Journal of Experimental Child Psychology 116 (2013) 216–233

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

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

Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments

Emmy Defever [a,b,*], Bert Reynvoet [a,b], Titia Gebuis [a]

[a] Laboratory of Experimental Psychology, University of Leuven, 3000 Leuven, Belgium
[b] Subfaculty of Psychology and Educational Sciences, University of Leuven, Kulak, 8500 Kortrijk, Belgium

ARTICLE INFO

Article history:
Received 1 November 2012
Revised 21 April 2013
Available online 13 July 2013

Keywords:
Visual properties
Numerosity judgment
Comparison
Same–different
Mathematics achievement

ABSTRACT

Researchers investigating numerosity processing manipulate the visual stimulus properties (e.g., surface). This is done to control for the confound between numerosity and its visual properties and should allow the examination of pure number processes. Nevertheless, several studies have shown that, despite different visual controls, visual cues remained to exert their influence on numerosity judgments. This study, therefore, investigated whether the impact of the visual stimulus manipulations on numerosity judgments is dependent on the task at hand (comparison task vs. same–different task) and whether this impact changes throughout development. In addition, we examined whether the influence of visual stimulus manipulations on numerosity judgments plays a role in the relation between performance on numerosity tasks and mathematics achievement. Our findings confirmed that the visual stimulus manipulations affect numerosity judgments; more important, we found that these influences changed with increasing age and differed between the comparison and the same–different tasks. Consequently, direct comparisons between numerosity studies using different tasks and age groups are difficult. No meaningful relationship between the performance on the comparison and same–different tasks and mathematics achievement was found in typically developing children, nor did we find consistent differences between children with and without mathematical learning disability (MLD).

---

[*] Corresponding author at: Subfaculty of Psychology and Educational Sciences, University of Leuven, Kulak, 8500 Kortrijk, Belgium. Fax: +32 56 24 60 52.

E-mail address: emmy.defever@kuleuven-kulak.be (E. Defever).

0022-0965/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jecp.2013.04.006

Introduction

Numerosity and its continuous visual properties are correlated in everyday life. For instance, when more marbles are added to a collection of marbles, numerosity as well as the size of the collection increases. Therefore, it is common practice that studies investigating numerosity processing manipulate the visual cues of the numerosity stimuli. These manipulations should prevent participants from relying on the visual cues when judging numerosity and, thus, allow researchers to study pure number processes. However, an increasing number of studies show that numerosity judgments are sensitive to the continuous visual properties of the numerosity stimuli. These studies show, for instance, that numerosity judgments are influenced by density (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Sophian & Chu, 2008; Tibber, Greenwood, & Dakin, 2012), the size of the individual elements (Gebuis & Reynvoet, 2011b; Hurewitz, Gelman, & Schnitzer, 2006; Rousselle & Noël, 2008; Tokita & Ishiguchi, 2010), or patch size (Gebuis & Gevers, 2011; Tokita & Ishiguchi, 2010).

Only a few researchers have investigated whether and how the visual stimulus manipulations they use in their experiments affect the measurements of numerosity processing. Researchers mostly believe that the visual confounds are sufficiently controlled and collapse the data from congruent trials (e.g., the larger numerosity has larger visual cues) and incongruent trials (e.g., the smaller numerosity has larger visual cues) without further examining whether the visual stimulus manipulations in these conditions lead to different findings (e.g., De Smedt & Gilmore, 2011; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). However, investigating the effect of the different visual manipulations in congruent and incongruent trials separately seems to be necessary because some studies have shown that participants respond to the visual properties even when controlling for the visual cues (Gilmore, Attridge, & Inglis, 2011; Inglis, Attridge, Batchelor, & Gilmore, 2011). For instance, approximately 30% of the 8-year-olds in a study by Inglis and colleagues (2011) and approximately 40% of the adults in a study by Gilmore and colleagues (2011) were removed. The performance of these participants in the congruent and incongruent conditions differed by more than 50% in accuracy, suggesting that they relied on the visual stimulus properties. Apparently, the performance on numerosity tasks is unavoidably affected by the manipulations of the continuous visual properties.

To allow comparisons across numerosity studies using different designs and including different age groups, it is essential to find out how the visual stimulus manipulations affect numerosity performance throughout development. It is not clear whether children of different ages are influenced differently by the visual stimulus properties. Infant studies have revealed inconsistent results about the effect of the visual stimulus properties on numerosity judgments. There are studies showing that infants might be more sensitive to the continuous visual properties at the expense of numerosity (Clearfield & Mix, 1999, 2001; Feigenson, Carey, & Spelke, 2002), that infants are equally sensitive to numerosity and visual cues (Brannon, Lutz, & Cordes, 2006; Cordes & Brannon, 2009; vanMarle & Wynn, 2006), and that infants prefer to attend to numerosity over visual properties (Brannon, Abbott, & Lutz, 2004; Cordes & Brannon, 2008, 2011; Xu, 2003; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). The few studies using explicit paradigms, which examined the effects of the visual stimulus manipulations, showed a more consistent pattern of findings. These studies point to a reliance on visual cues in young children at the expense of number, which seems to decrease with increasing age (Rousselle & Noël, 2008; Rousselle, Palmers, & Noël, 2004) but remains visible even in adults (Gebuis & van der Smagt, 2011; Gilmore et al., 2011; Halberda, Mazzocco, & Feigenson, 2008). However, the age at which children could compare numerosities while the visual cues were manipulated differed. Rousselle and colleagues (Rousselle & Noël, 2008; Rousselle et al., 2004) showed that 3-year-old children were unable to compare numerosities when the stimuli were controlled for surface area, whereas 4- and 5-year-olds performed significantly above chance. In contrast, Soltész, Szucs, and Szucs (2010) observed that 4-year-olds performed at chance level when the visual properties and numerosity were manipulated inconsistently, whereas performance was above chance from 5 years onward.

Research on the impact of the visual stimulus manipulations on numerosity judgments is necessary when considering the important conclusions that are drawn from the results of numerosity studies. It is, for instance, suggested that performance on numerosity tasks is related to mathematics achievement.

Typically developing children who performed worse on numerosity discrimination showed less proficiency in mathematics than children who performed better on this task (e.g., Halberda et al., 2008; Inglis et al., 2011; Mazzocco, Feigenson, & Halberda, 2011b). Similarly, children with mathematical learning disability (MLD) performed worse than controls on numerosity comparison (Mazzocco, Feigenson, & Halberda, 2011a; Mussolin, Mejias, & Noël, 2010; Piazza et al., 2010). However, there are also several studies that show no correlation between the performance on numerosity tasks and mathematics achievement in typically developing children (e.g., Holloway & Ansari, 2009; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Soltész et al., 2010) or that show no difference between children with and without MLD on numerosity discrimination ability (De Smedt & Gilmore, 2011; Iuculano, Tang, Hall, & Butterworth, 2008; Rousselle & Noël, 2007).

The idea has been proposed that the influence of the visual stimulus properties could explain the mixed findings regarding the relationship between numerosity processing and mathematics achievement (Gebuis & Reynvoet, in press). To date, only a few studies investigating the association between numerosity discrimination and mathematics achievement have taken the influence of the visual stimulus properties into account. Mussolin and colleagues (2010) suggested that children with MLD were more sensitive than controls to continuous visual properties, such as density and surface area, during comparison of random stick patterns. Similarly, Mejias, Mussolin, Rousselle, Grégoire, and Noël (2012) showed that, in contrast to control children, children with MLD relied more on the visual cues that correlated with numerosity because they made better estimations in a task with homogeneous-sized dots (where the cumulative area covaried with numerosity) than in a task with heterogeneous-sized dots (where the cumulative area was constant across numerosities). These studies suggest that children with MLD are more influenced by the visual stimulus properties than control participants.

