Acta Psychologica 176 (2017) 47–57

Contents lists available at ScienceDirect

Acta Psychologica

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

Eye-movement patterns during nonsymbolic and symbolic numerical magnitude comparison and their relation to math calculation skills

Gavin R. Price[a,*], Eric D. Wilkey[a], Darren J. Yeo[a,b]

[a] Department of Psychology & Human Development, Peabody College, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203, United States
[b] Division of Psychology, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Avenue, 637332, Singapore

A B S T R A C T

A growing body of research suggests that the processing of nonsymbolic (e.g. sets of dots) and symbolic (e.g. Arabic digits) numerical magnitudes serves as a foundation for the development of math competence. Performance on magnitude comparison tasks is thought to reflect the precision of a shared cognitive representation, as evidence by the presence of a numerical ratio effect for both formats. However, little is known regarding how visuo-perceptual processes are related to the numerical ratio effect, whether they are shared across numerical formats, and whether they relate to math competence independently of performance outcomes. The present study investigates these questions in a sample of typically developing adults. Our results reveal a pattern of associations between eye-movement measures, but not their ratio effects, across formats. This suggests that ratio-specific visuo-perceptual processing during magnitude processing is different across nonsymbolic and symbolic formats. Furthermore, eye movements are related to math performance only during symbolic comparison, supporting a growing body of literature suggesting symbolic number processing is more strongly related to math outcomes than nonsymbolic magnitude processing. Finally, eye-movement patterns, specifically fixation dwell time, continue to be negatively related to math performance after controlling for task performance (i.e. error rate and reaction time) and domain general cognitive abilities (IQ), suggesting that fluent visual processing of Arabic digits plays a unique and important role in linking symbolic number processing to formal math abilities.

1. Introduction

Recent years have witnessed an increase in attention paid to the relations between basic numerical capacities and the development of math skills. Humans possess the ability to process basic numerical magnitude information, allowing them to compare, order, add, and subtract quantities of objects (Feigenson, Dehaene, & Spelke, 2004). This so-called 'approximate number system' (ANS) is observable in infancy (Xu & Spelke, 2000) and is shared with non-human species (Cantlon & Brannon, 2006, 2007). The ANS shows individual differences in precision, which are, typically, indexed by the effect of numerical ratio on nonsymbolic number comparison tasks (i.e. deciding which of two sets of dots is the more numerous). The numerical ratio effect (NRE) refers to the robustly observed effect that as the ratio of the smaller over the larger number increases, error rates and response times for comparing those two numbers increase (Moyer & Landauer, 1967). In other words, the closer two numbers are to one another, the more difficult it is to compare their relative magnitude. This effect is thought to reflect a greater degree of representational overlap for quantities that are closer together on a logarithmically compressed mental number line (Dehaene, 2003).

The NRE is also observed (and in fact was originally observed) when individuals compare the relative numerical magnitude of Arabic digits (Moyer & Landauer, 1967). This overlap has led some researchers to believe that Arabic digits acquire their semantic referents by being mapped onto the ANS (Mundy & Gilmore, 2009) which in turn may lead to refining of the ANS itself (Mussolin, Nys, Leybaert, & Content, 2015; Piazza, Pica, Izard, Spelke, & Dehaene, 2013). An alternative perspective suggests that Arabic digit knowledge is acquired independently of the ANS and is instead based on a developmental interplay between linguistic and object-attention systems (Carey, 2001). Whether or not nonsymbolic and symbolic number processing are based on the same underlying neurocognitive mechanisms remains an issue of some debate (Cohen Kadosh, Lammertyn, & Izard, 2008). However, it seems reasonable to presume that behavioral performance on number comparison tasks is not solely driven by the precision of underlying cognitive representations, but instead, comprises a combination of visuo-perceptual processes, response selection mechanisms, and other

* Corresponding author.
E-mail address: gavin.price@vanderbilt.edu (G.R. Price).

http://dx.doi.org/10.1016/j.actpsy.2017.03.012
Received 1 September 2016; Received in revised form 24 March 2017; Accepted 24 March 2017
Available online 02 April 2017
0001-6918/

cognitive processing. Although some recent findings suggest that numerosity processing of multiple modalities and presentation conditions (sequential vs. simultaneous) may depend on a general representation of number (Arrighi, Togoli, & Burr, 2014), and that numerosity information is extracted from visual displays spontaneously (Cicchini, Anobile, & Burr, 2016), at present, the extent to which visual-perceptual processing contributes to number comparison performance for nonsymbolic versus symbolic number formats is unknown.

At the same time, a growing body of research suggests that individual differences in the processing of nonsymbolic (Halberda, Mazzocco, & Feigenson, 2008; Mazzocco, Feigenson, & Halberda, 2011) and symbolic (Bugden & Ansari, 2011; Holloway & Ansari, 2009) numerical magnitude are related to individual differences in the acquisition of math skills. There is some conflict across the extant literature as to whether nonsymbolic or symbolic skills are the stronger predictor of math performance (for a review see De Smedt, Noël, Gilmore, & Ansari, 2013), with recent evidence suggesting that symbolic skills may mediate the relation between nonsymbolic skills and math (Fazio, Bailey, Thompson, & Siegler, 2014; Lyons & Beilock, 2011; Price & Fuchs, 2016). Given the current lack of knowledge regarding the component mechanisms underlying numerical comparison performance, it is unclear what is driving the relation between number comparison and math outcomes at the mechanistic level. Despite this lack of knowledge, educational interventions have already been developed that seek to improve ANS precision as a method to improve math outcomes (Park & Brannon, 2013; Wilson et al., 2006). It is critical that the component mechanisms underlying numerical magnitude processing and their relation to math performance be elucidated if such interventions are to achieve optimal efficacy.

One approach to investigating visuo-perceptual processes during cognitive processing is to record and analyze eye-movement patterns during task performance. Eye-tracking data can provide subjective and sensitive information about attentional allocation (Duchowski, 2007), and has been used to gain insights into the mechanistic processes underlying a number of cognitive domains including reading, visual search, memory, language, and problem solving (Henderson, 2013). Eye-tracking has also been used to reveal a number of characteristics of numerical and arithmetic processing (for a review see Mock, Mock, & Huber, 2016) and examination of multiple eye-tracking measures, such as number of fixations and location of first fixation, can provide information regarding sensitivity to top-down cognitive processes or bottom-up stimulus salience respectively (Mock et al., 2016). The majority of numerical cognition eye-tracking studies have investigated multi-digit number processing, with a focus on understanding the ways in which multi-digit numbers are decomposed (Moeller, Neuburger, Kaufmann, Landerl, & Nuerk, 2009; Huber, Mann, Nuerk, & Moeller, 2014). While a handful of studies have investigated nonsymbolic numerical processing, they have largely focused on estimation and enumeration (Gandini, Lemaire, & Dufau, 2008; Godau, Wirth, Hansen, Haider, & Gaschler, 2014; Sophian & Crosby, 2008). Further, despite a widespread focus on the numerical ratio effect in behavioral and neuroimaging studies, only one study to date, to the best of our knowledge, has investigated the effect of ratio on eye-movement patterns during nonsymbolic comparison. Odic and Halberda (2015) reported a decreasing number of fixations on the correct stimulus, number of saccades, and probability of first fixation being on the correct stimulus, as task difficulty increased. In the case of symbolic number processing, to our knowledge, only two studies to date have investigated the effect of ratio on eye-movement patterns. Merkley and Ansari (2010) showed ratio effects for fixation dwell time (FD), fixation count (FC) and number of saccades (SC), with more difficult comparison trials eliciting more fixations, more saccades, and longer fixations. And, while several of their eye-movement measures correlated with task performance (i.e. reaction time (RT) and accuracy rate) when calculated as an overall mean, their ratio effects did not significantly correlate with performance ratio effects, leading the authors to suggest that eye-movement measures may index distinct processes from performance metrics. More recently, in an eye-tracking investigation of common and cross-notational comparison of whole numbers, fractions, and decimals, Hurst and Cordes (2016) found that ratio effects were only apparent for fixation dwell time on the smaller number, although participants generally looked longer at the larger number than the smaller number.

In summary, performance measures of both nonsymbolic and symbolic number comparison have been related to the development of arithmetic skills. The presence of a numerical ratio effect for both formats suggests a potentially shared underlying semantic representation, however, little is known regarding how visuo-perceptual processes are related to the numerical ratio effect, whether they are shared across numerical formats, and whether they relate to math competence independently of performance outcomes. Thus, the present study addresses the following two questions. First, do eye-movement patterns during symbolic and nonsymbolic numerical comparison indicate a shared underlying semantic representation, shared visuo-perceptual processing mechanisms, or both? Second, do eye-movement measures provide unique information about the processing of numerical magnitudes that relates to individual differences in math competence, beyond that accounted for by task performance (i.e., error rate and reaction time)?

