schooling supports the development of children's symbolic numerical skills is critical, as this information is potentially useful in its application with children for whom this same support is not available.

It is also noteworthy that performance on the symbolic comparison task, but not the non-symbolic task, was significantly correlated with children's number line estimation performance. Children who performed better on the symbolic comparison task also demonstrated more precise estimates of where the Arabic digits, 1–9, belong on a horizontal number line. This result provides some evidence of construct validity, as the number line estimation task is a widely used measure of numerical ability. However, as discussed in greater detail further below, symbolic comparison skills were found to be a stronger predictor of arithmetic and school mathematics achievement. Thus, although the two tasks are correlated, and may even tap a common magnitude processing system (e.g., see Siegler, 2016), our results suggest that two tasks are differentially related to mathematics achievement.

4.2. Links between magnitude processing skills and mathematics grades

An important, and yet rarely explored question, is the extent to which basic numerical skills predict children's school grades in mathematics. To date, the link between basic numerical skills and mathematics performance has almost exclusively focused on standardized math outcome measures, most notably those that emphasize arithmetic and calculations. (e.g., see our previous work with the Numeracy Screener, Nosworthy et al., 2013). While there are many advantages of using standardized mathematics assessments, including the establishment of empirically validated psychometric properties (e.g., good test-retest statistics), there are also some notable advantages to using teacher-assigned grades. For example, teacher-assigned grades in mathematics have the added advantage in that they represent an ecologically valid measure of student mathematics achievement, providing a comprehensive measure of mathematics skills (e.g., numeration, geometry, measurement, algebra, etc.) across varied and prolonged opportunities for student assessment (as opposed to a single brief assessment). Furthermore, given that a central aim of ours is to provide educators with a quick and easy-to-administer numerical assessment tool, it is important to uncover how performance on the measure relates to school grades in mathematics. For these reasons, teacher assigned grades in mathematics served as our primary outcome variable. Our analyses revealed that individual differences in kindergarten students' symbolic comparison skills and number line estimation performance both uniquely predicted children's 1st grade mathematics grade. There was an especially strong relation between symbolic comparison and children's mathematics grade. Bayesian analyses indicated the observed relation to be 3754 more times likely than there not being one, even after taking into account the influence of the other variables, such as age and non-symbolic comparison skills. Comparatively, the likelihood of there being a unique relation between number line performance and mathematics grades was valued at approximately 17 times more likely than not. Interestingly, as was the case with arithmetic, children's non-symbolic comparison skills were not found to predict children's mathematics grades. Taken together, these findings point to children's symbolic comparison and number line estimation skills as important independent and longitudinal predictors of school mathematics.

4.3. Incremental validity of the Numeracy Screener

In an attempt to further test the predictive and incremental validity of the Numeracy Screener, we carried out a follow-up analyses with a subsample of children for whom we had collected measures of language and processing speed. The inclusion of these variables allowed us to tentatively rule out the influence of domain-general (language/working memory) and task-specific (i.e., processing speed) cognitive demands which may be viewed as potentially confounding variables in the relations reported above. As expected, performance on these measures were strongly related to mathematics achievement (see Table 4) and thus provided initial support for their role as potentially moderating variables. Even with the inclusion of these variables, we found strong evidence for longitudinally predictive relations between performance on the symbolic comparison task and 1st grade mathematics achievement. Indeed, the likelihood of there being a relation was valued at 472 times more likely than there not being a relation. Interestingly, the relations between number line estimation and mathematics reported above was no longer statistically present in these analyses.

Of additional interest is the finding that symbolic comparison skills but not arithmetic continue to predict first grade math achievement after controlling for sentence recall and processing speed. This raises the question, what is so special about symbolic comparison that is not captured by arithmetic? It is possible that the arithmetic task may have relied more on verbal working memory than symbolic comparison, which is perhaps why we see the contribution of arithmetic substantially diminished with the inclusion of sentence recall. Support for this possibility comes from a rather large body of research showing strong relations between children's verbal working memory and arithmetic performance (e.g., De Smedt, Verschaffel, & Ghesquière, 2009; Friso-; Raghubar, Barnes, & Hecht, 2010; Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015; van den Bos, van der Ven, Kroesbergen, & van Luit, 2015). Future research efforts are needed to further test this possibility.

In sum, our findings provide additional support for longitudinal relations between symbolic numerical comparison skills and future mathematics achievement (e.g., Bartelet et al., 2014; De Smedt et al., 2009; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Xenidou-Dervou et al., 2017). Importantly, while other studies have used computerized lab-based assessments to uncover this relation, the current study demonstrates comparable findings but with the use of a two-minute paper-and-pencil measure (also see Brankaer et al., 2017). Furthermore, this is the first longitudinal study of which we are aware of that shows the association also generalizes to teacher assigned mathematics grades.