Our goals in this study were twofold. First, we investigated whether the visual cues influence numerosity judgments differently across development and across task designs. Insight into these variables will reveal whether data from different studies can be compared or whether caution is warranted. Second, we investigated whether the visual cues could explain the observed discrepancies between numerosity studies regarding the relationship between numerosity processing and mathematics achievement. To this end, we conducted an experiment in a group of typically developing children from first, second, third, and sixth grades (Experiment 1) and in children with MLD (Experiment 2). The performance of both typically developing children and children with MLD was investigated because difficulties in typically developing children with low math proficiency are not necessarily the same as the difficulties in children with MLD (e.g., Desoete, Ceulemans, De Weerdt, & Pieters, 2012; Mazzocco, Devlin, & McKenney, 2008). We administered both a comparison task and a same–different task in which we manipulated the visual properties of the numerosity stimuli using a more stringent method to control the visual cues than the method used in most previous studies (Gebuis & Reynvoet, 2011a). Similar to previous studies, in the comparison task, in half of the trials the numerically larger numerosity consisted of a larger surface, density, and diameter but a smaller convex hull (congruent condition), and in the other half the numerically larger numerosity consisted of a smaller surface, density, and diameter but a larger convex hull (incongruent condition) (for a similar design, see Condition 4 in Gebuis & Reynvoet, 2011a). The visual cues of the trials with two numerically different stimuli in the same–different task were manipulated in the same way as in the comparison task. The numerically same trials also consisted of a stimulus with larger visual cues and a stimulus with smaller visual cues. Differences in the reliance on the visual cues in a comparison task and a same–different task can be expected. Namely, in the comparison task, where participants need to indicate the numerically larger numerosity, participants might be prone to indicate the stimulus consisting of larger visual properties as numerically larger. In contrast, in the same–different task, where participants need to decide whether the numerosities are numerically the same or different, participants might be prone to judge the numerosities as different because the visual properties of the numerosities always differ (also in the numerically same trials, one of the two stimuli was always larger in visual stimulus properties). To examine the relationship between the performance on the tasks and mathematics achievement in Experiment 1, we administered both a general mathematics achievement test and a timed arithmetic test.

Experiment 1

Method

Participants

Participants were 26 first graders, 34 second graders, 34 third graders, and 30 sixth graders recruited from two elementary schools in Belgium. Among this sample, 6 participants were excluded from the analyses because they were too slow or made too many errors (>3 standard deviations above group average) or because their math achievement score was missing. This resulted in a final sample of 25 first graders (16 boys and 9 girls, mean age = 6.67 years, *SD* = 0.29), 32 second graders (20 boys and 12 girls, mean age = 7.19 years, *SD* = 0.41), 33 third graders (17 boys and 16 girls, mean age = 8.61 years, *SD* = 0.46), and 28 sixth graders (10 boys and 18 girls, mean age = 11.26 years, *SD* = 0.48). Written informed consent was obtained from all of the parents, and the study was approved by the ethics committee of the University of Leuven.

Apparatus, stimuli, and procedure

Experimental tasks. Stimulus presentation and recording of the data were controlled by E-Prime 1.1 (Psychology Software Tools, http://www.pstnet.com). Participants were presented with two arrays of gray dots on a black background. A vertical gray line separated the dot arrays. The stimuli of both experiments were generated using the program developed by (Gebuis & Reynvoet, 2011a). Four visual properties were manipulated: (a) the convex hull (i.e., smallest contour around the array of dots), (b) the aggregate surface of the dots, (c) the density (i.e., aggregate surface divided by convex hull), and (d) the average diameter. Regression analyses confirmed that there was no relationship between the difference in visual properties and the difference in numerosity across trials for the comparison task (all R^2s < .07, all *p*s > .001) and the same–different task (all R^2s < .07, all *p*s > .001). For the comparison task, in half of the trials, the more numerous array had a larger surface, density, and diameter but a smaller convex hull (hereafter referred to as *congruent trials*) and this was vice versa for the other half of the trials (hereafter referred to as *incongruent trials*) (these visual cue combinations resulted in the strongest congruency effects in a previous study; see Condition 2 in Gebuis & Reynvoet, 2011b). This way, participants could not rely on a single visual cue to decide which dot array is numerically larger. The same manipulation was applied to the "different trials" in the same–different task. For the "same trials" in the same–different task, half of the trials had the stimulus with the larger surface, density, and diameter and smaller convex hull on the left side of the screen, whereas in the other half it was presented on the right side of the screen. Consequently, participants could not decide that two stimuli were numerically different simply because the visual stimulus properties differed.

Comparison task. One of the two arrays always contained 16 dots, whereas the other display contained a smaller number (8, 11, 12, 13, or 14) or larger number (18, 19, 21, 24, or 32) of dots, resulting in 10 different number pairs. These 10 different number pairs can be divided over 5 different ratio conditions (2.0, 1.5, 1.3, 1.2, and 1.12). The 10 different number pairs were presented separately in blocks, each consisting of 16 trials (total of 160 trials, $10*16$) (see Gebuis & Van der Smagt, 2011, for an identical approach). The different blocks were fully randomized between participants, as were the 16 trials within each block.

In half of the trials the larger number of dots was presented on the left side of the screen, whereas in the other half it was presented on the right side. Participants were asked to indicate which of the two arrays contained more dots by pressing the left ("a") or right ("p") button on an AZERTY keyboard. To make children familiar with the task, 5 practice trials were given, during which feedback was provided, before the experiment started. Each trial started with a fixation cross (800 ms) followed by the stimuli (until a response was given), after which the intertrial interval (1000 ms) started. Children were seated approximately 50 cm from the screen. In total, the experiment took approximately 10 min.

Same–different task. One of the two arrays always contained 16 dots, whereas the other display contained an equal number (16), a smaller number (6, 7, 8, 11, or 12), or a larger number (21, 24, 32, 37, or 40) of dots. This resulted in 10 different number pairs with two numerically different stimuli (hereafter referred to as *different trials*) and 1 number pair with two numerically same stimuli (hereafter referred to as *same trials*). The 10 different number pairs were presented separately in blocks of 8 trials together with 8 same trials. In total, the experiment consisted of 160 trials, that is, 32 trials per ratio condition (2.5, 2.3, 2.0, 1.5, and 1.3).[1] By assigning the same trials to a specific ratio condition, we could include them in the analyses (for a similar approach, see Gebuis & Van der Smagt, 2011). This enabled us to account for a possible response bias. Children might show a bias toward responding "different" because the numerosity stimuli were always visually different. Analyzing only the different trials, therefore, would artificially lead to better performance. By including the same trials in the analyses, this problem can be circumvented (see Gebuis & Van der Smagt, 2011).

Participants were asked to indicate whether the two arrays contained an equal number or a different number of dots by pressing "a" (labeled with "=") or "p" (labeled with "≠") on an AZERTY keyboard. All other task characteristics were identical to those in the comparison task.