In regards to our first question, if cognitive mechanisms for processing symbolic and nonsymbolic number share semantic representation, we would expect to see significant, positive correlations between performance measures across formats and perhaps eye-movement measures as well. If they share visuo-perceptual processing mechanisms, eye-movement measures should be positively correlated, but if *only* visuo-perceptual processing mechanisms and not semantic representations are shared, then eye-movement measures would be expected to correlate across formats in the absence of cross-format performance correlations. Based on previous research calling into question the reliability of ratio effects (Lyons, Nuerk, & Ansari, 2015), we explore both total means and ratio effects for each measure.

Regarding the second question, if visuo-perceptual processing of dot arrays or Arabic digits is related to math competence, beyond cognitive aspects captured by task performance measures, we would expect to see a relationship between eye-movement patterns and math competence even after controlling for error rate and reaction time, and general cognitive ability. Beyond this, spatial distribution of eye-movement patterns may provide additional details about individual differences in perception. For example, number of fixations on singular stimuli during the nonsymbolic task may serve as a proxy for enumeration strategy. Or, fixation dwell time on single Arabic digits, while controlling for mean reaction time, may be one way to capture visual fluency in processing number symbols.

If eye-movement patterns do prove to be a unique predictor of math outcomes, they could serve as an additional tool for future research investigating individual differences in the cognitive mechanisms underlying representation of number and their relation to math, such as visual fluency with digits apart from the acuity of the semantic representation. Similar research has already begun to make progress for understanding the role of binocular coordination in dyslexia (Hutzler, Kronbichler, Jacobs, & Wimmer, 2006; Kirkby, Blythe, Drieghe, & Liversedge, 2011).

2. Methods

2.1. Participants

Seventy-three undergraduate students completed participation in the study. Three students were excluded due to incomplete or inaccurate eye-tracking data. Of those three, one student misunderstood instructions, one student did not respond in the appropriate response window, and there was distracting ambient noise during one experi-

Table 1
Task accuracy, standardized math measures, IQ, and demographic information.

N = 56 (36 female)	Mean	SD	Range
Age (years)	19.4	0.78	18.3–21.4
Symbolic magnitude comparison accuracy (%)	93.1	4.52	81–100
Nonsymbolic magnitude comparison accuracy (%)	78.1	7.72	60–93
Symbolic magnitude comparison RT (ms)	684.79	63.32	504.30–806.90
Nonsymbolic magnitude comparison RT (ms)	686.52	64.91	511.85–794.76
WCJ-III Math Fluency	109.3	12.3	86–142
WCJ-III Calculation	120.2	11.3	89–146
WCJ-III Math Composite Calculation Skills	118.5	11.1	89–146
Verbal IQ (KBIT-2)[a]	110.7	10.9	78–131
Nonverbal IQ (KBIT-2)[b]	107.3	11.4	86–132
Composite IQ (KBIT-2)[a]	111.0	9.86	84–132

Note. WCJ-III: Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001). RT: reaction time.
[a] $n = 54$.
[b] $n = 55$.

mental session. Three other participants were excluded due to low accuracy on the nonsymbolic comparison task. Their performance was not significantly above chance (50%). A binomial test on 72 trials of a two-alternative forced-choice task yields a threshold of 59.7% accuracy to be above chance level at $p < 0.05$. Eleven other participants were also excluded due to a lack of basal scores for the Woodcock-Johnson III Calculation subtest. A total of 56 subjects (36 female) were included in subsequent analyses of behavioral and eye-movement data. Participants' ages ranged from 18 to 21 years ($M = 19.4$, $SD = 0.78$). Task accuracy, standardized math measures, IQ, and demographic information are presented in Table 1.

2.2. Tasks, stimuli, and design

2.2.1. Nonsymbolic numerical magnitude comparison

In the nonsymbolic comparison task, two dot arrays were presented simultaneously on either side of a centrally presented fixation point. Participants were instructed to choose the set with more dots as quickly and as accurately as possible. Sets of black dots ranging from 6 dots to 15 dots were presented on a white background. Set sizes below 6 were not included to avoid the influence of subitizing on the eye-movement and behavioral measures. Fourteen different ratios were presented: 0.50, 0.56, 0.57, 0.60, 0.63, 0.67, 0.71, 0.75, 0.78, 0.80, 0.83, 0.86, 0.88, and 0.89. For analysis, the total 72 trials were binned into 36 small ratio trials (ratio < 0.7) and 36 large ratios trials (ratio > 0.7). Stimuli were presented for 1000 milliseconds (ms), followed by the presentation of a central fixation dot. The paradigm was programmed to proceed to the next trial once fixation on the central dot had been maintained for 1000 ms in order to ensure that fixation location started at the same location across all participants and all trials. Responses were made via left- and right-hand button press. Response side was fully counterbalanced. To control for the possibility that participants might choose a strategy based on visual cues rather than number of dots, the following visual properties of dot sets were varied using the MATLAB code recommended by Gebuis and Reynvoet (2011) to generate stimuli: convex hull (area extended by a stimulus), total surface area (aggregate value of dot surfaces), average dot diameter, and density (convex hull divided by total surface area). None of the visual parameters significantly correlated with number of dots presented. Trial order was randomized at the subject level.

2.2.2. Symbolic numerical magnitude comparison

The symbolic comparison task was identical to the nonsymbolic task in all details except that the stimuli consisted of pairs of single-digit Arabic numerals ranging from 2 to 9 presented simultaneously. The same fourteen ratios were presented as in the nonsymbolic task and the trials were again binned into 36 small ratio trials (ratio < 0.7) and 36 large ratios trials (ratio > 0.7) for analysis.

2.2.3. Mathematical competence

Mathematical competence was assessed using the Math Fluency and Calculation subtests of the Woodcock-Johnson III Tests of Achievement (WCJ-III) (Woodcock et al., 2001). The Math Fluency subtest requires participants to answer as many simple addition, subtraction, and multiplication problems as possible within a 3-minute period. The Calculation subtest, on the other hand, is untimed, and requires participants to complete as many calculation items as possible that increase in difficulty, ranging from simple arithmetic to calculus. A weighted, Composite Calculation Skills cluster score comprising both subtests was computed for each participant using the WCJ scoring software. Age-normed standard scores were used for all analyses. Kolmogorov-Smirnov test of normality with Lilliefors significance correction demonstrated that all the math measures were normally distributed (all p-values > 0.072).

2.2.4. IQ

Nonverbal IQ, Verbal IQ, and Composite IQ estimates were obtained for each participant based on the Kaufman Brief Intelligence Test, second edition (KBIT-2; Kaufman & Kaufman, 2004). IQ tests were not completed for four students. Due to technical issues with the eye-tracking equipment, time did not allow for IQ testing of one participant. For 3 participants, basal scores were not properly established in the verbal IQ subtest of the KBIT-2. Composite IQ scores were normally distributed (all p-values > 0.200) and centered slightly above average ($M = 111.0$, range = 84–132), suggesting a wide range of overall cognitive ability representative of a typically developing sample. Composite IQ scores were positively correlated with the Composite Calculation Skills [$r(52) = 0.327$, $p = 0.016$], but not with mean error rate in both symbolic and nonsymbolic comparison ($ps > 0.618$).

2.3. Eye-tracking apparatus

Eye-movement data were recorded at 1000 Hz using an EyeLink 1000 eye-tracking system (SR Research, Kanata, Ontario, Canada), which achieves a typical spatial resolution accuracy down to 0.5° of visual angle. All stimuli were presented on a 21.5″ monitor driven at a refresh rate of 60 Hz and resolution of 1920 × 1080 pixels using Eyelink's Experiment Builder software (Version 1.10.165). The 47.7 × 26.8 cm screen subtended 44.7° × 26.0° at the viewing distance of 58 cm. The stimuli were arrays of dots (nonsymbolic) and Arabic digits (symbolic), centered at 12.6° left and right of the center fixation point. Dot arrays were presented within square 420 × 420 pixel images (10.1° × 10.1°). Arabic digits were also presented at 12.6° left and right of center fixation, but were 18 × 27 pixels (0.42° × 0.65°) in size. Participants' viewing distance and viewing angle were controlled by the use of a chin rest. Two rectangular interest areas were defined that were slightly larger than each stimulus presentation area (left and right), remained the same size for both nonsymbolic and symbolic conditions, and were the defined areas of interest for all subsequent analyses. For example, fixation counts for a single trial would include all fixations landing within either of these two regions of interest. Responses were recorded via left and right trigger buttons on a Microsoft Sidewinder USB gamepad.

2.4. Procedure

Testing took place in a quiet room during a single testing session. The room was dimly lit during eye-tracking sessions. For each session the first two tasks were the nonsymbolic and symbolic magnitude comparison tasks, counterbalanced among subjects to avoid an order effect. The WCJ-III and the KBIT-2 were administered after the eye-tracking tasks were completed.