4.4. Test-retest reliability

A final objective of ours was to evaluate the test-retest reliability of the numerical measures included in the study. Of the four numerical measures evaluated, only the symbolic comparison measure met the recommended test-retest cut-off of 0.65 (Cohen & Swerdlik, 2009). Interestingly, our finding of relatively stable symbolic number comparison performance and less stable performance on the non-symbolic and number line task is consistent with prior research (Kolkman, Kroesbergen, & Leseman, 2013; Inglis & Gilmore, 2014). While the findings above speak to the validity of the symbolic comparison task, the test-retest results also indicate that the measure is reliable over time. Clearly, this is a desirable property of a measure that is intended to be used in the classroom and over multiple time points (e.g., at the beginning and end of the school year).

4.5. Interpretations, implications and next steps

4.5.1. Symbolic number comparison vs. non-symbolic comparison

Our findings indicate that children's symbolic comparison and, to a lesser degree, children's number line estimation skills at kindergarten are concurrent predictors of arithmetic and longitudinal predictors of overall mathematics achievement in 1st grade. Contrary to previous findings (e.g., Halberda et al., 2008), we found no evidence of unique relations between children's non-symbolic comparison skills and mathematics achievement. These findings, however, do not preclude the possibility that non-symbolic numerical skills may play a role in children's symbolic mathematical development. Indeed, it is possible, and some research does suggest it may be the case, that non-symbolic comparison skills play a fundamental role in grounding the meaning of symbolic number earlier in development (Mundy & Gilmore, 2009; Purpura, Baroody, & Lonigan, 2013). Future research is needed to test this possibility by assessing relations between performance on the non-symbolic comparison task and mathematics prior to kindergarten entry.

4.5.2. Symbolic number comparison vs. number line estimation

An interesting question that emerges from this research is in understanding why performance on the symbolic comparison task was a consistently more reliable predictor of mathematics than number line estimation. Based on a review of the literature, it appears as though the relations between number line estimation and mathematics is higher than those typically reported between symbolic comparison and mathematics (Schneider et al., 2018; Siegler & Booth, 2004). One reason for the relatively stronger contributions of symbolic comparison skills and mathematics might have to do with the poor reliability of the number line measure observed in the current study. Future research is needed that more fully evaluates the reliability of number line task performance. Furthermore, the relatively higher cognitive demands of the number line task compared to the symbolic comparison task might make it a less consistent and reliable predictor amongst young children. For example, prior research indicates that in addition to magnitude and ordinal processing requirements, performance on number line tasks may involve proportional reasoning (Barth & Paladino, 2011), spatial reasoning (Gunderson, Ramirez, Beilock, & Levine, 2012), and visuo-motor coordination (Simms, Clayton, Cragg, Gilmore, & Johnson, 2016). Paradoxically, it may be the recruitment and integration of these other cognitive skills and the use of more sophisticated strategies that contributes to poorer number line–math relations amongst young children but stronger relations amongst older children and adults (e.g., see Schneider et al., 2018). Thus, it is possible that some young children might struggle with the task not necessarily as a result of limited knowledge of numerical magnitudes but due to the complexities of the task. Future research efforts are needed to better understand potential interactions between symbolic comparison and number line estimation throughout development and to further disentangle when and whether each task is a better predictor of children's mathematics achievement. Based on the current study, however, it appears as though a brief assessment of children's symbolic number comparison is a stronger and more reliable predictor of future mathematics achievement in a sample of kindergarten students.

4.5.3. Symbolic number knowledge as a focus of early mathematics instruction

In terms of implications, our findings suggest the importance of attending to, assessing, and developing young children's symbolic number knowledge even prior to school entry. As others have found as well (Bartelet et al., 2014; Sasanguie, De Smedt, et al., 2012), children who enter formal schooling with stronger symbolic number knowledge are more likely to succeed in mathematics. Presumably, this is due to the hierarchical nature of mathematics, where earlier learned concepts and skills are needed to give rise to new and more advanced understandings. For example, the practical importance of being able to quickly access numerical symbols can be observed across a range of mathematical tasks, including arithmetic, where understanding the numerical magnitudes being manipulated is foundational to both estimating and correctly solving the given problem. It is difficult, if not impossible, for example, to estimate that 143 + 153 is about 300 without an understanding of the numerical magnitude of the two addends. Moreover, the development of increasingly more advanced and abstract arithmetic strategies depends on a child's ability to instantly recognize the numerical magnitudes to be manipulated (i.e., the underlying quantity). In the question 3 + 7, for instance, a child who can instantly recognize 7 as the larger addend can then use this knowledge to apply the "counting on" strategy; that is, the child starts with 7 and counts on three more, "8, 9, 10" (Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Carpenter, Fennema, Peterson, & Carey, 1988). In short, rapid access to symbolic numerical magnitudes may serve as a scaffold for mathematics learning and development.