Standardized mathematics achievement tests. General mathematics achievement. We assessed children's general mathematics achievement by means of curriculum-based standardized achievement tests from the Flemish Student Monitoring System (Deloof, 2005; Dudal, 2000, 2001, 2002). This student monitoring system provides every grade, apart from first and sixth grades, with three mathematics tests: one to administer at the beginning, the middle, and the end of the school year. Only two tests are provided for first grade (one to administer at the middle and end of the school year) and sixth grade (one to administer at the beginning and middle of the school year). In this study, the scores from the test administered at the middle of the school year were used.

Each test comprises 60 items covering number knowledge, understanding of operations, (simple) arithmetic, word problem solving, measurement, and geometry. The content of the test items is adapted to the mathematics curriculum of the specific grade. Thus, the same constructs are included in the tests for each grade but have a different difficulty level. To give an example, the test in third grade administered at the middle of the school year includes 20 number knowledge items (e.g., How many numbers are there between 91 and 94?), 15 items measuring the understanding of operations (e.g., 90 less than 400 is ...), 15 items measuring (simple) arithmetic (e.g., 80 + 209), 5 word problem solving items (e.g., Aunt Rose is this year exactly twice as old as her son Bruno. Bruno is 28 years old. How old is Aunt Rose?), 10 items on measurement (e.g., Our postman is neither fat nor skinny, tall nor short. How much could he weigh? 25 kg – 40 kg – 75 kg – 110 kg – 125 kg), and 5 items in the domain of geometry (e.g., A door has the shape of a: square – triangle – circle – rectangle – hexagon). The mathematics tests from the Flemish Student Monitoring System were developed to assess children's overall mathematics competence. Hence, several items in the tests measure more than one domain. For example, the word problem solving items also assess children's understanding of operations. The reliability indexes of Kuder–Richardson (KR 20) for this test were .90, .89, .90, and .88 for first, second, third, and sixth grades, respectively. For the analyses, we transformed the raw mathematics achievement scores to *Z*-scores per grade.

Timed arithmetic test. The second math achievement test was a speeded test for mental calculations (Tempo Test Arithmetic; de Vos, 1992). It consists of 5 subtests: one for each type of operation (addition, subtraction, multiplication, and division) and one with mixed operations. A total of 40 items of increasing difficulty are presented in each subtest, and participants have 1 min to solve as many problems as possible. The score on this subtest is the number of correctly solved items. Children in first grade conducted only the first two subtests because they had not learned how to multiply and divide yet. For the analyses, we transformed the raw mathematics achievement scores to *Z*-scores per grade.

General procedure

Children were tested in groups of approximately 5 to 7. They were seated in such a way that they could not distract each other and the experimenter could monitor them closely. The order of the experimental tasks was counterbalanced within each grade; half of the children started with the

---

[1] Performance on the same–different task is usually worse compared with the comparison task (Gebuis & Van der Smagt, 2011; Piazza et al., 2004). Therefore, two larger ratios were included in the same–different task and the two smallest ratios used in the comparison task were removed (ratios 2.3 and 2.5 instead of ratios 1.12 and 1.2).

same–different task, whereas the other half started with the comparison task. The mathematics achievement tests were administered collectively within each classroom.

Data analyses

Comparison task. The analysis was performed on the percentage of correct trials as a function of congruency and ratio condition. To investigate the visual congruency effects across age groups, we conducted a repeated measures analysis of variance (ANOVA) with congruency (congruent or incongruent) and ratio (2.0, 1.5, 1.3, 1.2, or 1.12) as within-participant factors. In addition, to examine the development of the congruency effects and its relationship with mathematics achievement, we included grade as a between-participant factor and mathematics achievement as a covariate. We conducted the ANOVA twice: once with the scores on the untimed curriculum-based math achievement test as a covariate and once with the scores on the timed fluency test. Post hoc analyses were conducted to further examine the visual congruency effects across age groups in more detail.

Same–different task. To investigate whether performance was influenced by the visual cues, we analyzed the response bias by comparing the numbers of different and same responses. We first examined whether the difference in the number of same and different responses differed significantly from zero. If children relied on the visual cues, we expected a significant bias toward responding “different” because the visual properties of the to-be-compared numerosities were always different (also in the numerically same trials). To examine whether the response bias differed as a function of grade and ratio, we conducted a repeated measures analysis on the difference in percentage between the numbers of different and same responses with ratio as a within-participant factor, grade as a between-participant factor, and mathematics achievement as a covariate.

Because the same–different task consisted of not only numerically different trials but also numerically same trials, we could not analyze the data as a function of number–visual congruency as was done in the comparison task. Therefore, to investigate the ratio effect across age groups, we conducted a second repeated measures ANOVA with ratio (2.5, 2.3, 2.0, 1.5, or 1.3) as a within-participant factor, grade as a between-participant factor, and mathematics achievement as a covariate. Similar to the comparison task, the analyses were run twice: once with the scores on the untimed curriculum-based math achievement test as a covariate and once with the scores on the timed fluency test.

Results

Comparison task

Analysis of visual congruency effect. The repeated measures ANOVA revealed a main effect of ratio, $F(4, 110) = 160.303$, $p < .0001$, $\eta_p^2 = .854$, indicating that children’s accuracy increased with increasing ratio (see Fig. 1A). Pairwise comparisons showed a significant difference between all levels of ratio (all

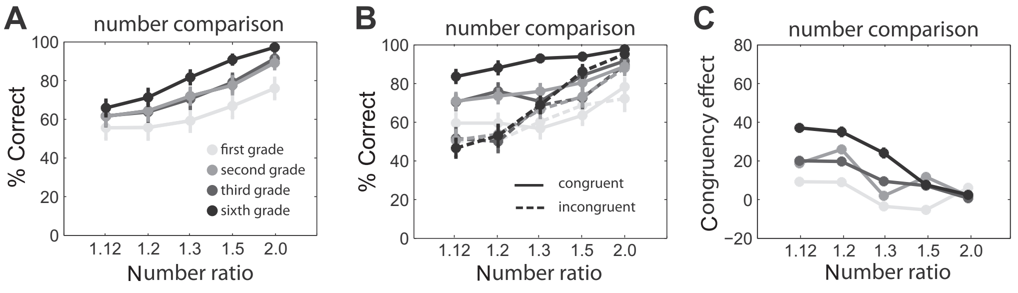

Fig. 1. Children’s mean accuracies on the comparison task in Experiment 1. (A) Mean accuracies as a function of ratio and grade. (B) Mean accuracies as a function of ratio, grade, and congruency. (C) Visual congruency effect (accuracy on congruent trials – mean accuracy on incongruent trials).

