



# Numerical estrangement and integration between symbolic and non-symbolic numerical information: Task-dependence and its link to math abilities in adults

Xueying Ren<sup>a,b,c,\*</sup>, Ruizhe Liu<sup>a,b,c</sup>, Marc N. Coutanche<sup>a,b,c,d</sup>, Julie A. Fiez<sup>a,b,c</sup>, Melissa E. Libertus<sup>a,b,c</sup>

<sup>a</sup> Department of Psychology, University of Pittsburgh, Pittsburgh 15260, PA, USA

<sup>b</sup> Learning Research and Development Center, University of Pittsburgh, Pittsburgh 15260, PA, USA

<sup>c</sup> Center for the Neural Basis of Cognition, Pittsburgh 15260, PA, USA

<sup>d</sup> Brain Institute, University of Pittsburgh, Pittsburgh 15260, PA, USA

## ARTICLE INFO

### Keywords:

Symbolic number  
Non-symbolic number  
Numerical estrangement  
Numerical integration  
Math  
Individual differences

## ABSTRACT

Most adults have access to two different number systems to represent numerical information: an exact number system, which relies on different forms of number symbols to represent exact numerical information, and an approximate number system, which allows for approximate estimates of numerical quantities. Here we investigate the integration between the symbolic and non-symbolic numerical information (i.e., “numerical integration”), and how numerical integration relates to adults’ formal math abilities. We administered two tasks to measure numerical integration. For a number comparison task with non-symbolic dot arrays and Arabic numerals, participants indicated the larger of two sequentially presented stimuli that were same-format (dot-dot or numeral-numeral), or mixed-format (dot-numeral or numeral-dot). For a number-letter discrimination task, participants identified Arabic numerals or letter pairs that co-occurred with dot arrays (matching or mismatching the quantity represented by the numeral). In the number comparison task, participants were significantly slower when comparing mixed-format stimuli, especially when Arabic numerals were presented first and dot arrays second, suggesting estrangement between symbolic and non-symbolic numerical information and an asymmetry depending on the order in which the numerical information is presented. In contrast, in the number-letter discrimination task, participants were significantly faster in number-letter discrimination for matching dot arrays and numerals, suggesting integration between symbolic and non-symbolic numerical information. Surprisingly, some measures of numerical estrangement derived from the number comparison task significantly correlated with adults’ performance on a standardized math assessment. Thus, we conclude that numerical integration or estrangement is task-dependent, and adults with greater levels of symbolic estrangement tend to have higher math skills.

## 1. Introduction

Most adults have access to two different systems to represent numerical information: an exact number system that relies on different forms of number symbols (e.g., Arabic numerals, number words) to represent exact numerical information, and an approximate number system that represents imprecise numerical magnitudes from non-symbolic stimuli. The exact number system requires explicit exposure to and instruction in the use of a symbolic number system. For example, the more parents talk about numbers with their toddlers, the better their

later understanding of the cardinality principle, i.e., that the last word in the count sequence refers to the total number of objects in the set that was counted (Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010).

In contrast, the approximate number system (ANS) seems to be present from birth (Izard, Sann, Spelke, & Streri, 2009). Its precision is typically measured with non-symbolic number comparison tasks. In these tasks, participants choose which member of a dot array pair contains the larger number of dots while controlling for perceptual information that often correlates with numerical information (e.g., density,

\* Corresponding author at: 3420 Forbes Avenue, Fifth floor, Pittsburgh, PA, USA.

E-mail address: [xur1@pitt.edu](mailto:xur1@pitt.edu) (X. Ren).

<https://doi.org/10.1016/j.cognition.2022.105067>

Received 3 July 2021; Received in revised form 20 January 2022; Accepted 15 February 2022

Available online 24 February 2022

0010-0277/© 2022 Elsevier B.V. All rights reserved.

area, convex hull; Dietrich, Huber, & Nuerk, 2015). Behavioral performance on these non-symbolic number comparison tasks is typically dependent on the ratio between the numbers. Specifically, when the ratio is closer to 1 (e.g., 15 dots vs 16 dots, a 15:16 ratio), participants tend to respond slower and less accurately than when the numbers are at distant ratios (e.g., 15 dots vs 30 dots, a 1:2 ratio).

What might be the relation between symbolic and non-symbolic number systems? According to the ANS mapping account, one acquires the meaning of symbolic numbers by mapping them onto approximate non-symbolic magnitude (Dehaene, 2001; Piazza, 2010). This notion is supported by evidence that similar behavioral effects (Defever, Sasanguie, Gebuis, & Reynvoet, 2011; Holloway & Ansari, 2009) and brain activation patterns (Cantlon et al., 2009; Dehaene, Izard, & Piazza, 2005; Eger, Sterzer, Russ, Giraud, & Kleinschmidt, 2003) are observed when processing non-symbolic and symbolic numbers. An alternative hypothesis posits that small numerical symbols such as number words are first mapped onto an object tracking system (OTS), a precise representation with a limited capacity of up to four objects (Carey, 2011). Then, with increasing knowledge of the count list, one might use that knowledge to infer principles about number relations such that numbers coming later in the count list are larger (Reynvoet & Sasanguie, 2016). This process gradually results in a separate system for symbolic numbers where symbolic numbers are represented through order associations with other symbolic numbers.

While the exact and approximate number systems are thought to distinctly represent different forms of numerical information, previous research has yielded mixed results regarding the extent to which these systems are integrated in adulthood (Liu, Schunn, Fiez, & Libertus, 2015; Liu, Schunn, Fiez, & Libertus, 2018; Lyons, Ansari, & Beilock, 2012; Schneider et al., 2017). More importantly, even though previous studies have found that both symbolic and non-symbolic number processing are related to math abilities (e.g., Libertus, Odic, & Halberda, 2012; Sasanguie, Lyons, De Smedt, & Reynvoet, 2017; Schneider et al., 2017), the way in which numerical integration between non-symbolic and symbolic numerical information relates to adults' formal math skills remains largely uncharacterized. Therefore, in this study, we implemented a mixed-format number comparison task and a number-letter discrimination task to investigate the integration between symbolic and non-symbolic numerical information (henceforth referred to as "numerical integration"), as well as a standardized math assessment to test how numerical integration relates to adults' formal math abilities.

### 1.1. Evidence for and against numerical integration using comparison tasks

Similar to the non-symbolic number comparison tasks described above, symbolic number comparison tasks have participants identify the larger of two Arabic numerals (or other symbolic number formats). In general, participants are slower at comparing two symbolic numbers with small numerical distance (e.g., judging 6 is smaller than 7) than with large numerical distance (e.g., judging 6 is smaller than 9). This is known as the distance effect (Defever et al., 2011; Sasanguie, De Smedt, Defever, & Reynvoet, 2012). Both distance and ratio effects are typically explained by the overlapping magnitude representations on a mental number line. On this mental number line, magnitudes are represented in a Gaussian distribution such that the closer the numbers are on this mental number line, the harder they are to discriminate. Thus, the similarity of the distance and ratio effects observed in symbolic and non-symbolic number processing tasks has been taken as evidence for an integration between exact and approximate numerical representations such that non-symbolic number representations are employed during numerical comparisons, even when the numerical information is symbolic (Dehaene, 2001; Dehaene & Akhavan, 1995).

Other evidence suggests that the two number systems might not be so tightly integrated (e.g., Lyons et al., 2012; Sasanguie, De Smedt, & Reynvoet, 2017). For instance, Lyons et al. (2012) asked adults to

compare numbers presented either in symbolic formats (Arabic numerals), non-symbolic format (dot arrays), or mixed formats (dots vs numerals). The authors argued that if the symbolic and non-symbolic numerical information were indeed integrated, mixed-format comparisons should result in comparable accuracy and response time to same-format comparisons. However, they found significantly higher response time and lower accuracy in mixed-format comparisons relative to same-format comparisons. The authors reasoned that additional cognitive effort was needed in the mixed-format comparison likely due to a lack of integration between non-symbolic and symbolic number representations, i.e., an estrangement between non-symbolic and symbolic number representations. A switch cost for mixed-format trials has also been found in an audio-visual comparison paradigm where participants indicated the numerically larger of two stimuli when presented with spoken number words, tone sequences, Arabic numerals or dot arrays (Marinova, Sasanguie, & Reynvoet, 2018). In a follow-up study, Marinova, Sasanguie, and Reynvoet (2021) manipulated three experimental factors (the number range, the ratio difficulty, and the presentation modality) in this audio-visual comparison task and found ratio effects in all tasks containing non-symbolic number stimuli, but not in the task containing symbolic numbers only, and a switch cost was also observed for mixed-format conditions. These findings thus further support two distinct number processing systems that are not tightly integrated.