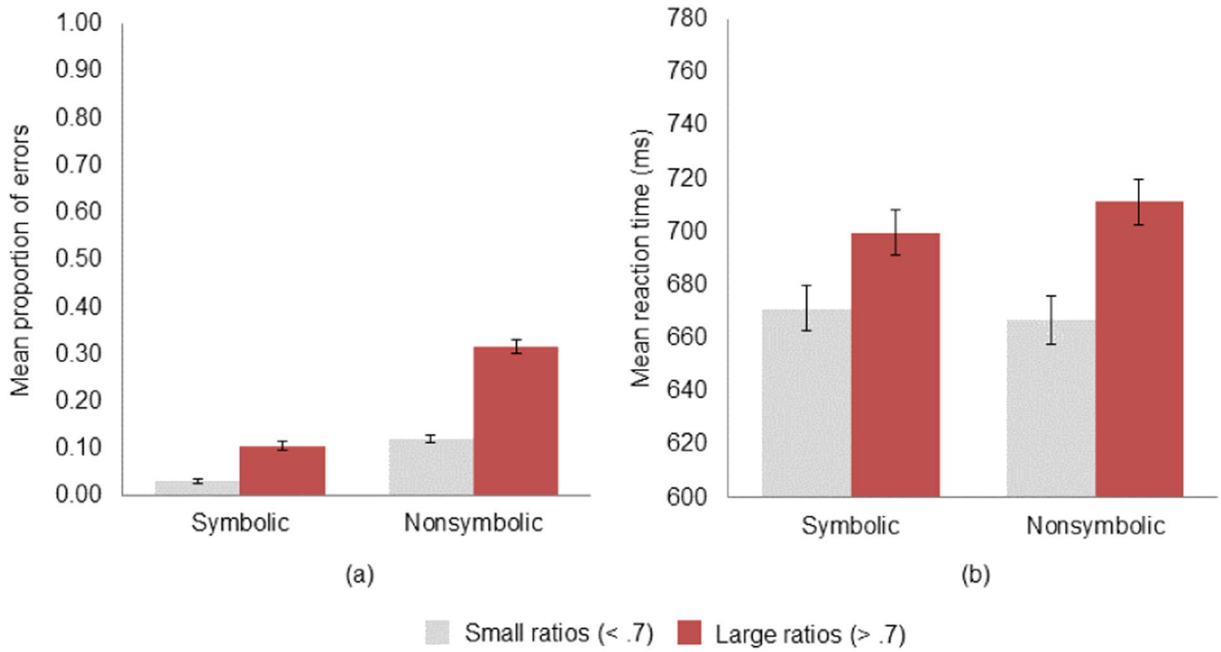

Fig. 1. Effects of ratio on mean error rate and RT. Lower mean error rate and RT were found for easier, small ratio trials than for harder, large ratio trials. A larger ratio effect on both (a) error rate and (b) RT was observed for nonsymbolic than for symbolic comparison. Error bars denote standard errors of the means.

2.5. Analyses

In order to investigate the effects of ratio and format on task performance, two 2 × 2 (small vs. large ratio × symbolic vs. nonsymbolic) repeated-measures Analysis of Variance (ANOVAs) were performed for mean error rate and mean RT. Post-hoc paired-samples *t*-tests were conducted to examine any significant interaction effects.

Two additional ANOVAs (2 × 2) were performed for mean fixation count (FC) and mean fixation dwell time (FD) on the interest areas (i.e., the average total number of fixations and average total time spent fixating on the numerical stimuli on a trial respectively) in order to investigate the effects of ratio and format on measures of eye-movements. Post-hoc paired-samples *t*-tests were also conducted to examine any significant interaction effects. Incorrect trials and trials in which the RTs exceeded three standard deviations of an individual's mean were excluded from RT analyses, eye-movement analyses, and all subsequent analyses (< 0.001% of all trials). We included only the fixations that last for at least 50 milliseconds (ms), and only analyzed eye-movement data that fell within a specified interest period extending from the onset of the stimuli until a response was made. Saccades between interest areas were not analyzed as they are dependent on and were correlated almost perfectly with mean total FC in both symbolic [$r(54) = 0.95$, $p < 0.001$] and nonsymbolic [$r(54) = 0.91$, $p < 0.001$] comparison tasks.

To explore the spatial distribution of eye-movement measures (i.e., fixations on the correct (numerically larger) and incorrect (numerically smaller) numerosities), we performed (i) one 2 × 2 repeated-measures ANOVA on the proportion of first fixations on the larger numerosity with format (symbolic, nonsymbolic) and ratio size (small, large), and (ii) four 2 × 2 repeated-measures ANOVAs for FC and FD with correctness of stimulus (fixation on correct stimulus, fixation on incorrect stimulus) and ratio size (small, large). Post-hoc paired-samples *t*-tests were conducted to examine any significant interaction effects.

To assess the relationships among the performance and eye-movement variables, as well as between their respective ratio effects, within and across each format, Pearson's product-moment bivariate correlations were computed. To correct for multiple hypothesis testing, the critical *p*-values for each set of correlations were adjusted using the Benjamini-Hochberg's (B-H) False Discovery Rate method with $\alpha_{FDR} = 0.05$ (Benjamini & Hochberg, 1995), which provides a good balance between controlling for false positives and power for detecting weaker, but significant relationships. Whenever an uncorrected *p*-value was < 0.05, but greater than the B-H critical *p*-value, both the uncorrected and the B-H critical *p*-values were reported. While it is common for the magnitude of the ratio effect to be computed using formulas such as (large ratio − small ratio) / small ratio (e.g., Holloway & Ansari, 2009) to account for individual differences in the baseline of each measure, in this study, we simply used the difference between large and small ratios for two reasons. Firstly, using the (large ratio − small ratio) / small ratio formula does not yield meaningful ratio effects on error rate because there are many instances of zero error rate on the easier small ratio trials that would render the denominator to be zero. Secondly, doing away with the denominator in the formula allowed us to compare our findings directly with those by Merkley and Ansari (2010). The ratio effects for RT and eye-movement measures computed with and without the denominator were very highly correlated in both formats (all $rs > 0.948$, $ps < 0.001$) suggesting no meaningful difference between the formula based vs. subtraction based calculation.

Finally, to examine the relationships among eye-movement measures, measures of task performance, and math competence, bivariate correlations were computed between standardized WCJ-III Math Calculation Skills cluster scores and each of the performance and eye-movement measures, as well as their respective ratio effects. To further clarify the unique relationship between calculation skills and the relevant eye-movement measures, a hierarchical regression was also performed controlling for task performance (i.e. error rate and RT) and composite IQ. This regression clarifies the relation between eye-movement measures and calculation skills beyond that of task performance measures and domain-general cognitive abilities.

3. Results

3.1. Performance measures

3.1.1. Error rate

Mean error rate was lower for symbolic compared to nonsymbolic trials [$F(1, 55) = 202.41$, $p < 0.001$, $\eta_p^2 = 0.79$] and lower for large ratio compared to small ratio trials [$F(1, 55) = 278.90$, $p < 0.001$, $\eta_p^2 = 0.84$] (Fig. 1a). The interaction between format and ratio size was also significant, [$F(1, 55) = 75.85$, $p < 0.001$, $\eta_p^2 = 0.58$], revealing that a larger ratio effect was observed for nonsymbolic compared to symbolic trials. Post-hoc analyses revealed a lower error rate for small ratio than for large ratio trials in symbolic comparison [$t(55) = 9.26$, $p < 0.001$, Cohen's $d = 1.36$], as well as in nonsymbolic comparison [$t(55) = 15.26$, $p < 0.001$, $d = 2.14$].

3.1.2. Reaction time (RT)

Mean RT on small ratio trials was shorter than that for large ratio trials [$F(1, 55) = 205.53$, $p < 0.001$, $\eta_p^2 = 0.79$], but there was no effect of format [$F(1, 55) = 0.22$, $p = 0.640$, $\eta_p^2 = 0.004$] (Fig. 1b). The interaction between format and ratio was significant [$F(1, 55) = 13.47$, $p = 0.001$, $\eta_p^2 = 0.20$], indicating a larger ratio effect for nonsymbolic than symbolic comparisons. Post-hoc analyses revealed shorter mean RT for small ratio than for large ratio trials in symbolic comparison [$t(55) = 10.10$, $p < 0.001$, $d = 0.45$], as well as in nonsymbolic comparison [$t(55) = 11.85$, $p < 0.001$, $d = 0.67$].

In sum, ratio modulated both error rate and RT, such that large ratio trials elicited more errors and longer RTs than small ratio trials. The effect of ratio was greater for nonsymbolic comparison than for symbolic comparison.

3.2. Eye-movement measures

3.2.1. Fixation count

There were fewer fixations on small ratio trials than on large ratio trials [$F(1, 55) = 46.96$, $p < 0.001$, $\eta_p^2 = 0.46$], but there was no effect of format [$F(1, 55) = 0.13$, $p = 0.718$, $\eta_p^2 = 0.002$] (Fig. 2a). The interaction between format and ratio size was also significant [$F(1, 55) = 12.01$, $p = 0.001$, $\eta_p^2 = 0.18$], indicating a larger ratio effect for nonsymbolic than for symbolic comparison. Post-hoc analyses revealed fewer fixations for small ratio than for large ratio trials in nonsymbolic comparison [$t(55) = 6.27$, $p < 0.001$, $d = 0.45$], and in symbolic comparison [$t(55) = 2.44$, $p = 0.018$, $d = 0.11$].