*p*s < .050, Cohen's *d*s ranging from 0.25 to 2.29). A significant main effect of congruency was also observed, $F(1, 113) = 17.931$, $p < .0001$, $\eta_p^2 = .137$, showing a higher accuracy for congruent trials (78%) compared with incongruent trials (67%), $p < .0001$, Cohen's $d = 0.68$ (see Fig. 1B). The significant main effect of grade, $F(3, 113) = 20.765$, $p < .0001$, $\eta_p^2 = .355$, indicated that accuracy increased with increasing grade (all *p*s < .0001, Cohen's *d*s > 1.02), with the exception that no significant difference was found between the second and third graders ($p = .867$). The effect of mathematics achievement was also significant, $F(1, 113) = 8.759$, $p < .010$, $\eta_p^2 = .072$. Partial correlations controlling for grade showed that the mean accuracy increased with increasing mathematics achievement, $r_{\text{partial}}(115) = .25$, $p = .007$. However, the explained variance is negligible ($R^2 = .06$), especially in the light of our large number of participants ($N = 118$). Finally, a significant two-way interaction was present between ratio and grade, $F(12, 291) = 2.473$, $p < .010$, $\eta_p^2 = .082$, and between congruency and ratio, $F(4, 110) = 13.329$, $p < .0001$, $\eta_p^2 = .326$, which were embedded in a three-way interaction among congruency, ratio, and grade, $F(12, 291) = 2.649$, $p < .010$, $\eta_p^2 = .087$. As depicted in Fig. 1C, the visual congruency effect increased with increasing grade and decreasing ratio. For the first graders, no visual congruency effect was present for the different ratios (all *p*s > .121). The second graders showed a visual congruency effect for ratios 1.12, 1.2, and 1.5 (all *p*s < .050, Cohen's *d*s > 0.56) but not for ratios 1.3 and 2.0 (both *p*s > .568). The third graders showed a visual congruency effect for ratios 1.12 and 1.2 (both *p*s < .05, Cohen's *d*s > 0.79) but not for the three easier ratios (all *p*s > .130). Finally, the sixth graders showed a significant visual congruency effect for ratios 1.12, 1.2, and 1.3 (all *p*s < .001, Cohen's *d*s > 1.61) but not for ratios 1.5 and 2.0 (both *p*s > .213). No other significant effects were observed (all *p*s > .217).

The same analysis as mentioned above was also conducted with the scores on the timed mathematics achievement test (TTR) as a covariate (note that we administered two different tests to assess mathematical proficiency). This analysis revealed identical results as the analysis with the untimed curriculum-based math achievement test scores, with the exception that no significant main effect of mathematics achievement was found, $F(1, 113) = 1.567$, $p = .213$. Together for both mathematics tests, no clear relationship between numerosity performance and mathematics achievement was obtained.

Post hoc analyses of visual congruency effect. The presence of a larger visual congruency effect for the older children compared with the younger children in the comparison task was unexpected. On the basis of previous studies, we would expect a decrease of the visual congruency effect with increasing age (Rousselle & Noël, 2008; Soltész et al., 2010). However, our observation could be explained by the inconsistencies in the direction of these visual congruency effects. Some previous studies have shown that participants perform better on congruent trials than on incongruent trials and, thus, associated larger visual cues with a larger numerosity (Dakin et al., 2011; Hurewitz et al., 2006; Rousselle & Noël, 2008), whereas others have shown the opposite pattern (Gebuis & Van der Smagt, 2011; Ginsburg & Nicholls, 1988; Sophian, 2007). Apparently, differences between participants in how they associate number and visual cues can be observed. Closer inspection of our data showed relatively large negative and positive congruency effects (see Fig. 2). In previous studies, such large congruency effects were considered to be indicative of a reliance on the visual cues (Gilmore et al., 2011; Inglis et al., 2011). We observed that the visual congruency effects were in opposite directions. Some children indicated a numerosity as numerically larger when it was a visually congruent stimulus, whereas others indicated a numerosity as numerically larger when it was a visually incongruent stimulus. This shows that the association between numerosity and its visual properties differs among children. These visual congruency effects in different directions were clearly visible in the first graders, were visible to a lesser extent in the second and third graders, and were absent in the sixth graders, suggestive of a developmental change in the association between numerosity and its visual cues. Apparently, with increasing age, children are more inclined to associate a visually congruent stimulus with the larger numerosity. The opposite congruency effects can explain the initial observation of an increase of the visual congruency effect with increasing age. Indeed, individual congruency effects that are in opposite directions, which might have cancelled out or reduced the overall visual congruency effect, were present mostly for the younger children.

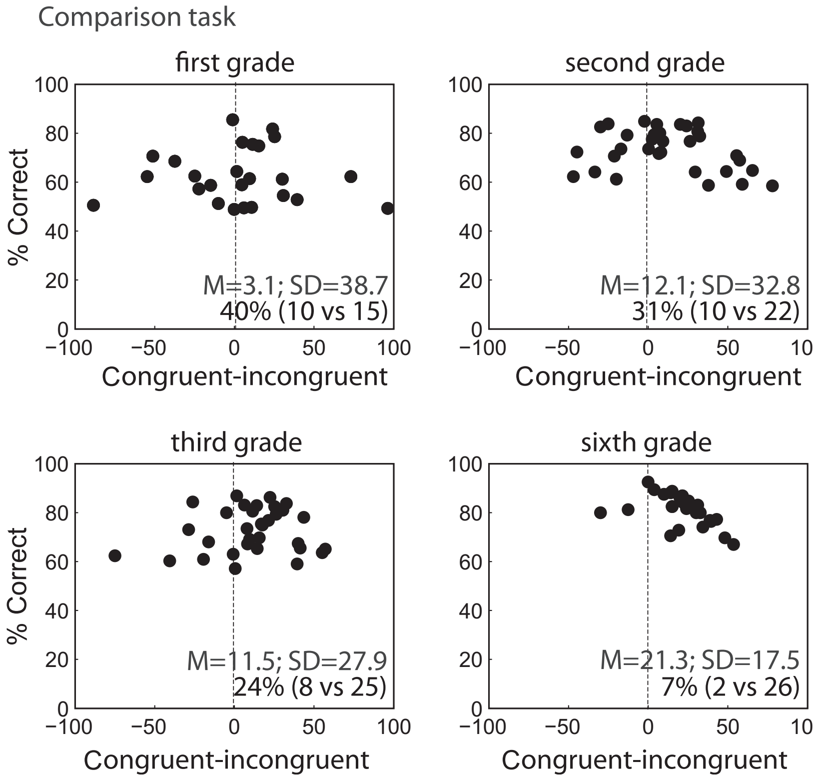

Fig. 2. Visual congruency effect in comparison task for each individual as a function of grade in Experiment 1. The panels show a developmental pattern in the reliance on the visual cues; the percentage of children showing a negative visual congruency effect decreased with increasing age. The visual congruency effects of the children in first grade were in opposite directions; this was less the case for the second and third graders. In sixth grade, nearly all children associated numerosity and visual cues in a similar manner; a larger surface, density, and diameter and a smaller convex hull were associated with the larger number.

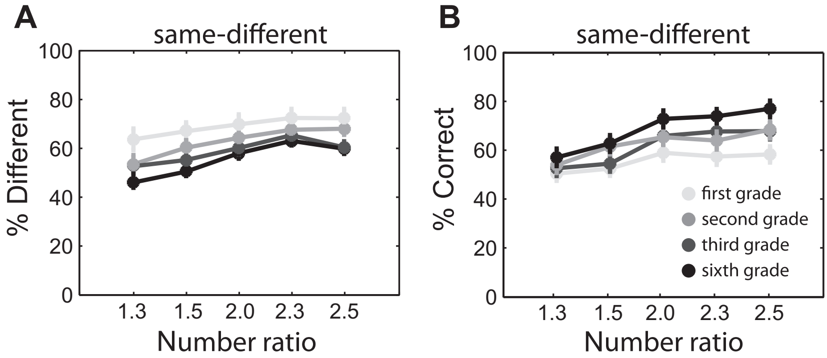

Fig. 3. Children's percentages of different responses and percentages correct on the same–different task in Experiment 1. (A) Percentages of different responses as a function of ratio and grade. (B) Mean accuracies as a function of ratio and grade.