### 1.2. Implicit numerical integration

The comparison tasks used by Lyons et al. (2012) explicitly asked participants to process the two formats of number to make the comparison. However, making explicit magnitude-based judgements in this comparison task may force the translation of symbolic number representations into non-symbolic ones or vice versa, leading to increased response time and reduced accuracy in the mixed-format comparisons. To explore whether numerical integration is evident without explicit magnitude comparisons, Liu et al. (2015) implemented a number-letter discrimination task. In this task, adult participants were asked to decide whether two-item symbol strings were composed of Arabic numerals or letters, with the stimuli superimposed on dot arrays designed to match or mismatch quantities denoted by the numeral strings. Importantly, the dot array (including its quantity) was irrelevant for completing the task. Nevertheless, participants responded more accurately and faster when the Arabic numerals matched (versus mismatched) the dot quantities, suggesting that numerical integration occurs between non-symbolic and symbolic numbers even when the task does not require decisions about number magnitude or the non-symbolic number to be processed. In a follow-up experiment using event-related potentials (ERPs), adult participants passively viewed the same images as in the above-mentioned study (Liu et al., 2018). The amplitude of the N1, an ERP component linked to number processing, was greater for matching than mismatching dot quantities and Arabic numerals. This suggests that the human brain readily integrates non-symbolic and symbolic number representation even in the absence of a task that requires magnitude judgements.

### 1.3. The link between symbolic and non-symbolic number processing and math abilities

Another important question in math cognition that has yet to be fully answered is whether number processing in symbolic, non-symbolic, or both formats is crucial for formal math achievement, especially in adults, and earlier findings are mixed. On the one hand, symbolic number knowledge has been consistently found to be correlated with formal math performance (see De Smedt, Noël, Gilmore, & Ansari, 2013 for a review). For example, Castronovo and Göbel (2012) found that adults with greater math achievement showed faster and more accurate performance in symbolic number comparisons. Moreover, Lyons and

Beilock (2011) found that symbolic number-ordering ability significantly predicted adults' complex mental arithmetic performance. In their study, adults identified whether triads of Arabic numerals (ranging from 1 to 9) were all in increasing order from left to right regardless of the numerical distance between numbers (e.g., "1, 2, 3" increases as does "1, 2, 5"). Participants' mental-arithmetic performance was evaluated using four different mental arithmetic tasks (i.e., addition, subtraction, multiplication, and division). The symbolic number-ordering and mental arithmetic task performance was correlated, even when controlling for numeral identification, performance on symbolic and non-symbolic number comparison tasks, and performance on letter ordering and working memory tasks. Finally, in a meta-analysis, Schneider et al. (2017) also found consistent associations between symbolic number processing and mathematical competence in children and adults.

On the other hand, support for a link between non-symbolic number processing and math abilities, especially in adults, is more mixed (Braham & Libertus, 2018; Libertus et al., 2012; Park & Brannon, 2013; Price, Palmer, Battista, & Ansari, 2012). For example, Libertus et al. (2012) found a positive association between precision of the approximate number system and performance on the quantitative portion of the Scholastic Aptitude Test (SAT), a standardized college entrance exam. To measure precision of non-symbolic number representations, college students completed a non-symbolic number comparison task, in which they decided whether there were more blue or yellow dots in a visual display. A robust correlation between precision of the approximate number system and quantitative scores on the SAT was found even when controlling for performance on the verbal portion of the SAT. In their meta-analysis, Schneider et al. (2017) also found a significant association between non-symbolic number processing and mathematical competence in children and adults, albeit a weaker one than observed between symbolic number processing and mathematical competence.

However, some other studies have not found a correlation between non-symbolic number processing and math achievement in adults (Castronovo & Göbel, 2012; Inglis, Attridge, Batchelor, & Gilmore, 2011; Price et al., 2012). For instance, Price et al. (2012) used three different methods to present stimuli in a non-symbolic number comparison task (i.e., simultaneously presented dot arrays that were either intermixed or spatially separated, and sequentially presented dot arrays) to assess the precision of participants' approximate number system. They did not find correlations between the performance on any of the three versions of the non-symbolic number comparison task and a measure of math fluency (Woodcock Johnson Math Fluency subtest, Woodcock, McGrew, Mather, et al., 2001).

One possible explanation for the inconsistencies in the literature regarding the link between the approximate number system and math abilities may be that it is mediated by the integration between symbolic and non-symbolic numerical information. Earlier studies have found that greater precision in mapping of non-symbolic to symbolic numerical representations positively correlates with children's and adolescents' math achievement (De Smedt, Verschaffel, & Ghesquière, 2009; Holloway & Ansari, 2009; Mazzocco, Feigenson, & Halberda, 2011). In other words, relatively weak integration between non-symbolic and symbolic numerical information may result in noisier mappings between them, which may lead to greater difficulty and less efficiency in accessing the exact representations of numerical information when solving math problems (Holloway & Ansari, 2009). Consistent findings have also been reported in adults, whose symbolic number-ordering ability mediates the link between approximate number system precision and complex mathematical skills (Lyons & Beilock, 2011). Moreover, prior work has suggested that the integration between symbolic and non-symbolic numerical information in adults may be supported by greater engagement of the parietal lobe compared to children who rely more heavily on frontal areas. This age-related shift to greater parietal engagement may reflect the maturation of underlying number representation and increasing flexibility in mapping between numerical symbols and the

magnitudes they represent (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005).

#### 1.4. The current study

As reviewed above, there are uncertainties in the literature regarding the integration between symbolic and non-symbolic numerical information, and to what extent this numerical integration relates to adults' formal math abilities. Therefore, the goals of the current study were twofold: 1) to determine whether evidence for numerical integration is task dependent, and 2) to explore how numerical integration relates to adults' math abilities. To this end, we administered a number comparison task previously used by Lyons et al. (2012) and a number-letter discrimination task previously employed by Liu et al. (2015) to the same group of adult participants. There are two main differences between our comparison task and the comparison task used by Lyons et al. (2012). First, they used a blocked design in which participants were aware of the format of the upcoming stimulus set, whereas our trial types were randomly intermixed such that participants did not know whether a trial was a same- or mixed-format trial ahead of time. Second, in their analyses, Lyons and colleagues did not separate performance based on presentation order of the mixed trials (i.e., dot first or numeral first). Instead, we examined participants' performances on mixed trials separately based on the order in which different formats are presented (i.e., dot first vs Arabic numeral first). The number-letter discrimination task was identical to the one used by Liu et al. (2015). Even though these two conceptually different tasks measure numerical integration differently, we hypothesized that their indices of numerical integration would correlate. Additionally, we administered a standardized assessment of math ability (Woodcock Johnson III Tests of Achievement, Woodcock et al., 2001), and predicted that the two indices of numerical integration would correlate with adults' math abilities.

## 2. Methods

### 2.1. Participants

One hundred twenty-two adults participated in this study (77 female, age range: 18–35 years, mean age ( $M_{age}$ ) = 23.39 years, standard deviation ( $SD$ ) = 4.72). Participants were recruited from the Pittsburgh community, and written informed consent was obtained from all participants prior to completing any research activities as approved by the local Institutional Review Board. All participants were native English speakers with normal or corrected-to-normal vision. Participants received monetary compensation for their participation. We excluded data from seven participants from the final analyses due to incomplete data: five participants were excluded due to a programming error and two participants failed to complete all tasks, which resulted in a final sample of 115 participants (73 female,  $M_{age}$  = 23.5) for the analyses described below.

### 2.2. Stimuli and tasks

#### 2.2.1. Number comparison task

In the number comparison task, participants decided which one of two sequentially presented stimuli represented the larger quantity. To generate the stimuli, 24 symbolic numbers/dot quantities (14, 18, 19, 23, 24, 27, 29, 31, 32, 34, 35, 39, 40, 43, 45, 46, 53, 55, 56, 65, 69, 75, 82, 95) were used repeatedly. Four conditions were created based on different orders and combinations of symbolic numbers and dot quantities: dots-dots (DD), numerals-numerals (NN), dots-numerals (DN), numerals-dots (ND). All conditions were randomly intermixed during testing. The same four ratios between the two stimuli (larger divided by smaller number: 1.3, 1.6, 1.9, 2.2) were used in all four conditions. Forty pairs of stimuli were generated for each condition, which yielded 160 trials in total. Each trial contained two stimuli presented sequentially

with a fixation point in between, followed by a blank screen. Each stimulus was presented for 600 ms. For half of the trials in each condition, the larger number was presented first, for the other half, the larger number was presented second. Participants were instructed to press “S” on a keyboard if the first image presented had a larger quantity and press “L” if the second one had a larger quantity. There was no time limit for a response, but participants were encouraged to respond as quickly and accurately as possible.

2.2.2. Number-letter discrimination task

The goal of the number-letter discrimination task was to measure participants’ implicit integration between numerical information presented in symbolic and non-symbolic formats. Since previous research has shown that this integration effect can be observed more reliably when participants’ subjective estimates of non-symbolic quantities are taken into account (Liu et al., 2018), we first measured participants’ non-symbolic estimation biases using a dot estimation task and then generated stimulus sets for the number-letter discrimination task that took each participant’s individual bias into account.