Recent research efforts indicate the potential benefits of developing young children's symbolic number knowledge through training studies. For example, early interventions targeting both symbolic comparison skills (e.g., see Honoré & Noël, 2016; Scalise, Daubert, & Ramani, 2017) as well numerical linear board games (e.g., see Ramani & Siegler, 2008; Siegler & Ramani, 2008) have been found effective at improving not only children's performance on the trained task but also higher level mathematics, such as arithmetic. More generally, both symbolic number comparison and number line estimation have been prominent features of successful early years mathematics interventions (e.g., see Number Worlds; Griffin, 2004). Taken together, finding and creating motivating and meaningful ways to engage *all* young children –even prior to school entry– in the learning of numerical symbols and their relations is a desirable goal.

4.6. Limitations

As reported above, we see much value in using children's school grades in mathematics as a measure of mathematics achievement. However, some caution is warranted as teacher-assigned grades are not standardized. That is, there is no uniform way of assessing students' mathematics achievement and as a result, teacher assessment of student mathematics performance is likely to vary somewhat depending on teacher. With that said, the Ontario curriculum has clear guidelines of what children are expected to demonstrate in mathematics at each grade level. Furthermore, our finding that teacher-assigned mathematics grades strongly correlated with all numerical tasks, including arithmetic, lends further support to the validity and use of school math grades as an outcome measure. Ideally, future research should aim to use both school grades and standardized mathematics measures as indicators of children's mathematics achievement. This would allow for a more complete picture of the individual learner.

With the exception of the symbolic number comparison task, test-retest coefficients were suboptimal and may have influenced the outcomes of the study. Future research efforts are needed to more carefully examine the reliability of basic numerical measures during this period of child development. It is possible that young children demonstrate greater fluctuations in task performance than older children potentially as a result of transitioning from informal to formal education. The large amount of missing data, especially on the number line task, is also a limitation of the current study and may have further affected the reliability of the selected measures. Future research efforts are needed to better understand why certain children were unable to complete various measures.

An additional limitation of the current study is the potential confound of motor-control issues related to the paper-and-pencil response format. Given the young age of participants, it is possible that fine-motor skills may have affected the results. Moreover, dyscalculia and fine-motor difficulties have been shown to co-occur (e.g., see Pieters, Desoete, Roeyers, Vanderwalmen, & Van Waelyelde, 2012). Future research efforts are needed to examine the potential influence of fine-motor skills on the Numeracy Screener.

Although we revealed statistically significant relations and correspondingly strong support for relations between symbolic comparison and mathematics as indicated through Bayes factors, the unique variance explained was quite small. For example, the final model (Table 7) only explained 41% of the total variance in mathematics, of which 6% was unique to symbolic comparison. This finding suggest that more comprehensive models are needed to more fully account for children's mathematics achievement (e.g., see LeFevre et al. (2010) 'Pathways to mathematics model'). It is also important to consider affective (socio-emotional) factors in working towards a more comprehensive model of mathematical development (e.g., attitudes and anxiety).

5. Conclusion

Our findings indicate that a two-minute paper-and-pencil measure of children's symbolic number comparison is a reliable measure and demonstrates concurrent, predictive, and incremental validity. The task was shown to predict children's arithmetic as well as children's future school mathematics achievement over and above other measures. Thus, this freely available task provides one means for educators and researchers to assess and better understand young children's symbolic number processing. Ultimately, it is our hope that the Numeracy Screener will prove useful in the planning and delivery of classroom mathematics instruction.

Author note

Zachary Hawes, Department of Psychology, Brain and Mind Institute, University of Western Ontario; Nadia Nosworthy, Department of Graduate Psychology and Counseling, Andrews University; Lisa Archibald, School of Communication Sciences and Disorders, Department of Psychology, University of Western Ontario; Daniel Ansari, Department of Psychology, Brain and Mind Institute, University of Western Ontario.

Declarations of interest

None.

Funding

This work was supported by operating grants from the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), the Canada Research Chairs Program, an E.W.R Steacie Memorial Fellowship from the Natural Sciences and Engineering Council of Canada (NSERC) to DA as well as Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship to ZH.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.learninstruc.2018.09.004.