Same–different task

Analysis of influence of visual cues. The difference between the numbers of different and same responses differed significantly from zero, $t(117) = 8.748$, $p < .0001$, Cohen's $d = 1.61$, showing that participants more often indicated a trial as being numerically different (61% of trials) than the same (39% of trials). The repeated measures ANOVA on the difference in numbers of different and same responses with ratio as a within-participant factor, grade as a between-participant factor, and mathematics achievement as a covariate showed a significant effect of grade, $F(3, 113) = 5.105$, $p < .010$, $\eta^2 = .119$. This bias toward a "different" response decreased with increasing age (the difference between the percentages different and same presses was 38%, 25%, 17%, and 11% for the first, second, third, and sixth graders, respectively) (see Fig. 3). The effect of ratio was also significant, $F(4, 110) = 32.841$, $p < .0001$, $\eta_p{}^2 = .544$, showing that the bias to indicate a trial as numerically different increased with increasing ratio (8%, 16%, 25%, 34%, and 29% for ratios 1.3, 1.5, 2.0, 2.3, and 2.5, respectively). The effect of mathematics achievement was not significant, $F(1, 113) = 2.936$, $p = .089$. A significant interaction between ratio and mathematics achievement was observed, $F(4, 110) = 2.777$, $p < .050$, $\eta_p{}^2 = .092$. Partial correlations controlling for grade showed a positive correlation between the bias toward pressing "different" (% different –% same responses) and mathematics achievement for ratio 1.3, $r_{\text{partial}}(115) = .19$, $p = .036$, and ratio 2.0, $r_{\text{partial}}(115) = .18$, $p = .048$, but not for ratio 1.5 ($p = .365$), ratio 2.3 ($p = .065$), or ratio 2.5 ($p = .724$). It appears that the relation between ratio and mathematics was not consistently present. Furthermore, the explained variance was again negligible (R^2s = .03–.04). Finally, the interaction between ratio and grade was not significant, $F(12, 291) = 1.076$, $p = .380$.

The same analysis was also conducted with the scores on the TTR as a covariate. This analysis revealed identical results, with the exception that no interaction between ratio and mathematics achievement was found ($F < 1$).

Analysis of ratio effect. The repeated measures ANOVA with ratio as a within-participant factor, grade as a between-participant factor, and mathematics achievement as a covariate revealed a significant main effect of ratio, $F(4, 110) = 33.229$, $p < .0001$, $\eta_p{}^2 = .547$, showing that performance improved with increasing ratio (see Fig. 3). Pairwise comparisons showed a significant difference between all levels of ratio (ps < .050, Cohen's ds ranging from 0.15 to 1.21) except between ratios 2.0 and 2.3 ($p = .990$) and between ratios 2.0 and 2.5 ($p = .063$). A main effect of grade was also observed, $F(3, 113) = 11.119$, $p = .0001$, $\eta_p{}^2 = .228$. Accuracy significantly increased with increasing grade (all ps < .010, Cohen's ds > 0.70) with the exception that no significant difference was observed between the second and third graders ($p = .661$). The interaction between ratio and grade was marginally significant, $F(12, 291) = 1.740$, $p = .058$, $\eta_p{}^2 = .059$; the ratio effect tended to be steeper for the older children compared with the younger children (see Fig. 3). We also obtained a significant effect of mathematics achievement, $F(1, 113) = 5.161$, $p < .050$, $\eta_p{}^2 = .044$. Partial correlations controlling for grade indicated that the mean accuracy increased with increasing mathematics achievement, $r_{\text{partial}}(115) = .21$, $p = .026$. However, the explained variance is again negligible ($R^2 = .04$). Finally, the interaction between ratio and mathematics achievement was not significant ($F < 1$).

The analysis with the scores on the TTR as a covariate showed identical results. The effect of mathematics achievement was also significant, $F(1, 113) = 6.785$, $p < .050$, $\eta_p{}^2 = .057$. Partial correlations controlling for grade indicated that the mean accuracy increased with increasing mathematics achievement, $r_{\text{partial}}(115) = .23$, $p = .011$. Similar to the analysis with the scores on the untimed mathematics achievement test as a covariate, the explained variance was again negligible ($R^2 = .05$).

Discussion

In line with expectations, the visual cues influenced the performance on both the comparison and same–different tasks. A bias toward pressing "different" was present in the same–different task, which decreased with increasing age. The visual congruency effect in the comparison task (i.e., better performance on congruent trials vs. incongruent trials) unexpectedly increased with increasing age. Closer inspection of the data from the comparison task suggested strong influences of visual cues in all age groups, but for the younger children these influences were in opposite directions. With increasing age, differences between participants in how they associate numerosity and its visual properties decreased. This developmental change can explain the increase of the visual congruency effect with increasing age. Overall, children's numerosity judgments were influenced differently by the visual properties depending on the task at hand and age.

No meaningful relationship was found between the performance on both tasks and mathematics achievement. This contradicts previous studies in which a strong relationship between the performance on nonsymbolic numerosity tasks and mathematics achievement was found in typically developing children (Halberda et al., 2008; Inglis et al., 2011; Mazzocco et al., 2011b). However, it is consistent with the studies that did not find such a (strong) relationship in typically developing children (Defever, Sasanguie, Vandewaetere, & Reynvoet, 2012; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Holloway & Ansari, 2009; Iuculano et al., 2008; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Soltész et al., 2010). On the basis of our findings, we cannot pinpoint the cause of the discrepancies between studies regarding the relationship with mathematics achievement. The absence of a meaningful effect of mathematics achievement does not allow us to examine whether an association between mathematics achievement and numerosity processing, which has been found in previous studies, could be confounded by the visual stimulus properties. Possibly, we did not find a meaningful effect of mathematics achievement due to our visual stimulus controls, which were stricter than the ones used in previous studies. However, this is speculative, and it remains a challenge for future studies to examine what drives the relationship between math achievement and numerosity processing, which is clearly not consistently present across different studies.

In Experiment 1, we examined whether the performance on the comparison and same–different tasks is related to mathematics achievement in typically developing children. To examine whether our results in typically developing children can be generalized to children with MLD, we conducted Experiment 2. It is suggested that children with MLD do not necessarily represent the lower end of a continuum of arithmetical ability but rather can represent a specific and definable impairment (e.g., Desoete et al., 2012; Mazzocco et al., 2008). Hence, the absence of meaningful associations between numerosity processing and mathematics achievement in typically developing children does not necessarily imply that no differences would be observed between children with MLD and controls. Moreover, as mentioned previously in the Introduction, two studies have suggested that children with MLD are more sensitive to the visual stimulus properties than typically developing children (Mejias et al., 2012; Mussolin et al., 2010). To examine whether children with MLD are indeed more sensitive to the visual stimulus properties than typically developing children, we used the same tasks as in Experiment 1 with a group of children with MLD and controls matched on age, gender, and IQ.

Experiment 2

Method

Participants

The MLD group consisted of 33 primary school children with MLD. To be part of the MLD group, children were indicated by their teacher or therapist as having a severe delay in mathematics despite receiving intensive remediation (in and/or outside school). Among this sample, 12 children attended regular schools and 21 children attended special needs schools for children with learning disorders. Moreover, all children with MLD scored below the 10th percentile of the population sample mean on a standardized mathematics achievement test (Tempo Test Arithmetic; de Vos, 1992). In addition, 11 children with MLD scored below the 10th percentile of the population sample mean on a speeded reading test (Brus & Voeten, 1999), indicating that they also had reading difficulties.[2]

The control participants were recruited from a standard elementary school; a large group of fourth, fifth, and sixth graders took part in an arithmetic and IQ assessment. The control participants who matched best with the MLD children on gender, age, and IQ were selected to take part in the experiment.