3.2.2. Fixation dwell time

Mean fixation dwell time was shorter for symbolic compared to nonsymbolic comparison [$F(1, 55) = 8.96$, $p = 0.004$, $\eta_p^2 = 0.14$] and for small ratio compared to large ratio trials [$F(1, 55) = 159.82$, $p < 0.001$, $\eta_p^2 = 0.74$] (Fig. 2b). The interaction between format and ratio size on mean fixation dwell time was also significant, [$F(1, 55) = 7.07$, $p = 0.010$, $\eta_p^2 = 0.11$], indicating a larger ratio effect for nonsymbolic than symbolic comparison. Post-hoc analyses revealed shorter fixation dwell times for small ratio than for large ratio trials in nonsymbolic [$t(55) = 9.99$, $p < 0.001$, $d = 0.50$], and in symbolic comparison [$t(55) = 7.25$, $p < 0.001$, $d = 0.30$].

Hence, similar to the performance measures, ratio modulated fixation count and fixation dwell time, such that large ratio trials elicited more and longer fixations than small ratio trials. The effect of ratio was also greater for nonsymbolic comparison than for symbolic comparison.

3.3. Spatial distribution of fixations

To gain additional insights on the visual processing mechanisms underlying magnitude comparison, we examined the spatial distribution of fixations. Firstly, we asked whether participants used peripheral visual processing to complete the task by analyzing the proportion of first fixations on the larger (correct) numerosity, as information from peripheral vision would be required in order to attend first to the correct stimulus. Secondly, we asked whether the observed effects on the mean total fixation count (FC) and fixation dwell time (FD) were driven by fixations on the larger or smaller numerosity in each format and whether this differed by ratio.

3.3.1. Proportion of first fixations on larger numerosity

The mean proportion of first fixations on the larger numerosity for nonsymbolic comparison was significantly higher than that for symbolic comparison [$F(1, 55) = 6.54$, $p = 0.013$, $\eta_p^2 = 0.11$] (Fig. 3). There was also a higher mean proportion of first fixations on the larger numerosity on small ratio trials than on large ratio trials [$F(1, 55) = 18.25$, $p < 0.001$, $\eta_p^2 = 0.25$]. The interaction between format and ratio size was not significant [$F(1, 55) = 0.65$, $p = 0.424$, $\eta_p^2 = 0.01$].

Proportion of first fixation on the larger numerosity greater than

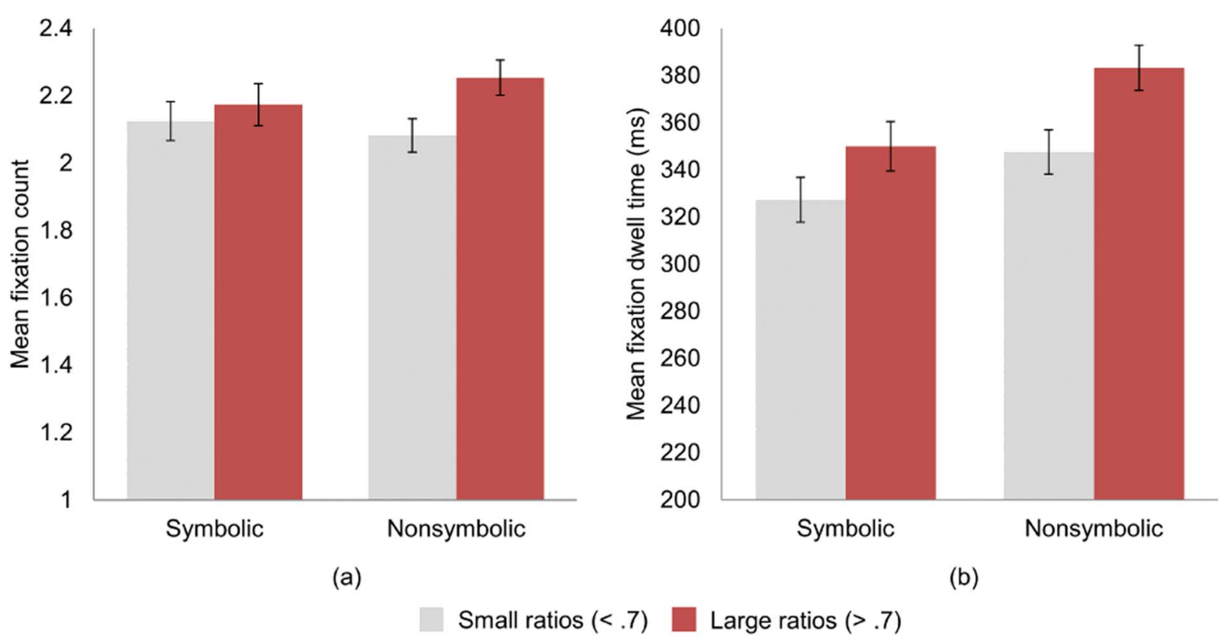

Fig. 2. Effects of ratio on mean fixation count and fixation dwell time. Lower mean fixation count and fixation dwell time were found for easier, small ratio trials than for harder, large ratio trials. A larger ratio effect on both (a) fixation count and (b) fixation dwell time was observed for nonsymbolic than for symbolic comparison. Error bars denote standard errors of the means.

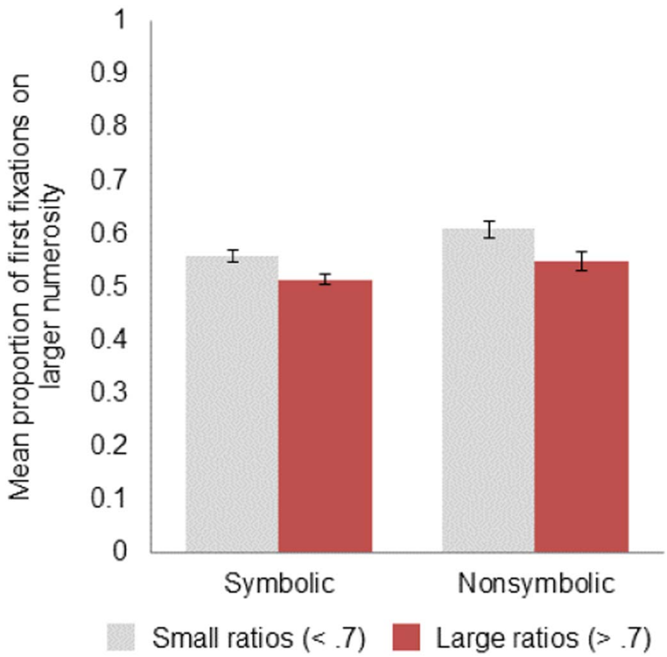

Fig. 3. Effects of ratio on mean proportion of first fixations on larger numerosity. Ratio effects on proportion of first fixations on larger numerosity were found for both symbolic and nonsymbolic comparisons, but they did not differ between formats. Error bars denote standard errors of the means.

chance (0.5) in both symbolic [$t(55) = 4.79$, $p < 0.001$, $d = 1.29$] and nonsymbolic comparisons [$t(55) = 6.22$, $p < 0.001$, $d = 1.68$], indicating that participants were receiving some information through peripheral vision about the numerical magnitude of the stimuli. Nonetheless, it is worthwhile to note that even though peripheral vision played a role in magnitude comparison at the group level, there were large individual variations, and the variation appeared to be greater for nonsymbolic comparison (from 0.37 to 0.82) than it was for symbolic comparison (from 0.39 to 0.65). Moreover, participants fixated on both stimuli at least once on 76.6% (SD = 18.0%, range = 10–100%) of trials for symbolic comparison, and on 77.3% (SD = 18.7%, range = 10–100%) of trials for nonsymbolic comparison. The fact that participants did not always fixate on both stimuli before making correct responses suggests that peripheral processing might have been sufficient for some trials, and for some participants. However, on a majority of the trials, most participants fixated on both stimuli before making a correct response, which suggests that foveal processing was the predominant visual strategy.

3.3.2. Fixation count on correct and incorrect numerosities (symbolic)

There were fewer fixations on the smaller, incorrect numerosity than on the larger, correct numerosity [$F(1, 55) = 134.10$, $p < 0.001$, $\eta_p^2 = 0.71$] and fewer fixations on small ratio trials [$F(1, 55) = 5.942$, $p = 0.018$, $\eta_p^2 = 0.10$] (Fig. 4a). The interaction between correctness and ratio size was also significant [$F(1, 55) = 12.01$, $p = 0.001$, $\eta_p^2 = 0.18$], indicating a larger ratio effect for FC on smaller, incorrect numerosity than larger, correct numerosity. Post-hoc analyses revealed fewer fixations on the smaller, incorrect numerosity for small ratio than for large ratio trials [$t(55) = 4.31$, $p < 0.001$, $d = 0.30$], but no effect of ratio on FC on the larger, correct numerosity [$t(55) = 1.63$, $p = 0.108$, $d = 0.11$].