References

Alloway, T. P., & Ledwon, F. (2014). Working memory and sentence recall in children. *International Journal of Educational Research, 65*, 1–8.
Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E. J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management, 41*(2), 521–543.
Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology, 117*, 12–28.
Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*(1), 125–135.
Beaver, J. (1968). *Developmental reading assessment.* Toledo, OH: Celebration Press.
Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016–1031.
Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP test). *Behavior Research Methods, 49*(4), 1361–1373.
Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology: Section A, 54*(4), 1005–1029.
Carpenter, T. P., Fennema, E., Peterson, P. L., & Carey, D. A. (1988). Teachers' pedagogical content knowledge of students' problem solving in elementary arithmetic. *Journal for Research in Mathematics Education,* 385–401.
Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148*, 163–172.
Cohen, R. J., & Swerdlik, M. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York, NY: McGraw-Hill.
De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*(2), 278–292.
De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48–55.
De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103*(4), 469–479.
Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* New York, NY: Oxford University Press.
Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.
Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44.
Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., et al. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology, 46*(6), 1731–1746.
Gebuis, T., & Reynvoet, B. (2012). The interplay between non-symbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642–648.
Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills*(6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
Gravetter, F. J., & Wallnau, L. B. (2014). Introduction to the t statistic. *Essentials of Statistics for the Behavioral Sciences, 8*, 252.
Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties.* London: Every Child a Chance Trust and KPMG.
Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's number knowledge. *Developmental Science, 14*(5), 1021–1032.
Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48*(5), 1229–1241.
Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668.
Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17–29.
Honoré, N., & Noël, M. P. (2016). Improving preschoolers' arithmetic through number magnitude training: The impact of non-symbolic and symbolic training. *PLoS One, 11*(11), e0166685.
Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica, 145*, 147–155.
Iuculano, T., Tang, J., Hall, C. W., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*(5), 669–680.
Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1), 2–9.
Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95–103.
Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology, 103*(4), 546–565.
Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*(6), 1723–1743.
LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development, 81*(6), 1753–1767.
LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science, 41*(2), 55–66.
Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From 'sense of number' to 'sense of magnitude'–The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences,* 1–62.
Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, e164.
Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300.
Lourenco, S. F., & Bonny, J. W. (2017). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental Science, 20*(4), https://doi.org/10.1111/desc.12418.
Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS One, 11*(3), e0149863.
Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237.

Mou, Y., Li, Y., Hoard, M. K., Nugent, L. D., Chu, F. W., Rouder, J. N., et al. (2016). Developmental foundations of children's fraction magnitude knowledge. *Cognitive Development, 39*, 141–153.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and non-symbolic representations of number. *Journal of Experimental Child Psychology, 103*(4), 490–502.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper- and-pencil test of symbolic and non-symbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS One, 8*(7), e67918.

Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends in Cognitive Sciences, 21*(6), 409–424.

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* National Research and Development Centre for Adult Literacy and Numeracy, Institute of Education.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41.

Pieters, S., Desoete, A., Roeyers, H., Vanderswalmen, R., & Van Waelvelde, H. (2012). Behind mathematical learning disabilities: What about visual perception and motor skills? *Learning and Individual Differences, 22*(4), 498–504.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*(2), 453–464.

Purpura, D. J., & Lonigan, C. J. (2015). Early numeracy assessment: The development of the preschool early numeracy scales. *Early Education & Development, 26*(2), 286–313.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology,* 111–163.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*(2), 110–122.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*(2), 375–394.

Redmond, S. M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics and Phonetics, 19*(2), 109–127.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*(7), 1301–1308.

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361–395.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012a). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*(2), 344–357.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*(3), 418–431.

Sasanguie, D., & Reynvoet, B. (2013). Number comparison and number line estimation rely on different mechanisms. *Psychologica Belgica, 53*(4), 17–35.

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012b). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain, and Education, 6*(3), 119–128.

Scalise, N. R., Daubert, E. N., & Ramani, G. B. (2017). Narrowing the early mathematics gap: A play-based intervention to promote low-income preschoolers' number skills. *Journal of Numerical Cognition, 3*(3), 559–581.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*(3), https://doi.org/10.1111/desc.12372.

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101*(2), 359.

Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., et al. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development.* https://doi.org/10.1111/cdev.13068.

Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2018). Associations of magnitude comparison and number line estimation with mathematical competence: A comparative review. In P. Lemaire (Ed.). *Cognitive development from a strategy perspective: A festschrift for Robert S. Siegler (Chapter 7)*London: Psychology Press (in press).

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science, 19*(3), 341–361.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428–444.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science, 11*(5), 655–661.

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology, 145*, 22–33.

Vanbinst, K., Ceulemans, E., Peters, L., Ghesquière, P., & De Smedt, B. (2018). Developmental trajectories of children's symbolic numerical magnitude processing skills and associated cognitive competencies. *Journal of Experimental Child Psychology, 166*, 232–250.

Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. (2015). Verbal and visual- spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition, 43*(3), 367–378.

Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. (2017). Non-symbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction, 50*, 1–13. https://doi.org/10.1016/j.learninstruc.2016.11.001.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11.