Thus, 33 children with MLD and 33 gender-, age-, and IQ-matched controls participated in the experiment. Among this sample, 4 children were excluded from the analyses because they scored at chance level on one of the tasks, and 8 children were excluded because they fell within the clinical

---

[2] We verified whether differences in results could be found between children with only MLD and children with MLD and reading difficulties. The analysis revealed no main effect of group ($p = .294$), nor were there any other significant interactions involving group (all $ps > .902$). Therefore, we did not consider them as two separate groups in the analyses.

Table 1
Descriptive statistics of sample in Experiment 2.

	MLD group	Control group
N	21	21
Gender	6 boys and 15 girls	6 boys and 15 girls
Age (years)	11.38 (1.16)	10.99 (0.92)
Math achievement[a]	64.23 (16.15)	116.38 (20.28)
IQ[b]	91.43 (10.51)	93.67 (9.88)

Note. Standard deviations are in parentheses.

[a] Mean raw score on the Arithmetic Tempo Test.

[b] IQ score on the Standard Progressive Matrices.

range for attention deficit/hyperactivity disorder (ADHD) as assessed by an ADHD questionnaire (Scholte & Van der Ploeg, 1998). The children with MLD were excluded to prevent the possibility that attentional difficulties influenced the performance on the experimental tasks. This resulted in a sample of 21 children with MLD (age range = 8.50–12.83 years, $SD = 1.15$) and 21 matched controls (age range = 9.42–12.25 years, $SD = 0.92$) (see Table 1 for detailed descriptive statistics). We used t tests to demonstrate that the MLD group had a significant lower mathematical ability than the control group, $t(40) = 9.217$, $p < .0001$, Cohen's $d = 2.84$. No significant differences in age, $t(40) = -1.217$, $p = .231$, or IQ ($t < 1$) were observed between the groups. Written informed consent was obtained from all of the parents, and the ethics committee of the University of Leuven approved the study.

Apparatus, stimuli, and procedure

Experimental tasks. Comparison and same–different tasks. The tasks were identical to those in Experiment 1 except that only half of the trials were used. As in Experiment 1, regression analyses confirmed that there was no relationship between each visual cue and numerosity for the comparison task (all R^2s < .06, all $ps > .03$) and the same–different task (all R^2s < .07, all $ps > .02$).

Processing speed task. We assessed general processing speed to exclude the possibility that observed differences in reaction time between the MLD and control groups are due to a slower processing speed of children with MLD (e.g., Censabella & Noël, 2005). We administered a similar task as the one used by Reigosa-Crespo and colleagues (2012). A black square was presented in the center of the screen, and children were asked to press the space bar as soon as they saw the square. The interstimulus presentation time varied between 500 and 1500 ms. The test consisted of 20 trials that were preceded by 5 practice trials. Median reaction times were used as a measure of processing speed.

Standardized tests. ADHD questionnaire. Children's teachers completed a Dutch questionnaire to assess whether the children showed behavioral symptoms of ADHD. This questionnaire contains 18 items that assess the presence of inattentive, hyperactive, and impulsive behavior on a 5-point Likert scale (Scholte & Van der Ploeg, 1998).

Speeded reading test. Word reading was assessed with the Version A of the standardized Dutch One-Minute Reading Test (Brus & Voeten, 1999). The test consists of a list of 116 unrelated words of increasing length and difficulty. Children were instructed to read aloud as many words as possible within 1 min without making errors.

Mathematics achievement. All children were tested with the Tempo Test Arithmetic (de Vos, 1992).

Intelligence test. All children completed the Standard Progressive Matrices (SPM; Raven, Court, & Raven, 1992) as a measure of intellectual ability.

General procedure

Data were collected in two sessions. The speeded reading test, mathematics achievement, and intelligence tests were administered in one session, whereas the computerized tests were administered in the other session. Children always started with the processing speed task, after which they performed the number comparison and same–different tasks. The order of the number processing tasks was counterbalanced across participants. Children with MLD performed the computerized tasks individually, whereas the control children were tested in groups of 5 to 7. Children were seated in such a way that they could not distract each other and the experimenter could monitor them closely.

Data analyses

The analyses were identical to those in Experiment 1 except that group was entered as a between-participant factor.

Results

General processing speed

No significant difference in general processing speed was observed between the MLD group ($M = 322$ ms, $SD = 77.46$) and the control group ($M = 299$ ms, $SD = 35.76$), $t(40) = -1.228$, $p = .227$.

Comparison task

The repeated measures ANOVA on children's mean accuracies with congruency and ratio as within-participant factors and group as a between-participant factor revealed a main effect of ratio, $F(4, 37) = 51.408$, $p < .0001$, $\eta_p^2 = .848$, indicating that children's accuracy increased with increasing ratio (see Fig. 4A). Post hoc pairwise comparisons showed a significant difference between all levels of ratio (all *p*s < .050, Cohen's *d*s ranging from 0.49 to 2.44). The significant main effect of congruency, $F(1, 40) = 16.790$, $p < .0001$, $\eta_p^2 = .296$, indicated that participants were more accurate on congruent trials (85%) compared with incongruent trials (70%) (see Fig. 4). As shown in Fig. 4C and D, most participants associated a visually congruent stimulus with a larger numerosity. The interaction between congruency

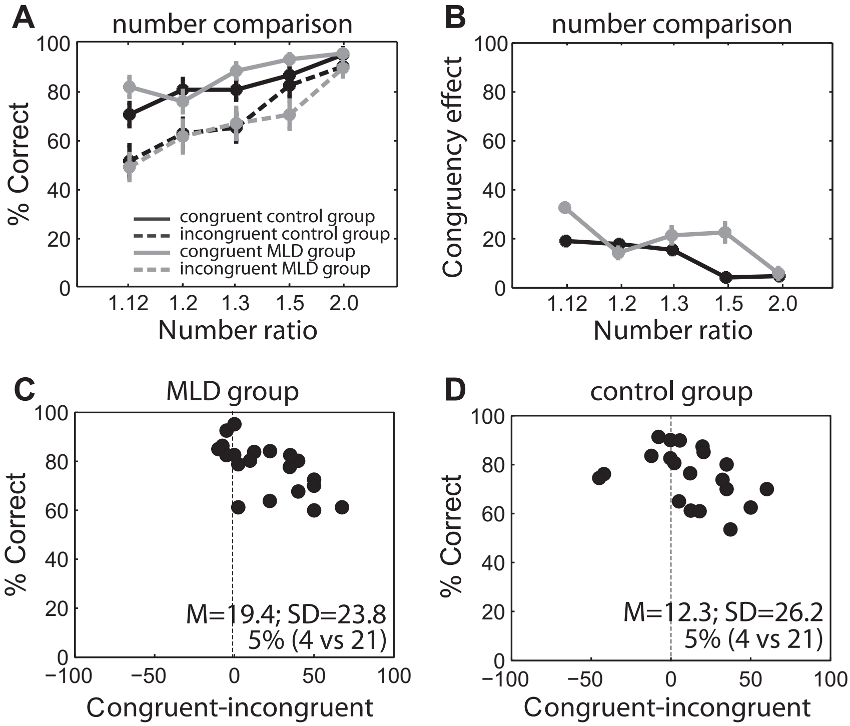

Fig. 4. Ratio and visual congruency effects in comparison task in Experiment 2. (A) Mean accuracies as a function of ratio and group. (B) Visual congruency effect (mean accuracy on congruent trials – mean accuracy on incongruent trials) for each group. (C, D) Visual congruency for each individual in the MLD group (C) and the control group (D). In both groups, only 19% of the children showed a negative visual congruency effect.

and ratio was also significant, $F(4, 37) = 3.111$, $p < .050$, $\eta_p^2 = .252$. As shown in Fig. 4B, the congruency effect decreased with increasing ratio. No effect of group was observed ($F < 1$), nor were there any other significant main effects or interactions (all *ps* > .070).