3.3.3. Fixation count on correct and incorrect numerosities (nonsymbolic)

Main effects were similar to those observed for symbolic comparison. Specifically, there were fewer fixations on the smaller, incorrect numerosity than on the larger, correct numerosity [$F(1, 55) = 204.95$, $p < 0.001$, $\eta_p^2 = 0.78$] and fewer fixations on small ratio trials than on large ratio trials [$F(1, 55) = 39.31$, $p < 0.001$, $\eta_p^2 = 0.42$] (Fig. 4b). Unlike symbolic comparison, however, the interaction between correctness and ratio size was not significant [$F(1, 55) = 1.58$, $p = 0.214$, $\eta_p^2 = 0.03$].

3.3.4. Fixation dwell time on correct and incorrect numerosities (symbolic)

Mean FD for smaller, incorrect numerosity was shorter than that for larger, correct [$F(1, 55) = 292.92$, $p < 0.001$, $\eta_p^2 = 0.84$] (Fig. 5a). Mean FD on small ratio trials was also shorter than on large ratio trials [$F(1, 55) = 52.56$, $p < 0.001$, $\eta_p^2 = 0.49$]. The interaction between correctness and ratio size on mean fixation dwell time was also significant [$F(1, 55) = 8.72$, $p = 0.005$, $\eta_p^2 = 0.14$], indicating that there was a larger ratio effect for the smaller, incorrect numerosity than for the larger, correct numerosity. Post-hoc analyses revealed shorter fixations on the smaller, incorrect numerosity for small ratio than for large ratio trials [$t(55) = 6.66$, $p < 0.001$, $d = 0.49$], but no effect of ratio on the fixation dwell time on the larger, correct numerosity [$t(55) = 1.15$, $p = 0.255$, $d = 0.09$].

3.3.5. Fixation dwell time on correct and incorrect numerosities (nonsymbolic)

Mean fixation dwell time for smaller, incorrect numerosity was shorter than that for larger, correct numerosity [$F(1, 55) = 260.60$, $p < 0.001$, $\eta_p^2 = 0.83$] (Fig. 5b). Mean fixation dwell time on small ratio trials was also shorter than on large ratio trials [$F(1, 55) = 99.87$, $p < 0.001$, $\eta_p^2 = 0.65$]. The interaction between correctness and ratio size on mean fixation dwell time was also significant, [$F(1, 55) = 18.09$, $p < 0.001$, $\eta_p^2 = 0.25$], indicating that there was a larger ratio effect for the smaller, incorrect numerosity than for the larger, correct numerosity. Post-hoc analyses revealed shorter fixations on the smaller, incorrect numerosity for small ratio than for large ratio trials [$t(55) = 9.81$, $p < 0.001$, $d = 0.72$], and also on the larger, correct numerosity for small ratio than for large ratio trials [$t(55) = 2.21$, $p = 0.032$, $d = 0.17$].

Taken together, our findings suggest that both peripheral processing and foveal processing were modulated by ratio and format to different extents. Participants possibly used peripheral processing to focus their attention on the larger numerosity, though to a greater degree for nonsymbolic than symbolic comparison, and also for small ratio (easier) trials than for the large ratio (harder) trials. However, effects of ratio on FC and FD were greater for the smaller, incorrect numerosity than for the larger, correct numerosity. These findings suggest that during subsequent foveal processing of both stimuli, the effects of ratio and format on total FC and FD were not due to a global increase in attention to both numerosities, but rather to a greater increase in attention to the smaller, incorrect numerosity rather than to the larger, correct numerosity when the ratios were larger and more difficult.

3.4. Relationships between performance and eye-movement measures

3.4.1. Error rate and RT

Error rate and RT were positively correlated in both symbolic [$r(54) = 0.318$, $p = 0.002$] and nonsymbolic comparisons [$r(54) = 0.459$, $p < 0.001$], suggesting that there was no speed-accuracy tradeoff.

3.4.2. Eye-movement measures and RT

Given the hypothesized inherent relationship between eye-movement measures and RT, we first examined if mean total FC and FD (i.e., on both numerosities), as well as proportions of first fixations, FC, and FD on the larger numerosity were related to one another, and to mean RT (see Table 2).

3.4.2.1. Symbolic. Most eye-movement measures were correlated with one another. An exception was that proportion of first fixations on the larger numerosity was not related to the proportion of FC and FD on the larger numerosity. RT was positively correlated with total FC and FD,

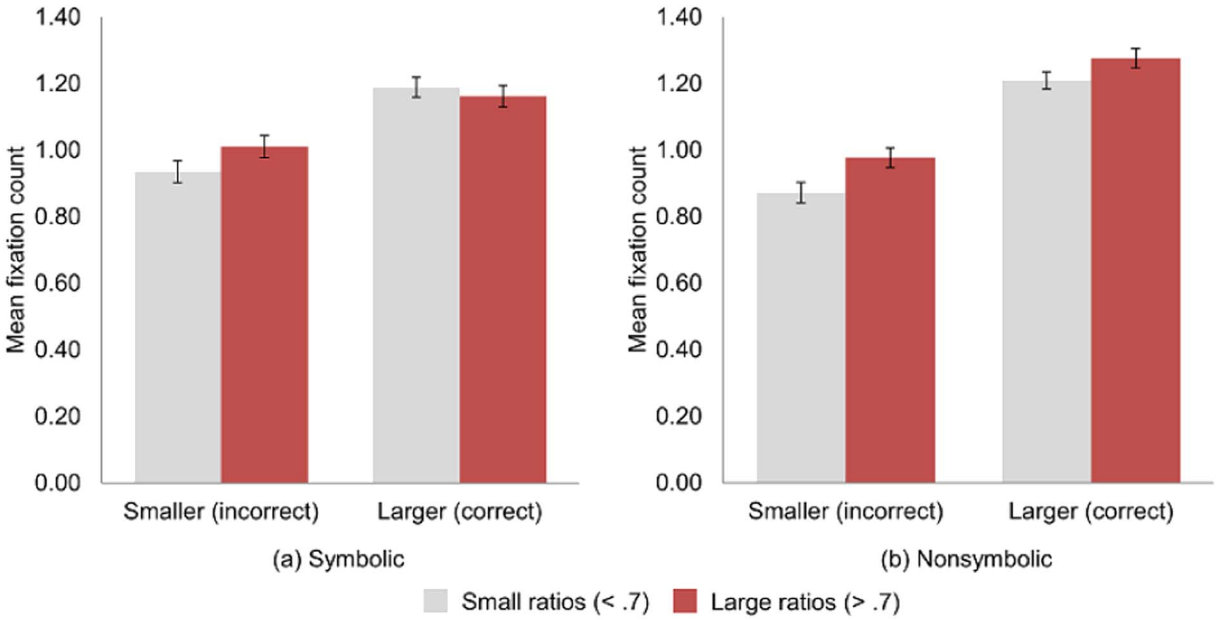

Fig. 4. Effects of ratio on mean fixation count on smaller and larger numerosities. (a) For symbolic comparison, a ratio effect was observed for number of fixations on the smaller, incorrect numerosity, but not for the larger, correct numerosity. (b) For nonsymbolic comparison, number of fixations on both the smaller, incorrect numerosity, and the larger, correct numerosity, were modulated by ratio to a similar extent. Error bars denote standard errors of the means.

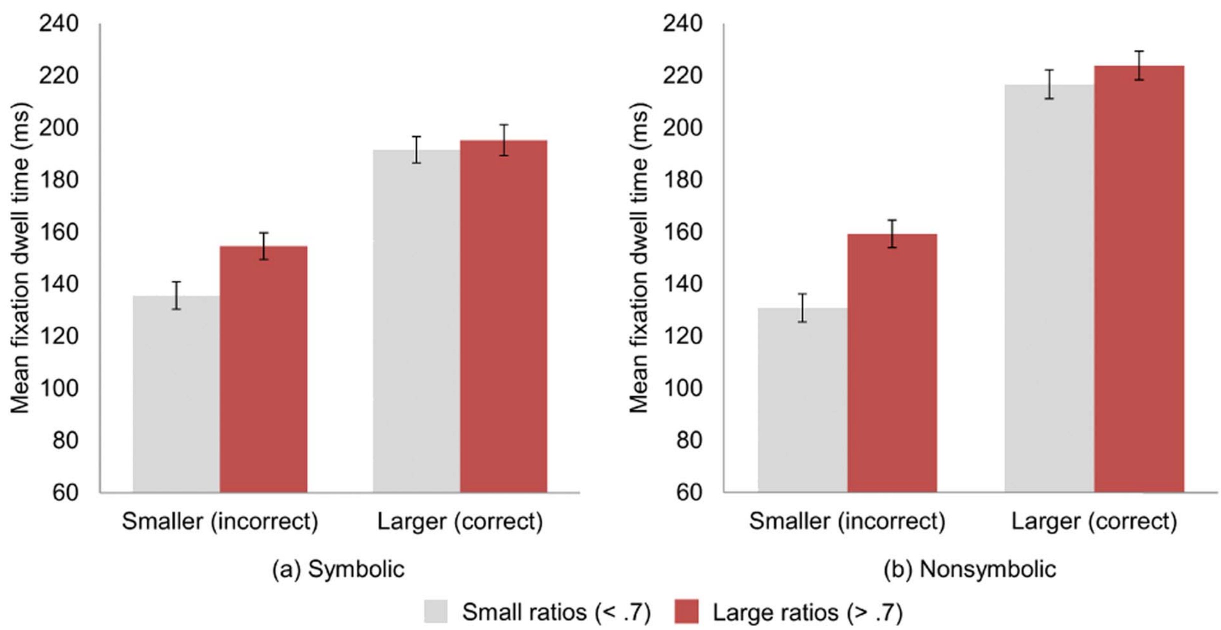

Fig. 5. Effects of ratio on mean fixation dwell time on smaller and larger numerosities. (a) For symbolic comparison, a ratio effect was observed for fixation dwell time on the smaller, incorrect numerosity, but not for the larger, correct numerosity. (b) For nonsymbolic comparison, a ratio effect was observed for fixation dwell time on both the smaller, incorrect numerosity, and the larger, correct numerosity. The size of the ratio effect was greater for the smaller, incorrect numerosity than for the larger, correct numerosity. Error bars denote standard errors of the means.

but not with any of the fixation spatial measures.