2.2.2.1. Dot estimation task. To measure participants’ non-symbolic estimation biases, we used a dot estimation task identical to the task used in Liu et al. (2015), and the dot stimuli were generated using the script created by Dehaene et al. (2005). Participants were asked to estimate the number of dots in each image presented for only 400 ms (so that participants were unable to count the dots). Importantly, stimuli were similar to those in the number-letter discrimination task, i.e., dot arrays were superimposed with double-digit Arabic numerals or letter pairs, from 25 dots quantities (7, 9, 11, 13, 14, 17, 19, 20, 21, 25, 26, 28, 30, 32, 38, 39, 42, 48, 57, 59, 63, 72, 86, 89, 95), 12 Arabic numerals (ranging from 11 to 63), and 12 letter pairs. Dot quantities and their pairings with either Arabic numerals or letters are listed in Table 1. For each dot quantity, six images were generated with three variations in the total area occupied by the dots, and two variations in the dot size, to control for potential effects of perceptual differences in the stimuli. Letter pairs were randomly linked to a specific Arabic numeral such that specific letter pairs were always paired with the same dot quantities (e.g., “RC” was matched with Arabic numeral “11”, so both “RC” and “11” were always paired with dot quantities of 11, 7, 17), though the actual relations between the letters and dot quantities were random designations. Three categories were created for each Arabic numeral and its corresponding letter pair: match with dot quantity, mismatch with dot

quantity where dot quantity is less than Arabic number, and mismatch with dot quantity where dot quantity is greater than Arabic number. Therefore, 432 dot images were created in total – half with Arabic numerals, and half with letters. The estimation task consisted of two sessions of 216 trials. Participants were instructed to estimate the number of dots in each image by typing their answer on the keyboard. Each image was presented for 400 ms, followed by a black screen with no time limit for a response. Participants were encouraged to respond as quickly and accurately as they could. Each participant’s estimates were fitted with a power function  $y_{symbolic\ number} = \alpha^x x_{dot\ quantity}^\beta$ , where  $\alpha$  is the scaling factor and  $\beta$  is the exponent of the power function, and they were obtained using a R-based PsiMLE 1.0 package (Odic, Im, Eisinger, Ly, & Halberda, 2016) with likelihood function  $L(\alpha, \beta, \sigma|x, y) = \prod_{i=1}^n (2\pi\alpha^*x^{\beta*}\sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \alpha^*x^{\beta*})^2 (\alpha^*x^{\beta*}\sigma)^{-2}\right)$ . To ensure that mismatch and match between Arabic numerals and dot arrays did not affect participants’ estimation performance, we conducted a paired samples *t*-test of  $\alpha$  and  $\beta$  obtained from the power function fitting, and confirmed that there was no significant differences between match and mismatch conditions ( $\alpha$ :  $t(125) = -0.29, p = .77$ ;  $\beta$ :  $t(125) = 0.18, p = .85$ ). Thus, the estimation function was calculated across all trials.

2.2.2.2. Number-letter discrimination task. In the number-letter discrimination task, participants were instructed to judge whether each stimulus contained an Arabic numeral or a letter pair. To generate the stimuli for the number-letter discrimination task, 23 dot quantities (7, 9, 11, 13, 14, 17, 19, 20, 21, 25, 26, 28, 32, 38, 39, 42, 48, 57, 59, 63, 72, 89, 90) were selected, covering a similar range as the dot quantities used in the estimation task. Additionally, as in the dot estimation task, dots varied on three different total areas occupied by the dots and two different dot sizes. Based on each participant’s estimation function, the perceived “matching” symbolic number for each dot quantity was generated for each participant. Therefore, the exact Arabic numerals used in this number-letter discrimination task varied across participants based on their estimation biases. For each dot quantity, three categories were generated: match with dot quantity, mismatch with dot quantity where the dot quantity was less than the Arabic numeral, and mismatch with dot quantity where the dot quantity was greater than the Arabic numeral. A ratio of 1.35 was used to generate those two “mismatching” conditions (a 1.35 ratio of symbolic number to dot quantity and of dot quantity to symbolic number). Each Arabic numeral was associated with

Table 1

Combinations of Arabic numerals, letters, and dot quantities used to create match and mismatch conditions in the dot estimation task.

Dot estimation task			
Arabic numeral (Match)	Letter	Mismatch Dot<Num	Mismatch Dot>Num
11	RC	7	17
13	PH	9	20
17	CF	11	26
21	LR	14	32
25	QX	17	38
28	GM	19	42
32	KJ	21	48
38	XR	25	57
42	YG	28	63
48	JD	32	72
59	PN	39	89
63	FW	42	95

The first column indicates the Arabic numerals and dot quantities used to create the match trials. Letter indicates the letter pairs matched with the corresponding numerals. The last two columns list the dot quantities used in the mismatch conditions with dot quantities being either smaller or larger than the Arabic numerals.

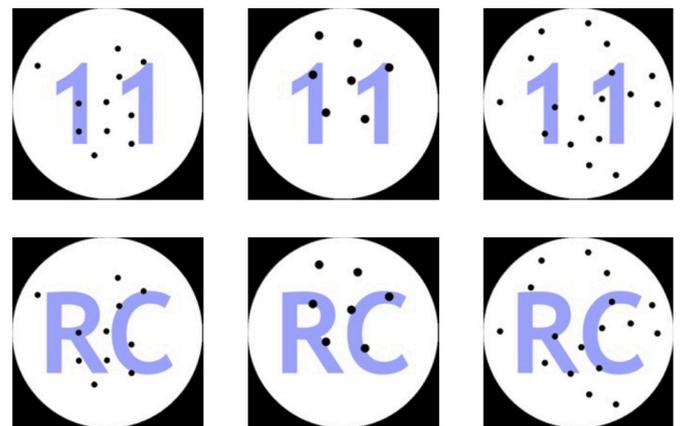


Fig. 1. Sample stimuli for the number-letter discrimination task. From left to right: Examples of Arabic numerals and letter pairs in the “match”, “mismatch (numerals < dots)”, and “mismatch (numerals > dots)” conditions. Note that the exact stimuli for each participant were generated individually before the number-letter discrimination task based on each participant’s estimation biases derived from the dot estimation task.

a randomly chosen letter pair (e.g., “11” with “RC”, see Fig. 1 for illustration). In total, 828 images were created for 23 dot quantities with 414 trials in each of two sessions. Each stimulus was presented for 400 ms, followed by a blank screen with no time limit for a response. Participants were instructed to press “S” if a numeral was presented, and “L” if a letter pair was presented. Participants were encouraged to answer as quickly and accurately as possible.

### 2.2.3. Math abilities

Participants’ math abilities were tested using the nationally normed Woodcock Johnson III Tests of Achievement, which contain three math subtests: Calculation, Math Fluency, and Applied Problems (Woodcock et al., 2001). The Calculation subtest uses a written format and includes problems that involve arithmetic, trigonometry, geometry, logarithms, and calculus. The Math Fluency subtest measures participants’ performance in timed mental arithmetic from solving as many simple addition, subtraction, and multiplication problems as possible within 3 min. During the Applied Problems subtest, the experimenter verbally reads word problems to the participants, and participants are required to select the relevant information, recognize the procedure, and perform the necessary calculations to arrive at the answer. Participants are asked to give answers verbally, but they are allowed to use scratch paper when needed.

## 2.3. Data analysis

### 2.3.1. Number comparison task

For the number comparison task, two behavioral measures were obtained: response time (RT hereafter) and accuracy. For each subject, trials with RT faster than 200 ms or 3 *SD* above the mean were excluded from the analyses (2.1% of trials on average, range: 0.6% - 4.4%). Additionally, two subjects were excluded due to low accuracy (i.e., below 50%), which resulted in useable number comparison data from 113 participants. To investigate how RT and accuracies differed in the four conditions, mean RT and mean accuracies were submitted to two separate repeated-measures analyses of variance (ANOVAs) with Condition (four levels: DD, NN, DN, ND) as the within-subject variable. In the case of significant main effect of Condition, pairwise comparisons between the four conditions were conducted to gain further insight into how conditions differed from one another.

### 2.3.2. Number-letter discrimination task

For the number-letter discrimination task, two behavioral measures were obtained: RT and accuracy. Before the analysis, for each subject, trials with RT faster than 200 ms or 3 *SD* above the mean were excluded (1.6% of trials on average, range: 0.5% - 4.0%). All participants’ mean accuracy was above chance (range: 70.5% - 99.8%), so data from all 115 participants remained in this analysis. To investigate how Symbol type (numeral, letter pair) and Match type (match, mismatch) influenced participants’ RT and accuracies, we ran two separate repeated-measures ANOVAs for RT and accuracy respectively.