Same–different task

Analysis of influence of visual cues. The difference between the numbers of different and same responses differed significantly from zero, $t(41) = 7.432$, $p < .0001$, Cohen's $d = 2.29$, showing that participants more often indicated a trial as being numerically different (60% of trials) than the same (40% of trials). The repeated measures ANOVA on the difference between the numbers of different and same responses with ratio as a within-participant factor and group as a between-participant factor showed a significant effect of group, $F(1, 40) = 9.141$, $p < .010$, $\eta_p^2 = .186$, indicating that children with MLD had a larger bias toward responding "different" than controls (i.e., the difference between percentages different and same presses was 12% and 28% for the control and MLD groups, respectively) (see Fig. 5A). A significant effect of ratio was also observed, $F(4, 37) = 10.577$, $p < .0001$, $\eta_p^2 = .533$, showing that the bias toward pressing "different" increased with increasing ratio (4%, 12%, 24%, 29%, and 30% for ratios 1.3, 1.5, 2.0, 2.3, and 2.5, respectively). The interaction between ratio and group was not significant ($F < 1$).

Analysis of ratio effect. The repeated measures ANOVA revealed a main effect of ratio, $F(4, 37) = 15.411$, $p < .0001$, $\eta_p^2 = .625$, which was embedded in a significant interaction between ratio and group, $F(4, 37) = 2.964$, $p < .050$, $\eta_p^2 = .243$ (see Fig. 5). Pairwise comparisons showed that the MLD group were less accurate than the control group for ratio 1.5 ($p = .015$, Cohen's $d = 0.78$) and tended to have a lower accuracy for ratio 2.5 ($p = .064$, Cohen's $d = 0.59$). No significant differences between the groups were present for all other levels of ratio (all *ps* < .102). No significant effect of group was observed, $F(1, 40) = 3.096$, $p = .086$.

Discussion

Similar to Experiment 1, children's judgments were influenced by the visual cues in both the comparison and same–different tasks. In the comparison task, no difference in the ratio effect or the visual congruency effect was observed between the MLD and control groups. This suggests that children with MLD were as efficient as controls in numerosity discrimination and that the influence of the visual properties on numerosity judgment was also comparable to that of controls. In both groups, children predominantly associated a congruent stimulus with a larger numerosity. Only in 19% of the cases (4 of 21 children) was an opposite congruency effect shown.

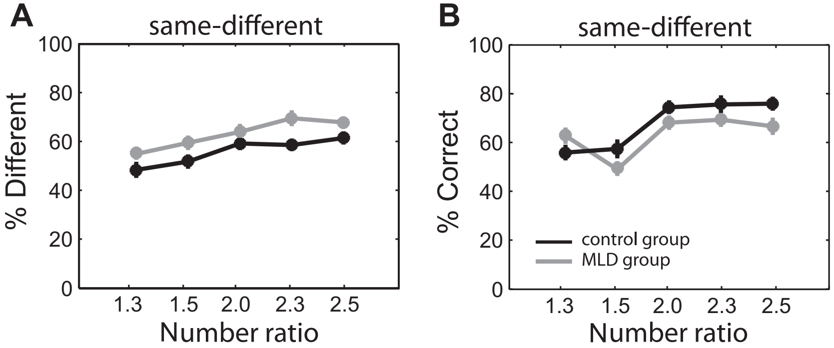

Fig. 5. Children's percentages of "different" responses and mean percentages correct on the same–different task in Experiment 2. (A) Percentages of different responses as a function of ratio and group. (B) Mean accuracies as a function of ratio and group.

The finding that controls and children with MLD are equally sensitive to the visual cues in the comparison task is not in line with the findings of Mussolin and colleagues (2010). These authors observed that, in contrast to controls, children with MLD made more errors when comparing the numerosity of two random stick patterns when the visual properties (e.g., surface area, density) were incongruent with numerosity than when they were congruent. Task differences might explain the differences between our findings and those of Mussolin and colleagues. These authors manipulated surface in half of the trials and manipulated density in the other half, whereas we manipulated four different visual cues across all trials. It should also be noted that we used magnitudes above the subitizing range, whereas Mussolin and colleagues used magnitudes from 1 to 9. It remains to be investigated whether comparison judgments on small and large magnitudes are influenced differently by the visual cues.

In the same–different task, children with MLD showed a lower accuracy for ratio 1.5 and performed slightly worse on ratio 2.5 compared with controls. However, it cannot be concluded that these results point to a clear numerosity processing deficit in children with MLD because a similar performance between the groups was observed for ratios 2.3, 2.0, and 1.3. Children with MLD, however, were more biased toward indicating two stimuli as numerically different than controls, suggesting that they were more influenced by the visual cues in the same–different task. This result is opposite from the results of our comparison task and shows that the results from the same–different and comparison tasks cannot be directly compared. Taken together, our findings show that observed differences between groups might depend on the task at hand.

General discussion

In the current study, we examined how manipulations of the continuous visual properties affect children's numerosity judgments. In Experiment 1, we investigated the development of the interaction between numerosity and its continuous visual properties in first, second, third, and sixth graders. Both a comparison task and a same–different task were administered as well as two standardized mathematics achievement tests. This enabled us to examine whether the effects of the visual stimulus properties on children's numerosity judgments are dependent on task and age and whether they are related to children's mathematics achievement. In Experiment 2, we administered the same numerosity tasks in children with MLD and controls matched on age, gender, and IQ to examine whether children with MLD are indeed more influenced by the visual stimulus properties than typically developing children (e.g., Mejias et al., 2012; Mussolin et al., 2010).

Our findings demonstrated that numerosity judgments were strongly influenced by the manipulations of the visual cues and that these influences changed throughout development. In the comparison task, older children tended to indicate the visually congruent stimulus as the largest in numerosity. In the younger age groups, more variability between participants was observed; half of the children indicated the visually congruent stimulus as the larger in numerosity, whereas this was reversed for the other half. The opposite visual congruency effects in younger children can explain why the effect of the manipulations of the visual cues increased with increasing age. Except for the first graders in Experiment 1, who did not show a significant visual congruency effect, participants performed better on congruent trials compared with incongruent trials and this difference was larger for the older children. The opposite visual congruency effects in younger children cancelled out (for the first graders) and reduced (for the second and third graders) the overall visual congruency effects. Although differences between participants in how they associate numerosity and its visual properties have also been shown in previous comparison studies (e.g., Gebuis & Reynvoet, 2011a, 2011b; Gebuis & Van der Smagt, 2011; Ginsburg & Nicholls, 1988; Tibber et al., 2012), caution is warranted here. We cannot definitely dissociate random variance in performance from opposite visual congruency effects due to different associations between numerosity and its visual cues. Nevertheless, the relatively large congruency effects and the developmental changes in the direction of the congruency effects suggest an influence of the visual cues in the comparison task rather than random variance. Therefore, it seems crucial for future studies to take the age of children and the direction of children's visual congruency effects into account before collapsing congruent and incongruent trials and to report only conclusions about the overall results.