3.4.2.2. Nonsymbolic. Most of the associations observed in symbolic comparison were also observed in nonsymbolic comparison. However, in nonsymbolic comparison, proportion of first fixations on the larger numerosity was not related to total FC, but with proportions of FC and FD on the larger numerosity. RT was also related to proportion of fixation count on larger numerosity in nonsymbolic comparison.

In summary, as predicted, most eye-movement measures were related with one another, and with RT, in both formats. However, there were some associations involving the fixation spatial patterns that were distinct between symbolic and nonsymbolic comparisons. These suggest that fixation spatial patterns were not as inherently related to RT as total FC and FD did, and the extent of dissociation between them appeared to be greater in symbolic than nonsymbolic comparison. RT, FC, and FD were correlated across formats, but spatial measures of eye movements were not related across formats.

3.4.3. Eye-movement measures and error rate

To examine the relationship between error rate and eye-movement measures, while accounting for their individual relationships with RT, we performed partial correlations between mean error rate and mean total FC and FD, as well as proportions of first fixations, FC, and FD on the larger numerosity, while controlling for mean RT.

Table 2
Bivariate correlations of RT and eye-movement measures for symbolic and nonsymbolic comparison.

Measure (N = 56)	1	2	3	4	5	6
1. RT	**0.580***	0.638*	0.691*	− 0.220	− 0.237	− 0.186
2. FC	0.600*	**0.566***	0.840*	− 0.436*	− 0.508*	− 0.491*
3. FD	0.830*	0.757*	**0.563***	− 0.383*	− 0.522*	− 0.519*
4. Prop. first fixations on larger numerosity	− 0.269†	− 0.236	− 0.419*	**− 0.166**	0.237	0.221
5. Prop. FC on larger numerosity	− 0.283*	− 0.412*	− 0.362*	0.581*	**0.207**	0.938*
6. Prop. FD on larger numerosity	− 0.202	− 0.417*	− 0.287*	0.435*	0.908*	**0.116**

Note. RT: reaction time. FC: total fixation count. FD: total fixation dwell time. Upper triangle comprises correlations of the measures for symbolic comparison, and lower triangle for nonsymbolic comparison. The diagonal, which is in bold, comprises correlations of the same measures across format.
* $p < 0.05$ with Benjamini-Hochberg FDR correction.
† $p < 0.05$ without Benjamini-Hochberg FDR correction.

Table 3
Bivariate correlations of ratio effects on RT and eye-movement measures for symbolic and nonsymbolic comparison.

Measure (N = 56)	1	2	3	4	5	6
1. RT RE	**0.182**	0.321†	0.247	− 0.085	0.236	0.168
2. FC RE	0.485*	**− 0.104**	0.642*	− 0.017	− 0.073	− 0.081
3. FD RE	0.707*	0.620*	**− 0.055**	− 0.083	0.043	0.046
4. Prop. first fixations on larger numerosity	− 0.077	− 0.178	− 0.128	**0.181**	0.257	0.104
5. Prop. FC on larger numerosity	− 0.085	− 0.214	− 0.056	0.587*	**0.043**	0.844*
6. Prop. FD on larger numerosity	− 0.132	− 0.210	− 0.046	0.354*	0.830*	**− 0.061**

Note. RT: reaction time. RE: ratio effect. FC: total fixation count. FD: total fixation dwell time. Upper triangle comprises correlations of the measures for symbolic comparison, and lower triangle for nonsymbolic comparison. The diagonal, which is in bold, comprises correlations of the same measures across format.
* $p < 0.05$ with Benjamini-Hochberg FDR correction.
† $p < 0.05$ without Benjamini-Hochberg FDR correction.

3.4.3.1. Symbolic. Error rate did not correlate with any of the eye-movement measures after controlling for RT (all *p*-values > 0.157).

3.4.3.2. Nonsymbolic. Unlike in symbolic comparison, error rate correlated positively with both proportions of FC [$r(53) = 0.355$, $p = 0.008$] and FD [$r(53) = 0.452$, $p = 0.001$] on the larger numerosity after controlling for RT. The other associations between error rate and eye-movement measures were not significant (all *p*-values > 0.085).

3.4.4. Cross-format

Error rate did not correlate across formats [$r(54) = 0.263$, $p = 0.05$, Benjamini-Hochberg (B-H) critical *p*-value < 0.029]. Only RT, total FC, and total FD correlated positively across formats (see Table 2). Total FC ($r(52) = 0.642$, $p < 0.001$) and FD ($r(52) = 0.588$, $p < 0.001$) still remained positively correlated across formats even after controlling for RTs for both symbolic and nonsymbolic comparisons. The fixation spatial patterns also did not correlate across formats (all *p*-values > 0.125).

In summary, error rate was largely unrelated to the eye-movement measures in both formats, with the exception of proportions of FC and FD on the larger numerosity for nonsymbolic comparison. The dissociation between visual perceptual and post-perceptual cognitive mechanisms was also substantiated by the cross-format associations for RT, total FC and FD, but not for error rate.

3.5. Relationships between performance and eye-movement ratio effects

3.5.1. Error rate and RT ratio effects

Error rate ratio effect and RT ratio effect were not correlated in symbolic comparison [$r(54) = -0.140$, $p = 0.302$], but were negatively correlated in nonsymbolic comparison [$r(54) = -0.431$, $p < 0.001$].

3.5.2. Eye-movement and RT ratio effects

Similar to the raw performance and eye-movement measures above, we first examined if the ratio effects on mean total FC, total FD, and proportions of first fixations, FC, and FD on the larger numerosity were related with one another, and with the ratio effect on mean RT (see Table 3).

3.5.2.1. Symbolic. FC ratio effect was positively correlated with FD ratio effect. Similarly, proportions of FC and FD on the larger numerosity ratio effect were positively correlated. The other eye-movement measures did not correlate with one another (all *p*-values > 0.056). RT ratio effect did not correlate with any of the eye-movement ratio effects (all *p*-values > 0.067). At an uncorrected threshold of $p < 0.05$, however, RT ratio effect was positively correlated with FC ratio effect [$r(54) = 0.321$, $p = 0.016$, B-H critical *p*-value < 0.01].

3.5.2.2. Nonsymbolic. All of the associations observed in symbolic comparison were also observed in nonsymbolic comparison. However, in nonsymbolic comparison, RT ratio effect correlated positively with both FC and FD ratio effects. Ratio effects on the proportions of first fixations, FC, and FD on larger numerosity were also correlated with one another. All other associations were not significant (all *p*-values > 0.114).

In summary, ratio effects on eye-movement measures were rarely related with one another, and with RT, but more so for symbolic than nonsymbolic comparison.

3.5.3. Eye-movement and error rate ratio effects

To examine the relationship between the ratio effects on error rate and eye-movement measures, while accounting for their individual relationships with RT ratio effect, we performed partial correlations between the ratio effects on mean error rate and mean total FC, total FD, and proportions of first fixations, FC, and FD on the larger numerosity, while controlling for the ratio effect on RT.

3.5.3.1. Symbolic. Error rate ratio effect did not correlate with any of the eye-movement ratio effects (all *p*-values > 0.023, B-H critical *p*-value < 0.01). At an uncorrected threshold of $p < 0.05$, however, error rate ratio effect was positively correlated with the ratio effects on the proportions of FC [$r(53) = 0.306$, $p = 0.023$, B-H critical *p*-value < 0.01] and FD on the larger numerosity [$r(53) = 0.293$, $p = 0.030$, B-H critical *p*-value < 0.02].

3.5.3.2. Nonsymbolic. Similar to symbolic comparison, error rate ratio effect did not correlate with any of the eye-movement ratio effects (*ps* > 0.312).

3.5.4. Cross-format ratio effects

Error rate ratio effect did not correlate across formats [$r(54)$

= 0.159, p = 0.243]. The ratio effects of RT and eye-movement measures also did not correlate across formats (all p-values > 0.179) (see Table 3).