### 2.3.3. Correlational analysis between number comparison task and number-letter discrimination task

To explore whether numerical integration across symbolic and non-symbolic number stimuli in the number comparison and number-letter discrimination tasks were linked, we calculated Pearson correlations between number comparison and number-letter discrimination task performance. Since accuracy did not differ between match and mismatch trials in the numeral condition for the number-letter discrimination task (see Results Section 3.2), these correlational analyses were only conducted on RT. Specifically, for the number comparison task, we calculated the mean RT difference between each of the two mixed-format conditions and the same-format conditions (i.e., DN-DD, DN-NN, ND-DD, and ND-NN). As such, a greater positive difference indicates numerical estrangement. For the number-letter discrimination

task, we calculated the mean RT difference between mismatch and match trials for the numeral condition (i.e., mismatch-match). In this case, a greater positive difference indicates numerical integration. For each correlation analysis, participants whose data were beyond three standard deviations from the mean of each numerical integration index were excluded from the 113 subjects before running the analysis (see Results Section 3.3 for detailed information).

### 2.3.4. Correlation analysis between number comparison and number-letter discrimination task performance and math abilities

To investigate the role of numerical estrangement/integration for adults’ math abilities, we conducted Pearson correlations between number comparison and number-letter discrimination task performance with math abilities. Specifically, we used the same indices of numerical estrangement/integration derived from the number comparison and number-letter discrimination tasks (described in Section 2.3.3), and math abilities were measured using standardized scores derived from the three Woodcock Johnson subtests. For the correlational analyses, two participants’ data were removed from the 113 subjects for the number comparison task, and two were excluded from the 115 subjects for the number-letter discrimination task due to missing math scores. Participants whose data were beyond three standard deviations from the mean of each measurement were also excluded before running the analyses (see Results Section 3.4 for detailed information).

## 3. Results

### 3.1. Number comparison task

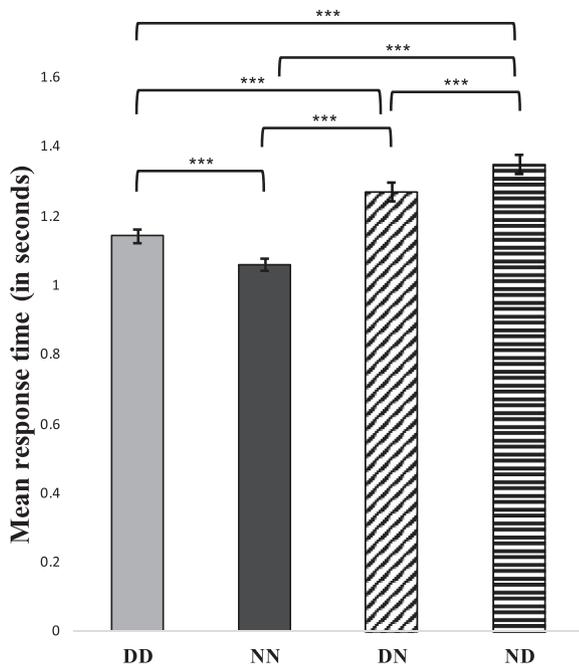
For the number comparison task, we were interested in how different combinations of number formats (dots and numerals) affect participants’ RT and accuracy. To this end, RT and accuracy were submitted to two separate repeated-measures ANOVAs with condition (four levels: DD, NN, DN and ND) as the within-subject variable. For RT, there was a significant main effect of condition,  $F(3, 336) = 125.5, p < .001, \eta_p^2 = 0.53$ . Participants responded faster for same-format conditions (DD, NN) than mixed-format conditions (DN, ND; Fig. 2). Pairwise comparisons were conducted to further investigate the differences in RT across the four conditions. All pairwise comparisons showed significant differences after Bonferroni correction,  $ps < 0.001$  (Fig. 2).

For accuracy, we also found a significant main effect of condition,  $F(3, 336) = 169.7, p < .001, \eta_p^2 = 0.60$ . Participants performed more accurately in single-format conditions (DD, NN) than in mixed-format conditions (DN, ND; Fig. 3). All pairwise comparisons showed significant differences between the four conditions after Bonferroni correction,  $ps < 0.01$  (see Fig. 3).

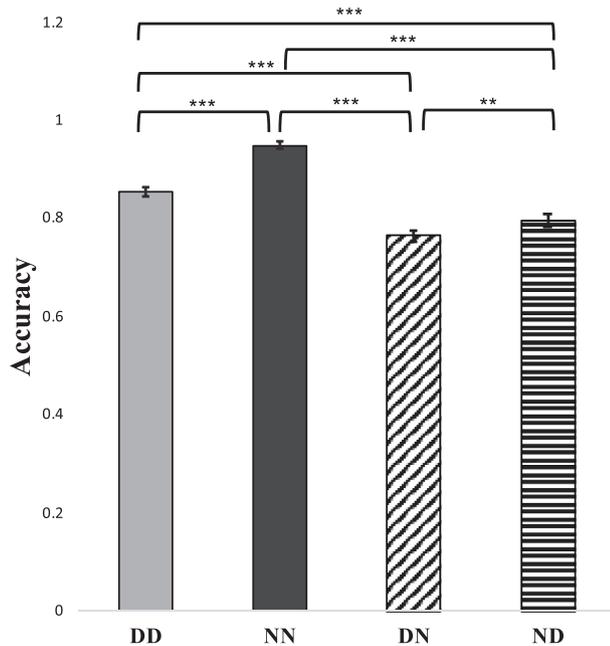
### 3.2. Number-letter discrimination task

For the number-letter discrimination task, we ran two separate 2 (text type: numeral, letter pair) X 2 (match type: match, mismatch) repeated-measures ANOVAs for mean RT and accuracy. For RT, there was a significant main effect of text type,  $F(1, 114) = 64.18, p < .001, \eta_p^2 = 0.59$  with participants performing faster during numeral trials ( $M = 0.66, SD = 0.14$ ) than letter pair trials ( $M = 0.68, SD = 0.14$ ). A significant main effect of match type was also observed,  $F(1, 114) = 7.46, p = .007, \eta_p^2 = 0.06$ . Participants responded faster during match trials ( $M = 0.66, SD = 0.14$ ) than mismatch trials ( $M = 0.67, SD = 0.14$ ). Importantly, there was also a significant interaction between text type and match type,  $F(1, 114) = 7.99, p = .006, \eta_p^2 = 0.07$ . Specifically, a difference in RT between match and mismatch trials was observed in the numeral trials ( $F(1, 114) = 11.48, p < .001, \eta_p^2 = 0.09$ ), and there was no difference in RT in the letter trials ( $F(1, 114) = 0.001, p = .98, \eta_p^2 = 0$ ; Fig. 4).

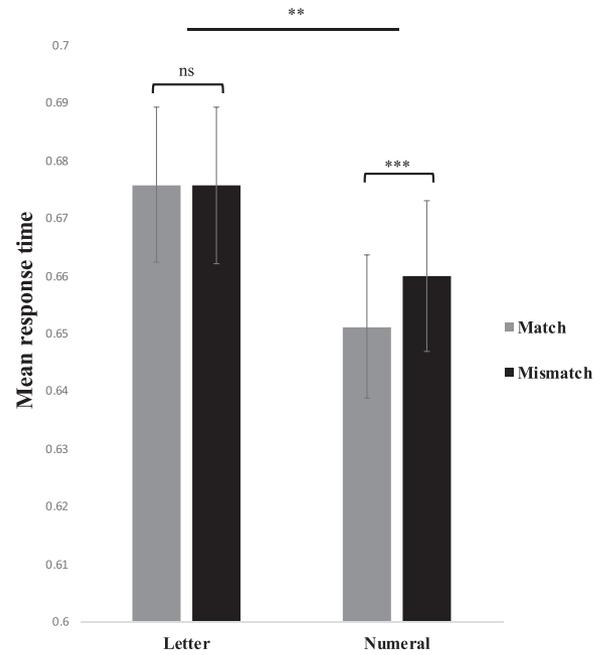
For accuracy, participants performed more accurately during mismatch trials ( $M = 0.953, SD = 0.051$ ) than match trials ( $M = 0.949$ ,



**Fig. 2.** Mean RT for the four conditions in the number comparison task. DD = dots-dots comparison, NN = numerals-numerals comparison, DN = dots-numerals comparison, ND = numerals-dots comparison. Pairwise comparisons between the four conditions were Bonferroni corrected. Error bars denote standard errors. Asterisks reflect a significant difference between two conditions, \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ .



**Fig. 3.** Mean accuracy for the four conditions in the number comparison task. DD = dots-dots comparison, NN = numerals-numerals comparison, DN = dots-numerals comparison, ND = numerals-dots comparison. Pairwise comparisons between the four conditions were Bonferroni corrected. Error bars denote standard errors. Asterisks reflect a significant difference between two conditions, \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ .



**Fig. 4.** Mean RT (in seconds) on the number-letter discrimination task separated by text type (numeral, letter pair) and match type (match, mismatch). Asterisks reflect a significant difference between two conditions, \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ . Error bars denote standard errors, and ns indicates non-significant difference.

$SD = 0.055$ ),  $F(1, 114) = 5.29, p = .02, \eta_p^2 = 0.04$ . However, no main effect of text type was observed,  $F(1, 114) = 0.22, p = .64, \eta_p^2 = 0.004$ , and there was no significant interaction between text type and match type,  $F(1, 114) = 3.66, p = .06, \eta_p^2 = 0.03$  (Fig. 5).<sup>1</sup>

### 3.3. Correlation between number comparison task and number-letter discrimination task

To investigate the numerical integration across symbolic and non-symbolic number stimuli in the number comparison and number-letter discrimination tasks, we first calculated indices of numerical integration for each task. Since only RT showed the expected significant difference between match and mismatch trials in the numeral trials for the number-letter discrimination task, we only focused on RT measurements for these analyses. For the number comparison task, four different indices were calculated: the RT differences between each mixed-format condition and each same-format condition (i.e., DN-DD, DN-NN, ND-DD,

<sup>1</sup> We also ran additional Pearson correlations to check whether there was a positive correlation between RTs and accuracy for numeral-match and numeral-mismatch conditions. We did find positive correlations between RTs and accuracy for both conditions ( $ps < 0.001$ ), which suggested a speed-accuracy trade-off. Thus, we calculated efficiency scores (ES; Townsend & Ashby, 2014) which integrate both RT and accuracy ( $ES = \text{average (correct) RT} / (1 - \text{error rate})$ ) and re-ran the repeated-measures ANOVA with efficiency scores. There was a significant main effect of text type,  $F(1, 114) = 38.59, p < .001, \eta_p^2 = 0.25$ . There was also a significant interaction between text type and match type,  $F(1, 114) = 14.36, p < .001, \eta_p^2 = 0.11$ . Significant difference between match and mismatch in numeral trials was observed as well,  $F(1, 114) = 8.45, p = .004$ . Therefore, our results still hold despite the speed-accuracy trade-off. Also, note that measurement of accuracy carries relatively small variations across different conditions, which may not capture the individual differences in the task performance well. Thus, results for accuracy in the number-letter discrimination task should be interpreted with caution. For ease of interpretation and consistency with other analyses, we continue to use RT throughout the paper.

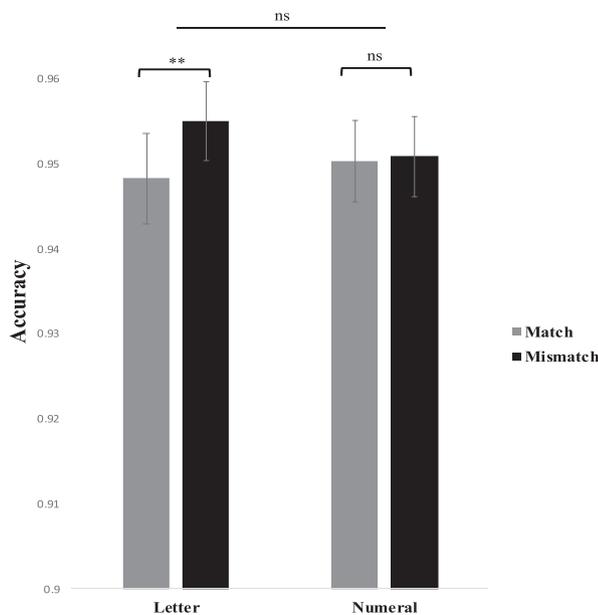


Fig. 5. Mean accuracy on the number-letter discrimination task separated by text type (numeral, letter pair) and match type (match, mismatch) on accuracy. Error bars denote standard errors.

and ND-NN). Here, a greater positive difference indicates greater numerical estrangement. For the number-letter discrimination task, only one index of numerical integration was calculated: the RT difference between mismatch and match trials in the numeral trials (i.e., mismatch-match). Here, a greater positive difference indicates greater numerical integration. Participants' performance differences (RT) between mixed-format and same-format conditions in the comparison task, and between the numeral mismatch and match trials in the number-letter discrimination task are summarized in Table 2.

Pearson correlations were conducted between the four indices of numerical estrangement in the number comparison task and the one index of numerical integration in the number-letter discrimination task performance. No significant correlations were observed between number comparison and number-letter discrimination task performances,  $p > 0.20$  (Table 3).

### 3.4. Correlation between number comparison and number-letter discrimination task performance and math abilities

To investigate the relation between numerical estrangement as measured in the number comparison task and math abilities, we used the same indices of numerical estrangement derived from the number comparison task described in Section 2.3.3 (i.e., DN-DD, DN-NN, ND-DD, and ND-NN). The mean math score across the three Woodcock Johnson subtests was calculated to represent overall math abilities for each participant. Two participants' data were excluded due to missing

Table 2  
Participants' performance (RT) differences between different trial conditions in the number comparison task and number-letter discrimination task.

	t-statistic	p-value	Cohen's d
DN vs DD	8.87	< 0.001	0.46
DN vs NN	12.12	< 0.001	0.76
ND vs DD	10.87	< 0.001	0.74
ND vs NN	14.86	< 0.001	1.02
Mismatch vs Match	3.33	= 0.001	0.06

Mismatch vs Match indicates the RT differences between numeral mismatch and match conditions in the numeral trials of the number-letter discrimination task.

Table 3  
Pearson correlations between indices derived from the number-letter discrimination task (mean RT difference between mismatch and match in numeral trials) and comparison task (mean RT difference between each of the two mixed-format conditions (i.e., DN and ND) and the same-format conditions (i.e., DD and NN)).

	DN-DD	DN-NN	ND-DD	ND-NN
Mismatch-match	$r(107) = -0.03$ ( $p = .76$ )	$r(109) = -0.12$ ( $p = .20$ )	$r(108) = -0.07$ ( $p = .50$ )	$r(109) = -0.11$ ( $p = .26$ )

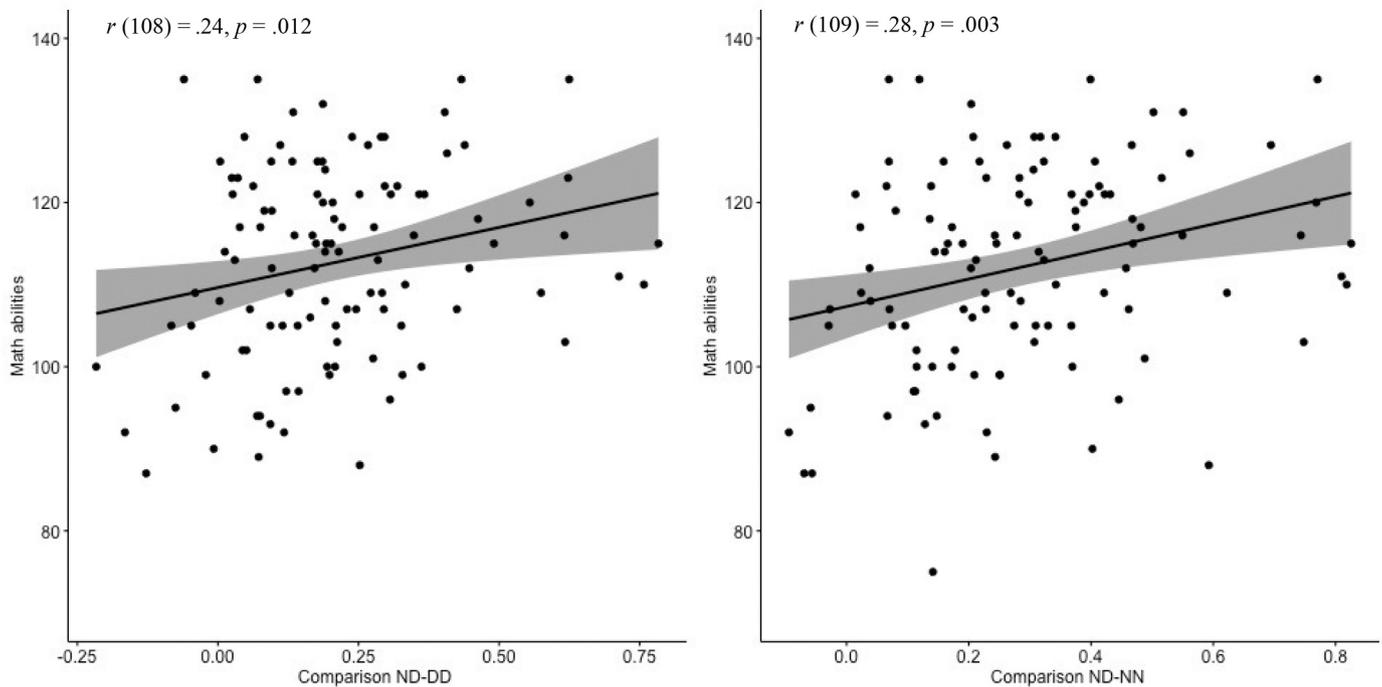
math scores, and participants whose data were beyond three standard deviations from the mean of each measurement were also excluded before running the correlations. No significant Pearson correlations were observed between DN-DD ( $M = 0.13$ ,  $SD = 0.15$ ) and math abilities ( $r(107) = 0.04$ ,  $p = .72$ ), or between DN-NN ( $M = 0.21$ ,  $SD = 0.18$ ) and math abilities ( $r(109) = 0.08$ ,  $p = .38$ ). However, there were significant correlations between ND-DD ( $M = 0.20$ ,  $SD = 0.20$ ) and math abilities ( $r(108) = 0.24$ ,  $p = .012$ ), and between ND-NN ( $M = 0.29$ ,  $SD = 0.21$ ) and math abilities ( $r(109) = 0.28$ ,  $p = .003$ ; see Fig. 6). In other words, participants with stronger math abilities had greater differences in completing the mixed numeral-first trials vs. the same-format trials. To ensure the robustness of our results, we re-ran the correlational analyses (between ND-DD, ND-NN and math abilities) by controlling for participants' performance (RT) on DD and NN trials separately, and the significant correlations remained even when accounting for response times on same-format trials ( $p < 0.01$ ). Correlations between the same indices derived from the number comparison task and each of the three Woodcock Johnson subtests can be found in Table A1 in the Appendix.