In summary, the ratio effects on performance measures (i.e. error rate and RT) were unrelated to the eye-movement ratio effects measures during symbolic comparison. RT ratio effects did correlate significantly with FC and FD ratio effects for nonsymbolic comparison however. Importantly, ratio effects were unrelated across formats. These findings further support the dissociation between visual perceptual and post-perceptual cognitive mechanisms at least within symbolic comparison, and a lack of shared visual perceptual and post-perceptual cognitive mechanisms between symbolic and nonsymbolic comparisons.

3.6. Relations between math competence, task performance, and eye-movement measures

3.6.1. Task performance and eye-movement measures

After correcting for multiple comparisons, the math calculation skills cluster score was significantly correlated only with total FD [r(54) = − 0.385, p = 0.003] in the symbolic comparison task, and not symbolic RT or symbolic error rate. At the uncorrected threshold of $p < 0.05$, math calculation skills was also correlated with symbolic error rate [r(54) = − 0.290, p = 0.030, B-H critical p-value < 0.011], symbolic total FC [r(54) = − 0.309, p = 0.020, B-H critical p-value < 0.007], and proportion of FD on the larger numerosity during the symbolic task [r(54) = 0.280, p = 0.037, B-H critical p-value < 0.014]. No significant correlations were observed for the nonsymbolic measures.

3.6.2. Task performance and eye-movement ratio effects

Math calculation skills was not correlated with any of the ratio effects on task performance measures or eye-movement measures (all p-values > 0.248) for either format.

To further investigate if visual processing fluency in symbolic comparison accounts for unique variance in math calculation skills over and above task performance measures and general cognitive ability, we performed hierarchical linear regression analyses of math calculation skills on total FD and FC separately, controlling for IQ, error rate, and RT during symbolic comparison on a sample of 54 participants who had the Composite IQ scores (Table 4). Given that only the symbolic comparison measures correlated with math competence, this analysis was not conducted for nonsymbolic comparison measures.

Results (Table 4 & Fig. 6) demonstrated that error rate and RT during symbolic comparison did not account for unique variance in calculation skills over and above IQ. Total FD, on the other hand, accounted for unique variance in calculation skills over and above IQ and performance measures, but total FC did not. Taken together, only total FD, and not performance measures such as error rate, was a robust predictor of calculation skills.

Table 4
Hierarchical linear regression analyses for IQ and symbolic comparison variables predicting calculation skills (N = 54).

Step/model	Predictor	β	R^2 (full model)	ΔR^2
Step 1			0.107*	0.107*
	Composite IQ	0.327*		
Step 2			0.187*	0.080
	Composite IQ	0.297*		
	Error rate	− 0.246		
	RT	− 0.084		
Step 3 Model 1			0.271*	0.084*
	Composite IQ	0.215		
	Error rate	− 0.256		
	RT	0.188		
	Fixation dwell time	− 0.412*		
Step 3 Model 2			0.235*	0.048
	Composite IQ	0.265*		
	Error rate	− 0.261		
	RT	0.095		
	Fixation count	− 0.285**		

Note. RT: reaction time.
* $p < 0.05$.
** $p < 0.01$.

4. Discussion

The current study investigated the relation between eye-movement and performance indices of nonsymbolic and symbolic numerical magnitude comparison and their relation to math calculation skills in order to address two principle questions. First, do eye-movement patterns indicate a shared underlying semantic representation across number formats, shared visuo-perceptual processing mechanisms, or both? Second, do eye-movement measures provide unique information about the processing of numerical magnitudes that relates to individual differences in math competence, beyond that accounted for by task performance (i.e. error rate and reaction time).

In terms of eye-movements, we investigated mean fixation count (FC) and fixation dwell time (FD) at the trial-level. We further analyzed the spatial distribution of eye-movement patterns by investigating the proportion of first fixations on the larger (i.e. correct) numerosity, as well as FC and FD on the larger and smaller stimulus considered independently. Consistent with a large body of previous literature we observed significant effects of ratio on error rate and RT, with a larger ratio effect for nonsymbolic comparison in both cases. In line with two previous studies that investigated nonsymbolic and symbolic comparison independently (Merkley & Ansari, 2010; Odic & Halberda, 2015), we observed significant ratio effects for FC and FD in both nonsymbolic and symbolic comparisons. The current analyses of the spatial distribution of eye movements also fit with previous results showing a ratio effect for proportion of first fixations on the large number, but further, that ratio effects were only apparent for FC and FD on the smaller number. As such, our tasks appear to elicit eye-movement and behavioral response profiles consistent with previous literature.

The current study is, however, the first to directly investigate differences and similarities in eye-movement patterns between nonsymbolic and symbolic formats. Performance measures and total mean measures of FC and FD were correlated across formats, but spatial measures of eye movements (i.e. proportion of first fixation, FC, and FD on the larger number) were not. The fact the spatial distribution of eye-movement measures was not related across formats is unsurprising given they are thought to reflect the effects of bottom-up visual processes, and the two formats were entirely different in terms of their visual representations. Further, nonsymbolic comparison elicited a higher proportion of first fixations on the larger (correct) stimulus than symbolic comparison despite having a comparable mean number of fixations. Our results also revealed that nonsymbolic comparison was associated with longer duration of fixations and a greater ratio effect on both FD and FC. These general format differences likely reflect the fact that the nonsymbolic task was subjectively more difficult than the symbolic task for the current sample. Error rate was largely unrelated to the eye-movement measures in both formats, with the exception of proportions of FC and FD on the larger numerosity for nonsymbolic comparison. Taken together, these results suggest that some aspects of visuo-perceptual processes underlying numerical magnitude comparison are shared across formats, and are associated with the speed, but not the accuracy of solutions. How accurately an individual compares numerical magnitudes, however, does not appear to be related to eye-movement patterns, suggesting that eye-movement patterns may be capturing important visuo-perceptual mechanisms involved in magnitude comparison that are not reflected in performance accuracy. While these explanations are inevitably speculative and require further empirical investigation, the results do suggest a divergence between formats in the influence of eye-movement patterns on task perfor-

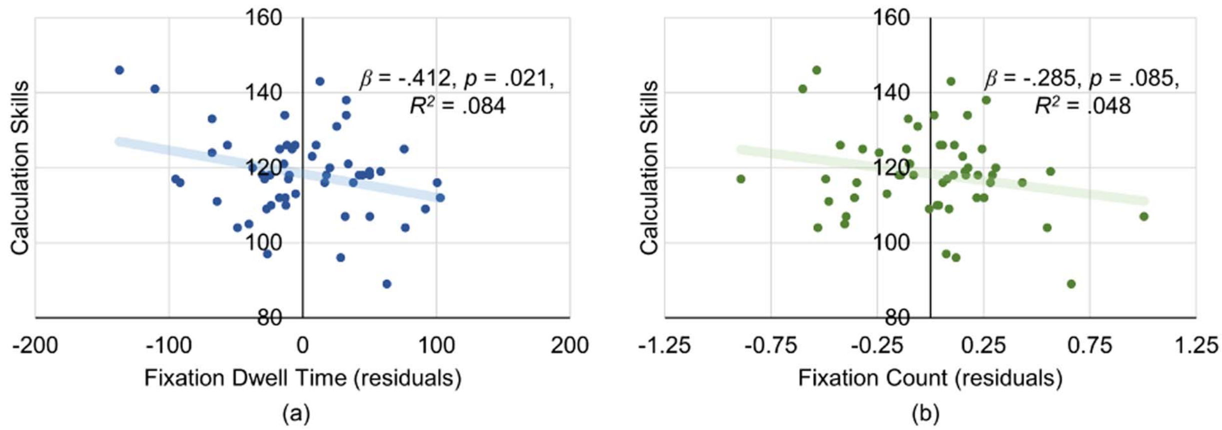

Fig. 6. After controlling for IQ, in addition to error rate and RT, (a) shorter fixations, during symbolic comparison predicted higher calculation skills. (b) Fixation count during symbolic comparison, however, was no longer predictive of calculation skills.

mance. Further divergence was evident in the relation between eye-movement and performance measure ratio effects.

The present study was also the first to investigate the relation between eye-movement patterns during numerical magnitude comparison and math performance. Our results demonstrate that eye-movement variables were related to math calculation skills even after controlling for task performance measures, but only for symbolic comparison. In fact, the only variables that were associated with math calculation skills were the eye-tracking measures of mean FC and FD during symbolic comparison. Because these relationships were negative, we interpret this to mean that fewer fixation and shorter durations were associated with more efficient perception of visual symbols. The correlation between symbolic error rate and calculation skills was no longer significant in the current study after correcting for multiple comparisons, but the effect size was in line with Schneider et al.'s (2016) meta-analysis demonstrating a small but consistent, significant relationship. Error rate, we assume, reflects the combined action of a number of cognitive and perceptual mechanisms, including visual processing, semantic processing, and decision making. Eye-movement patterns, on the other hand, we assume primarily reflect the efficiency of fluency of visual perception, although some influence of cognitive processing on eye-movements is likely. Therefore, to examine the extent to which eye-movement patterns during symbolic comparison, as a proxy for visual processing fluency of Arabic digits, were related to math performance over and above the range of cognitive processes reflected in task performance, we carried out a hierarchical regression predicting calculation skills from FC and FD after first entering IQ, error rate, and RT. Error rate and RT during symbolic comparison did not account for unique variance in calculation skills over and above IQ. Total FD, on the other hand, accounted for unique variance in calculation skills over and above IQ and performance measures, but total FC did not. It is interesting to note that FD was a unique predictor of calculation skills, despite the fact that FD, FC, and RT were highly correlated. We suggest that despite the high correlations among these variables, some distinction between them exists, indicated by the absence of perfect correlations. We propose that FC largely reflects the visuo-perceptual comparison process, and is possibly related to the difficulty of the trials and/or format – some require more fixations and saccades than others. Hence, it should be related to the number of saccades. As mentioned in the Methods section, FC and saccades were highly correlated ($rs > 0.91$ in both formats), which prompted us to exclude number of saccades from further analyses. FD, on the other hand, may reflect the amount of time spent extracting the semantic content (i.e., the numerosity) of the stimuli. Although FC and FD correlated in the present study, such a relation is not inevitably true. For instance, one may spend more time looking at each stimulus to extract the semantics, but only looked at each once, or they may spend less overall time looking, but make more comparisons between the two numerosities. Finally, RT is thought to reflect a combination of perceptual (FC), post visuo-perceptual semantic (FD) processes, and decision-making processes. Such delineation of the possible shared and distinct processes underlying each measure is congruent with the findings of our hierarchical regression analyses. Taken together, these results suggest that fluent visual processing of Arabic digits, over and above semantic and decision making related processes, plays an important role in linking symbolic number processing to formal math abilities.

In the current study, none of the ratio effects, whether symbolic or nonsymbolic, eye-movement or task-performance related, correlated with math competence. This result is part of a large body of findings that challenge previous assumptions about the extent to which the ratio effect indicates individual differences in the representational acuity of number systems, and the relevance of the ratio effect metric to math competence. For example, Lyons et al. (2015) published a detailed analysis of cross-format ratio effect comparisons and found that ratio effect did not correlate across formats or relate to child math competency in a large sample ($N = 1719$) across all primary school grades. Further, Schneider et al.'s (2016) meta-analysis of symbolic and nonsymbolic comparison tasks found that overall task performance, and not ratio effects, correlated with math across a wide range of studies ($k = 284$).

The fact that symbolic but not nonsymbolic comparison were associated with math skills in the current study is consistent with an emerging body of literature (De Smedt et al., 2013), and the current study adds to that literature by showing that the pattern extends to measures of eye-movement patterns. It is important to note, however, that previous research suggests that relation between nonsymbolic magnitude comparison and Math Fluency may be stronger in children than adults (Inglis, Attridge, Batchelor, & Gilmore, 2011; Schneider et al., 2016), and therefore the absence of a relation in the present study may in part be related to the fact that our sample included only adult participants. Therefore, the current study ought to be replicated with participants of varying ages to better understand the respective relations of nonsymbolic and symbolic processing to math skills. A second issue that we suggest be addressed in future research is the discrepancy in task difficulty between nonsymbolic and symbolic comparison in the present study. This could be, in part, due to the size effect. That is, the nonsymbolic task included slightly larger numbers overall (6–15) than the symbolic task (1–9) in order to avoid stimuli in the subitizing range. The nonsymbolic task was significantly more difficult overall and elicited stronger ratio effects than the symbolic task. Similarly, future studies should address differences in stimulus properties across formats not controlled for in the current study where possible. For example, dot arrays extended across a wider area than symbolic stimuli. Adjusting the task parameters, such as ratio range and visual degree, to achieve equivalent difficulty and more visual similarities across formats may reveal associations across formats not observed in the present study.

In summary, this is the first study to compare eye-movement patterns during nonsymbolic and symbolic magnitude comparison, and to relate those eye-movement patterns to math performance. Our results reveal a pattern of associations between eye-movement measures, but not their ratio effects, across formats. This suggests that ratio-specific visuo-perceptual processing during magnitude processing is different across nonsymbolic and symbolic formats. Furthermore, eye movements are related to math performance only during symbolic comparison, supporting a growing body of literature suggesting symbolic number processing is more strongly related to math outcomes than nonsymbolic magnitude processing. Finally, eye-movement patterns, specifically fixation dwell time, continues to be related to math performance after controlling for error rate, RT, and IQ, suggesting a unique role for fluent visual recognition of Arabic digits in the development of math competence. Taken together, the present results provide novel insights into the mechanisms underlying numerical magnitude processing across formats, as well as the relation between magnitude processing and math competence.

References

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20141791 http://doi.org/10.1098/rspb.2014.1791.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* http://doi.org/10.2307/2346101.

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*(1), 32–44 http://doi.org/10.1016/j.cognition.2010.09.005.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5), 401–406 http://doi.org/10.1111/j.1467-9280.2006.01719.x.

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology*, *5*(12), e328 http://doi.org/10.1371/journal.pbio.0050328.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenisis. *Mind & Language*, *16*(1), 37–55 http://doi.org/10.1111/1468-0017.00155.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*, 12536 http://doi.org/10.1038/ncomms12536.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology*, *84*(2), 132–147 http://doi.org/10.1016/j.pneurobio.2007.11.001.

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 http://doi.org/10.1016/j.tine.2013.06.001.

Dehaene, S. (2003). The neural basis of the c law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147 http://doi.org/10.1016/S1364-6613(03)00055-X.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.

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 http://doi.org/10.1016/j.jecp.2014.01.013.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and younger adults' strategies in approximate quantification. *Acta Psychologica*, *129*(1), 175–189 http://doi.org/10.1016/j.actpsy.2008.05.009.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*(4), 981–986 http://doi.org/10.3758/s13428-011-0097-5.

Godau, C., Wirth, M., Hansen, S., Haider, H., & Gaschler, R. (2014). From marbles to numbers—Estimation influences looking patterns on arithmetic problems. *Psychology*, *5*(2), 127–133 http://doi.org/10.4236/psych.2014.52020.

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668.

Henderson, J. M. (2013). Eye movements. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*. New York, NY: Oxford University Press.

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 http://doi.org/10.1016/j.jecp.2008.04.001.

Huber, S., Mann, A., Nuerk, H. C., & Moeller, K. (2014). Cognitive control in number magnitude processing: evidence from eye-tracking. *Psychological Research*, *78*(4), 539–548.

Hurst, M., & Cordes, S. (2016). Rational-number comparison across notation: Fractions, decimals, and whole numbers. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 281–293 http://doi.org/10.1037/xhp0000140.

Hutzler, F., Kronbichler, M., Jacobs, A. M., & Wimmer, H. (2006). Perhaps correlational but not causal: No effect of dyslexic readers' magnocellular system on their eye movements during reading. *Neuropsychologia*, *44*(4), 637–648 http://doi.org/10.1016/j.neuropsychologia.2005.06.006.

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 http://doi.org/10.3758/s13423-011-0154-1.

Kaufman, A., & Kaufman, N. (2004). *Kaufman brief intelligence test-2*. London: Pearson.

Kirkby, J. A., Blythe, H. I., Drieghe, D., & Liversedge, S. P. (2011). Reading text increases binocular disparity in dyslexic children. *PLoS ONE*, *6*(11), 1–7 http://doi.org/10.1371/journal.pone.0027105.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261 http://doi.org/10.1016/j.cognition.2011.07.009.

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology. General*, *144*(5), 1021–1035 http://doi.org/10.1037/xge0000094.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), e23749 http://doi.org/10.1371/journal.pone.0023749.

Merkley, R., & Ansari, D. (2010). Using eye tracking to study numerical cognition: The case of the ratio effect. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *206*(4), 455–460 http://doi.org/10.1007/s00221-010-2419-8.

Mock, J., Mock, J., & Huber, S. (2016). Insights into numerical cognition - Considering eye-fixations in number processing and arithmetic. *Psychological Research* http://doi.org/10.1007/s00426-015-0739-9.

Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K., & Nuerk, H. C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive Development*, *24*(4), 371–386. http://dx.doi.org/10.1016/j.cogdev.2009.09.007.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(109), 1519–1520.

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 http://doi.org/10.1016/j.jecp.2009.02.003.

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2015). How approximate and exact number skills are related to each other across development: A review. *Developmental Review*, 1–15 http://doi.org/10.1016/j.dr.2014.11.001.

Odic, D., & Halberda, J. (2015). Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *15*(2015), 1–15 http://doi.org/10.1167/15.15.5.doi.

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science* (August). http://doi.org/10.1177/0956797613482944.

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, *24*(6), 1037–1043 http://doi.org/10.1177/0956797612464057.

Price, G. R., & Fuchs, L. S. (2016). The mediating relation between symbolic and nonsymbolic foundations of math competence. *PLoS ONE*, *11*(2), e0148981 http://doi.org/10.1371/journal.pone.0148981.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 1–16 http://doi.org/10.1111/desc.12372.

Sophian, C., & Crosby, M. E. (2008). What eye fixation patterns tell us about subitizing. *Developmental Neuropsychology*, *33*(March 2015), 394–409 http://doi.org/10.1080/87565640801982460.

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(1), 19 http://doi.org/10.1186/1744-9081-2-19.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, 1–11.