We also investigated the correlation between number-letter discrimination task performance and math abilities. Specifically, we used the same index of numerical integration (i.e., mismatch - match) derived from the number-letter discrimination task described in Section 2.3.3, and the mean math score of the three Woodcock Johnson subtests used in the analyses above. No significant correlation was observed between the numerical integration index derived from the number-letter discrimination task and adults' math abilities ( $r(109) = -0.02$ ,  $p = .87$ ).

## 4. Discussion

In the current study, we had two aims: to test whether numerical integration or estrangement is task-dependent and associated with adults' math abilities. To answer those questions, we administered two tasks to measure numerical integration/estrangement, and a standardized math assessment to the same group of adult participants. In the number comparison task with non-symbolic dot arrays and Arabic numerals, participants indicated which of the sequentially presented stimuli was numerically larger. Half of the trials consisted of the same format (dots vs. dots or numerals vs. numerals); the other half consisted of mixed formats (dots vs. numerals or numerals vs. dots). We found that participants were significantly slower when comparing mixed-format stimuli compared to same-format stimuli regardless of whether these were dots or numerals, suggesting that different formats are not immediately integrated. More interestingly, within the mixed-format stimuli, participants were significantly slower when Arabic numerals were presented first and dot arrays second compared to the opposite order, suggesting that the order in which different formats are presented has an additional impact on the estrangement between different numerical stimuli.

Making explicit magnitude-based judgements in the number comparison task might force the translation of symbolic number representations into non-symbolic ones or vice versa. To overcome this issue, we also administered a number-letter discrimination task to explore whether numerical integration is evident without explicit magnitude processing. Specifically, in the number-letter discrimination task, participants judged whether Arabic numerals or letter pairs were presented



**Fig. 6.** Left panel: Scatterplot depicting the relation between mean RT difference for ND and DD conditions in the number comparison task and math abilities. Right panel: Scatterplot depicting the relation between mean RT difference for ND and NN conditions in the number comparison task and math abilities. Shaded area indicates 95% confidence interval.

on dot arrays, where the number of dots sometimes numerically matched the overlain numeral. Participants were significantly faster in number-letter discriminations when the dot arrays matched the Arabic numeral, suggesting that task-irrelevant numerical information embedded in the dot arrays affected a simple judgement about whether the symbolic stimulus was a numeral or letter string. Finally, we found that measures of numerical estrangement derived from the number comparison task significantly correlated with adults' overall math abilities, specifically in mixed-format trials where Arabic numerals were presented first.

#### 4.1. Evidence against numerical integration using a number comparison task

In the number comparison task, we contrasted participants' performance on the same-format and mixed-format trials. Consistent with previous findings (Lyons et al., 2012), participants performed significantly slower and less accurately when comparing mixed-format stimuli than single-format stimuli. Lyons et al. (2012) reasoned that an additional processing cost is accrued when participants compare mixed-format stimuli than when they compare two same-format stimuli suggesting that the two number formats (i.e., dot arrays and Arabic numerals) are not immediately integrated, and thus demonstrate "symbolic estrangement". More importantly, in contrast to Lyons et al. (2012), we found that participants were significantly slower when Arabic numerals were presented first and dot arrays second compared to the opposite order in the mixed trials. One potential explanation of these inconsistent findings is that they used a blocked design in which participants were aware of the format of the upcoming stimulus set, while our trial types were randomly intermixed, and participants did not know whether a trial was a same-format or mixed-format trial ahead of time. Our findings instead suggest that the order in which different formats are presented has an additional impact on the estrangement between different numerical stimuli, if participants do not know ahead of time what the stimulus format will be.

The order effect we observed in the comparison task suggests that

there may be different cognitive mechanisms involved when making comparisons between symbolic and/or non-symbolic stimuli. We propose two possible explanations. The first possible explanation is based on the assumption that the representational system in which participants make comparisons is determined by the format of the first stimulus. Specifically, if the first stimulus is an Arabic numeral, participants might prepare to use a symbolic representational system, in which the magnitude representation of the numerals could be based on symbolic associations such as the relative position of the numerals between each other (similar to symbol-symbol association account; Reynvoet & Sasanguie, 2016). If the Arabic numeral is followed by another Arabic numeral, processing of the second Arabic numeral will be highly efficient as the first numeral already prompted participants to use the symbolic representational system. However, if the first Arabic numeral is followed by a dot array, the dot array might be verbally re-coded and transferred into the symbolic representational system to make the comparison to the Arabic numeral. This transfer may cause greater cognitive load leading to the observed increase in reaction time for ND trials compared to NN trials. In the case that the first stimulus is a dot array, participants might prepare to use an analog representational system to perform the comparison. Thus, if the second stimulus is another dot array, it will take less effort for participants to make a comparison as they are already prepared to use the analog representational system. In contrast, if the second stimulus is an Arabic numeral, it needs to be translated into an analog representation to make a comparison with the dot array. Interestingly, trials in which this process is necessary are faster than ND trials suggesting that translating a dot array into a symbolic representation is the most effortful.

In contrast, the second explanation is based on the assumption that symbolic numbers need to be represented in an analog representational system to be compared with a non-symbolic stimulus. In other words, whenever a dot array is involved in the comparison – regardless of its position in the presentation – an Arabic numeral involved in the comparison needs to be translated into an analog representation. Comparing two Arabic numerals would be the only condition that allowed participants to compare the numbers directly without accessing the magnitude

information of the stimuli (Sasanguie, De Smedt, & Reynvoet, 2017; Van Hoogmoed & Kroesbergen, 2018). In contrast, if the Arabic numeral is followed by a dot array, participants may retroactively retrieve the Arabic numeral and activate its associated numerical magnitude in the analog representational system to make the comparison. Our results of the estimation task also provided evidence for this possibility such that match or mismatch between Arabic numerals and dot arrays did not affect participants' dot estimation, possibly due to the fact that Arabic numerals are only translated into analog representations when needed for comparison purposes.

This interpretation is also in line with earlier electrophysiological findings. Van Hoogmoed and Kroesbergen (2018) asked adults to indicate whether a prime number and subsequent target number matched while they recorded participants' event-related potentials (ERPs). Primes and targets could either be presented as non-symbolic dot arrays or Arabic numerals. They observed ERP differences in processing non-symbolic primes that predictably preceded non-symbolic targets compared to when they predictably preceded symbolic targets. In addition, ERP differences also emerged when a non-symbolic target was preceded by a non-symbolic prime compared to being preceded by a symbolic prime. These results suggest that processing of symbolic stimuli do not immediately activate their numerical magnitudes, and non-symbolic stimuli are affected by the context in which it occurs. It is important to note though that unlike in our study where all trial types were randomly intermixed, van Hoogmoed and Kroesbergen also used a blocked design where participants were aware of the stimulus types and the order in which they would encounter them. Future studies should examine to what extent knowing what stimuli will be presented affects participants' strategies and may lead to different cognitive processes.

While both explanations aforementioned seem reasonable, the current study could not differentiate between them, and further investigation is needed to disambiguate them. In sum, our findings - that the presentation-order of different formats impacts the estrangement between different numerical stimuli - shed light on and solicit a broader discussion on understanding the relation between symbolic and non-symbolic number processing and the strategies used to perform numerical comparisons.

It is also worth noting that switching between different visual formats (i.e., from Arabic numeral to dot array or vice versa) in our comparison task might contribute to the increased processing time as well. One way that to circumvent the problem is to use an audio-visual paradigm (Marinova et al., 2018; Sasanguie, De Smedt, & Reynvoet, 2017). For example, Marinova et al. (2018) used an audio-visual paradigm (Experiment 3), in which participants were required to compare pairs of symbolic numbers (i.e., Arabic numerals and spoken number words) and/or non-symbolic quantities (i.e., dot arrays and tone sequences). Therefore, four audio-visual comparison tasks were created: (1) a number word-digit task, (2) a tones-dots task, (3) a tones-digit task, and (4) a number word-dots task. Consistent with our findings, they also observed switch costs for trials in which symbolic and non-symbolic number formats were mixed supporting the notion of two distinct representation systems for non-symbolic and symbolic numbers. The advantage of this paradigm is that performance differences observed between the pure and mixed conditions are less likely due to the notation switch that participants have to make in mixed trials, because notation switches are present in all types of trials. This paradigm thus has been suggested to be better suited for investigating numerical integration/estrangement. Future work using this audio-visual paradigm is needed to investigate the influence of the presentation order of stimuli with different formats on the integration/estrangement between symbolic and non-symbolic representations.

#### 4.2. Implicit numerical integration

Comparison tasks have been widely used to examine the relation between symbolic and non-symbolic number representations. However,

explicitly asking participants to process two formats of number for the purposes of making comparisons may potentially prime people to attend primarily to either symbolic or non-symbolic numerical information (Liu et al., 2015). Researchers previously have also argued that distance and ratio effects observed in comparison tasks may not necessarily indicate the overlapping representations of numerical magnitudes, instead, they might reflect a response-selection process that occurs at the level of comparing two magnitudes (Van Opstal, Gevers, De Moor, & Verguts, 2008). According to this view, the numerical distance effect observed in comparison tasks is due to monotonic connection weights such that the activation of the correct output node decreases (i.e., increased RT) as the numbers become closer (Van Opstal et al., 2008). Similarly, Cohen Kadosh, Brodsky, Levin, and Henik (2008) observed distance effects in an auditory pitch comparison task, and reasoned that it was caused by a sensorimotor transformation such that it does not necessarily reflect magnitude representations. Therefore, numerical distance or ratio effects observed in comparison tasks may index the comparison process and not necessarily the numerical representations per se.

To overcome the limitations of the number comparison task, we implemented a number-letter discrimination task that does not require explicit judgement about magnitude information and is thus thought to tap into implicit numerical integration. The goal of the number-letter discrimination task is to investigate whether symbolic integration can be observed when symbolic and non-symbolic number information are presented simultaneously, but the non-symbolic information is irrelevant to the task that participants are asked to complete. We hypothesized that if participants immediately integrate symbolic and non-symbolic numerical information, participants would perform better when symbolic and non-symbolic number stimuli match than when they mismatch. Replicating previous studies (A. Liu et al., 2015; Liu et al., 2018), we indeed found that participants responded significantly faster in trials when the dot array quantities matched the Arabic numerals. One possible explanation is that participants could process task-irrelevant information simultaneously and activate the numerical information embedded in both dot arrays and Arabic numerals, and this activation could facilitate (in match conditions) or interfere (in mismatch conditions) with the decision making (i.e., whether or not the stimulus is an Arabic numeral). In other words, with the support of congruent numerical magnitude, participants respond faster in this number-letter discrimination task, whereas participants respond slower with incongruent numerical information embedded in dots and Arabic numerals. Moreover, we likely would not observe such an integration effect if the number-letter discrimination task focused on the dot arrays and the Arabic numerals were task-irrelevant. This is because the magnitude of Arabic numerals is not immediately accessed. In sum, our findings suggest that non-symbolic numerical information influences adults' judgement about symbolic numerical information in an implicit way, and that non-symbolic and symbolic number representations can be readily integrated in the absence of a task that requires processing the magnitude of the numerical stimuli.

Our findings from the number-letter discrimination task fit with results from previous neuroimaging studies. For example, using transcranial magnetic stimulation (TMS), Sasanguie, Göbel, and Reynvoet (2013) found that stimulation to the left intraparietal sulcus (IPS) interfered with the mapping between non-symbolic and symbolic representations in a priming task. Thus, it is possible that left IPS serves as the neural foundation for the integration between non-symbolic and symbolic number representations that we observed in the number-letter discrimination task. In other words, we would expect to observe no integration effect in the number-letter discrimination task when stimulating the IPS using TMS. Previous studies also suggested that the same brain regions are engaged when processing symbolic and non-symbolic numerical stimuli passively (Piazza, Pinel, Le Bihan, & Dehaene, 2007). Specifically, in an fMRI adaption task, Piazza et al. (2007) found that once subjects were adapted either to dot arrays or Arabic numerals, IPS and prefrontal regions showed a rebound response when the number

was changed by a sufficient amount, regardless of the notation changes, which demonstrated a convergence of symbolic and non-symbolic representations of numbers in IPS and prefrontal cortex. Therefore, the aforementioned common substrates activated during both symbolic and non-symbolic number processing might serve as the underlying mechanism that supports the numerical integration observed in the number-letter discrimination task. However, previous studies also have suggested different neural representations for symbolic and non-symbolic number processing (Bulthé, De Smedt, & Op de Beeck, 2014, 2015; Sokolowski, Fias, Bosah Ononye, & Ansari, 2017). For example, Bulthé et al. (2014, 2015) used multi-voxel pattern analysis (MVPA) to unravel the neural representations of symbolic and non-symbolic numbers, finding no overlapping representations. However, their task required participants to explicitly assess numerical magnitude, which differs from our experiment design which did not require explicit judgement about magnitude. Therefore, the differences in the methodology may have led to the discrepancy between our findings with theirs. One future research direction would be to use tasks that implicitly assess numerical integration (e.g., number-letter discrimination) combined with neuro-imaging techniques to investigate the relation between brain activation in certain brain regions (e.g., IPS) and participants' performance on such tasks.

One caveat of our findings is that even though we found significant differences between match and mismatch conditions in the numeral trials, the effect was small, which may not be well-suited for capturing robust individual differences in performance (RT). Thus, true correlations between number-letter discrimination performance and other measures may be masked by low variability.

#### 4.3. Discrepancy between explicit and implicit numerical integration tasks

Why does the number comparison task seem to suggest symbolic estrangement whereas the number-letter discrimination task suggests symbolic integration? These inconsistencies might be driven by the different task demands and experimental designs. Asking participants to make an explicit number comparison likely activates different cognitive processes than determining whether Arabic numerals or letter pairs are presented, and different strategies will be selected according to the task demands as well, which may also influence the number representation and processing. In addition, in the comparison task, participants were presented with symbolic and non-symbolic stimuli sequentially. Sequential stimulus presentation requires participants to keep the information in mind, thus relying heavily on short-term memory. In the comparison task, the use of short-term memory might take up cognitive resources to maintain the information such that not all numerical stimuli are immediately associated with their approximate magnitude, and this association may only be performed on symbolic number stimuli when explicitly needed (e.g., acquire its numerical magnitude via an analog representation to make comparison with another non-symbolic stimulus), as in the case of mixed-format trials. This explanation might also hold for non-symbolic stimuli that they are not associated with a symbolic representation unless needed. However, the current study could not differentiate these two possibilities as per our earlier discussion. In the case of the number-letter discrimination task, when symbolic and non-symbolic numerical information are presented simultaneously in the context of a simple number-letter discrimination, there is limited reliance on short-term memory such that more cognitive resources are available for accessing the numerical magnitude, which might have caused the discrepancy in the findings between these two tasks. Therefore, in order to fully understand *when* and *how* symbolic and non-symbolic numerical information are integrated, it is important for future studies to explore a broader range of experimental paradigms and different stimulus modalities.

#### 4.4. Numerical integration and math abilities in adults

The second research question we investigated was whether any of our indices of numerical estrangement/integration correlated with individuals' math skills. We found that some indices of the numerical estrangement derived from the number comparison task were correlated with participants' performance on the standardized math achievement tests. Specifically, the performance cost (longer RT) in the numeral-dot (ND) trials relative to both types of same-format trials (numeral-numeral or dot-dot) positively correlated with adults' overall math abilities. That is, the greater the difference between the time it took participants to complete the ND comparison compared to a same-format comparison, the better these participants were at math. Interestingly, we did not find any significant correlations between the performance cost for DN trials and participants' math achievement, further confirming the asymmetry in the direction of the numerical estrangement.

In an earlier study, Guillaume, Nys, Mussolin, and Content (2013) administered numerical comparison and continuous (cumulative area) comparison tasks with dot arrays to investigate the relation between the ANS acuity and adults' math abilities. They found that adults with greater math ability were more affected by the numerical dimension in the continuous comparison task compared to adults with lower math ability. They thus argued that adults with greater math ability can more quickly access the numerical magnitude from the visual input (i.e., higher ANS acuity), which creates greater interference when judging continuous property. However, our findings suggest the opposite that quick access of the numerical magnitude from the visual input is not of great importance for adults' math abilities. Specifically, when processing the non-symbolic numerical information seems less relevant, people with greater math skills are less likely to activate the magnitude of number symbols in processing the symbolic numerical information. More generally, previous research has shown that symbolic number processing is more consistently and more strongly associated with math abilities in both adults and children. For example, Castronovo and Göbel (2012) showed that adults' performance in a symbolic number comparison task was significantly associated with higher math achievement. Similarly, Lyons and Beilock (2011) suggested that symbolic number-ordering ability in adults significantly predicted participants' complex mental arithmetic performance. Evidence was also found in developmental studies such that children's performance on symbolic comparison tasks significantly correlates with their math achievement (De Smedt et al., 2009; Holloway & Ansari, 2009; Mundy & Gilmore, 2009). For instance, Holloway and Ansari (2009) found that children showed individual differences in the numerical distance effect, which related to their math achievement. Specifically, children who showed larger distance effects tended to have relatively lower math test scores. Our results added another important piece of evidence that symbolically estranged number processing is related to adults' math abilities. It is worth noting that our measure of math skills is heavily weighted towards the ability to represent, retrieve, and calculate exact quantities, so other standardized measures of math competency are needed to broadly investigate the associations between non-symbolic and symbolic number processing and math abilities.

## 5. Conclusions

The goals of this current study were twofold: 1) to determine whether evidence for numerical integration/ estrangement is task dependent, and 2) to explore the relation between numerical integration/estrangement and adults' math abilities using different measures. To answer those questions, we administered both a number comparison task similar to the one used by Lyons et al. (2012) and a number-letter discrimination task adapted from Liu et al. (2015) to the same group of adult participants. In the number comparison task, participants were significantly slower when comparing mixed-format stimuli than when comparing same-format stimuli, which agrees with the notion of

symbolic estrangement, i.e., additional processing is needed for mixed-format conditions. More importantly, we found that the presentation order of the mixed-format trials matters for the size of the cognitive cost, suggesting that participants generally do not activate the associated magnitude of Arabic numerals but that this translation is required when comparing an Arabic numeral to a dot array. Our findings regarding the association between symbolic estrangement and math abilities suggest that adults who show greater levels of symbolic estrangement, especially when Arabic numerals are presented first, tend to have higher math skills. Symbolic integration as indexed on a number-letter discrimination task does not seem to relate to adults' math abilities. Thus, we conclude that numerical integration or estrangement is task-dependent, and greater symbolic estrangement is related to greater math abilities in

adults.

## Declarations of interest

None.

## Acknowledgement and funding

This work was supported by the National Science Foundation (grant number: 1734735). We would like to thank Taylor Casteel for collecting the data, and Corrine Durisko and Griffin Koch for their helpful discussion.

## Appendix

**Table A1**

Correlations between number comparison task indices and Woodcock Johnson subtests.

	Calculation	Math Fluency	Applied Problems
DN – DD	$r(106) = 0.05$ $p = .637$	$r(106) = -0.02$ $p = .813$	$r(109) = 0.05$ $p = .620$
DN – NN	$r(108) = 0.11$ $p = .244$	$r(108) = -0.02$ $p = .807$	$r(109) = 0.07$ $p = .489$
ND – DD	$r(107) = 0.22$ $p = .024$	$r(107) = 0.13$ $p = .187$	$r(108) = 0.26$ $p = .007$
ND – NN	$r(108) = 0.30$ $p = .002$	$r(108) = 0.15$ $p = .118$	$r(109) = 0.25$ $p = .008$

## References

- Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport*, *16*(16), 1769–1773.
- Braham, E. J., & Libertus, M. E. (2018). When approximate number acuity predicts math performance: The moderating role of math anxiety. *PLoS One*, *13*(5), Article e0195696.
- Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–322. <https://doi.org/10.1016/j.neuroimage.2013.10.049>
- Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: Format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1376–1387. [https://doi.org/10.1162/jocn\\_a\\_00787](https://doi.org/10.1162/jocn_a_00787)
- Cantlon, J. F., Libertus, M. E., Pined, P., Dehaene, S., Brannon, E. M., & Pelphrey, K. A. (2009). The neural development of an abstract concept of number. *Journal of Cognitive Neuroscience*, *21*(11), 2217–2229.
- Carey, S. (2011). Précis of *The Origin of Concepts*. *Behavioral and Brain Sciences*, *34*(3), 113–124. <https://doi.org/10.1017/S0140525X10000919>
- Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS One*, *7*(4), Article e33832.
- 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.
- Defever, E., Sasanguie, D., Gebuis, T., & Reynvoet, B. (2011). Children's representation of symbolic and nonsymbolic magnitude examined with the priming paradigm. *Journal of Experimental Child Psychology*, *109*(2), 174–186.
- Dehaene, S. (2001). Précis of the number sense. *Mind & Language*, *16*(1), 16–36.
- Dehaene, S., & Akhavan, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 314.
- Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*.
- Dietrich, J. F., Huber, S., & Nuerk, H.-C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS) – A research review. *Frontiers in Psychology*, *6*, 295.
- Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., & Kleinschmidt, A. (2003). A Supramodal number representation in human intraparietal cortex. *Neuron*, *37*(4), 719–726. [https://doi.org/10.1016/S0896-6273\(03\)00036-9](https://doi.org/10.1016/S0896-6273(03)00036-9)
- Guillaume, M., Nys, J., Mussolin, C., & Content, A. (2013). Differences in the acuity of the approximate number system in adults: The effect of mathematical ability. *Acta Psychologica*, *144*(3), 506–512. <https://doi.org/10.1016/j.actpsy.2013.09.001>
- 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.
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, *18*(6), 1222–1229.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385.
- Kadosh, R. C., Brodsky, W., Levin, M., & Henik, A. (2008). Mental representation: What can pitch tell us about the distance effect? *Cortex*, *44*(4), 470–477.
- Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., & Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Developmental Psychology*, *46*(5), 1309.
- Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–379.
- Liu, A. S., Schunn, C. D., Fiez, J. A., & Libertus, M. E. (2015). *Symbolic integration, not symbolic estrangement*. *CogSci: For Double-Digit Numbers*.
- Liu, R., Schunn, C. D., Fiez, J. A., & Libertus, M. E. (2018). The integration between nonsymbolic and symbolic numbers: Evidence from an EEG study. *Brain and Behavior: A Cognitive Neuroscience Perspective*, *8*(4), Article e00938. <https://doi.org/10.1002/brb3.938>
- Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, *141*(4), 635.
- Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261.
- Marinova, M., Sasanguie, D., & Reynvoet, B. (2018). Symbolic estrangement or symbolic integration of numerals with quantities: Methodological pitfalls and a possible solution. *PLoS One*, *13*(7), Article e0200808.
- Marinova, M., Sasanguie, D., & Reynvoet, B. (2021). Numerals do not need numerosities: Robust evidence for distinct numerical representations for symbolic and non-symbolic numbers. *Psychological Research*, *85*(2), 764–776. <https://doi.org/10.1007/s00426-019-01286-z>
- 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.
- Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502.
- Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (2016). PsiMLE: A maximum-likelihood estimation approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavior Research Methods*, *48*(2), 445–462.

- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*(10), 2013–2019.
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551. <https://doi.org/10.1016/j.tics.2010.09.008>
- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305.
- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*(1), 50–57.
- Reynvoet, B., & Sasanguie, D. (2016). The symbol grounding problem revisited: A thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol–symbol associations. *Frontiers in Psychology*, *7*, 1581.
- Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357.
- Sasanguie, D., De Smedt, B., & Reynvoet, B. (2017). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychological Research*, *81*(1), 231–242.
- Sasanguie, D., Göbel, S. M., & Reynvoet, B. (2013). Left parietal TMS disturbs priming between symbolic and non-symbolic number representations. *Neuropsychologia*, *51*(8), 1528–1533.
- Sasanguie, D., Lyons, I. M., De Smedt, B., & Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition*, *165*, 26–38.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, *20*(3), Article e12372.
- Sokolowski, H. M., Fias, W., Bosah Ononye, C., & Ansari, D. (2017). Are numbers grounded in a general magnitude processing system? A functional neuroimaging meta-analysis. *Neuropsychologia*, *105*, 50–69. <https://doi.org/10.1016/j.neuropsychologia.2017.01.019>
- Townsend, J. T., & Ashby, F. G. (2014). Methods of modeling capacity in simple processing systems. In *Cognitive theory* (pp. 211–252). Psychology Press.
- Van Hoogmoed, A. H., & Kroesbergen, E. H. (2018). On the difference between numerosity processing and number processing. *Frontiers in Psychology*, *9*, 1650.
- Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425.
- Woodcock, R. W., McGrew, K. S., Mather, N., et al. (2001). *Woodcock-Johnson III tests of achievement*.