In the same–different task, children showed a bias toward responding that the numerosities were different. This is not surprising because the visual cues of the numerosities always differed for both trials with numerically identical and numerically different numerosity stimuli. We also observed a developmental change in the influence of the visual cues on the same–different task. The bias toward responding "different," which was present for all age groups, decreased with increasing age. The decreasing bias toward responding "different" could suggest that children become more capable to attend to numerosity and ignore irrelevant visual cues with increasing age (e.g., Holloway & Ansari, 2009; Soltész et al., 2010). However, it is also plausible that children have a reduced bias toward pressing "different" because they become more proficient in interpreting the visual cues to judge numerosity. Indeed, Gebuis and Gevers (2011) proposed that interpreting or integrating the information from multiple visual cues might become more efficient with age as children start to get a better understanding of the informative value of each visual cue within a display.

Both Experiments 1 and 2 showed that the visual congruency effect in the comparison task was most prominent for the more difficult (i.e., small) ratios. The visual congruency effect increased with decreasing ratio. This might suggest that children rely more on the visual cues when numerical comparison becomes difficult (see Jordan & Baker, 2011, for a similar idea). In the same–different task, the opposite was found. The bias toward responding "different" was smaller for the small ratio trials compared with the large ratio trials. These results suggest that children found the easier ratio conditions (i.e., large ratios) more visually dissimilar and the more difficult ratio conditions (i.e., small ratios) more visually alike. Thus, the visual cues influenced performance differently in both tasks. Consequently, results from both tasks cannot be directly compared to inform us about numerosity processes. Discrepancies in the influence of visual stimulus properties between the same–different and comparison tasks have been shown in a previous study as well, albeit with numerosities from within the subitizing range and different visual controls (Cantlon, Fink, Safford, & Brannon, 2007). Cantlon and colleagues (2007) showed that the visual similarity of the stimuli (i.e., heterogeneous vs. homogeneous stimuli) influenced the performance on a same–different task. The stimulus heterogeneity did not influence the results in a comparison task. This is not surprising because in the comparison task no judgment on similarity (i.e., same vs. different) needs to be made, but participants need to judge the numerical size (i.e., smaller or larger). However, an influence of visual cues does appear in the comparison task when visually congruent trials are contrasted with visually incongruent trials (e.g., Gilmore et al., 2011; Inglis et al., 2011). In this case, the manipulations of the visual cues (visual cue size) do overlap in response options with those of numerosity (smaller or larger numerosity).

Finally, we examined whether the performance on both tasks was related to children's mathematics achievement and whether performance differs between children with and without MLD. No meaningful association between the performance on the tasks and mathematics achievement was observed in typically developing children, nor were there any differences between children with and without MLD on the comparison task. In the same–different task, children with MLD had a larger bias toward responding "different" than controls. These results add to the large number of previous studies that show inconsistent results regarding the relationship between numerosity processing and mathematics achievement (e.g., Halberda et al., 2008; Iuculano et al., 2008; Mazzocco et al., 2011a, 2011b; Soltész et al., 2010). Our findings make the evidence for a relationship between mathematics achievement and performance on a nonsymbolic discrimination task even less compelling. Considering that we did not find meaningful associations with mathematics achievement in the comparison task, we could not investigate whether the visual stimulus properties play a role in the relationship between math achievement and numerosity discrimination shown by previous studies. However, our findings do not exclude the possibility that discrepancies between studies could be explained by the influence of the visual stimulus manipulations. From our results, it is clear that numerosity judgments do not reflect pure number processing but instead reflect, or are at least influenced by, strategies in processing the visual properties. Therefore, different methods used to control the visual properties of the numerosity stimuli might evoke different visual processing strategies, possibly contributing to the inconsistent results between studies. In addition, the influence of the visual stimulus manipulations differs depending on children's interpretation of the visual cues and on the numerosity task that is used. This might explain why we did not observe differences between children with and without MLD on the comparison task, whereas children with MLD showed a larger bias than controls toward responding "different" in the same–different task.

To conclude, we have examined whether the effects of the visual stimulus properties on children's explicit numerosity judgments are dependent on task and age and whether they are related to children's mathematics achievement. Our results showed that an influence of the visual cues in numerosity tasks is inevitable and that the visual stimulus manipulations directly affect performance. The effects of the visual stimulus manipulations on numerosity judgments changed throughout development and depended on the task at hand (i.e., comparison task or same–different task). Hence, future studies examining how visual cue information affects numerosity processing in more detail should take into account the differential reliance on the visual cues between participants. Our results showed no meaningful associations between the performance on the tasks and mathematics achievement in typically developing children, nor where there consistent differences between children with and without MLD in their performance on the tasks. Therefore, it remains to be seen whether studies that do obtain such an association do find a relation between mathematics achievement and visual processing. Future studies need to shed light on the specific details of the interaction between visual cues and numerosity.

Acknowledgments

Emmy Defever and Bert Reynvoet were supported by the Fund for Scientific Research–Flanders (Grant G.0451.10) and the Research Fund K.U. Leuven, and Titia Gebuis was supported by the Marie Curie Intra European Fellowship (Grant 255140). The authors thank Steffi De Cnodder and Sam Van Snick for assistance with the data collection, Roi Cohen Kadosh for helpful discussion, and all participating children and their parents.

References

Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition, 93*, B59–B68.

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science, 9*, F59–F64.

Brus, B. T., & Voeten, M. J. M. (1999). *Eén-Minuut-Test: Vorm A en B—Verantwoording en handleiding* [One-Minute Reading Test: Forms A and B–Manual]. Nijmegen, The Netherlands: Berkhout.

Cantlon, J., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. *Developmental Science, 10*, 431–440.

Censabella, S., & Noël, M.-P. (2005). The inhibition of exogenous distracting information in children with learning disabilities. *Journal of Learning Disabilities, 38*, 400–410.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*, 408–411.

Clearfield, M. W., & Mix, K. S. (2001). Infants use continuous quantity—not number—to discriminate small visual sets. *Journal of Cognition and Development, 2*, 243–260.

Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. *Child Development, 79*, 476–489.

Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Developmental Science, 12*, 453–463.

Cordes, S., & Brannon, E. M. (2011). Attending to one of many: When infants are surprisingly poor at discriminating an item's size. *Frontiers in Psychology, 2*, 65. http://dx.doi.org/10.3389/fpsyg.2011.00065.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 19552–19557.

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292.

De Vos, T. (1992). *TTR: Tempotest rekenen [Tempo Test Arithmetic]*. Lisse, The Netherlands: Swets & Zeitlinger.

Defever, E., Sasanguie, D., Vandewaetere, M., & Reynvoet, B. (2012). What can the same–different task tell us about the development of magnitude representations? *Acta Psychologica, 140*, 35–42.

Deloof, G. (2005). *LVS–VCLB leerlingvolgsysteem: Wiskunde toetsen 6—basisboek [student monitoring system: Mathematics tests 6—manual]*. Leuven, Belgium: Garant.

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82*, 64–81.

Dudal, P. (2000). *LVS–VCLB leerlingvolgsysteem: Wiskunde toetsen 1—Basisboek [student monitoring system: Mathematics tests 1—manual]*. Leuven, Belgium: Garant.

Dudal, P. (2001). *LVS–VCLB Leerlingvolgsysteem: Wiskunde Toetsen 2—basisboek [student monitoring system: Mathematics tests 2—manual]*. Leuven, Belgium: Garant.

Dudal, P. (2002). *LVS–VCLB Leerlingvolgsysteem: Wiskunde Toetsen 3—basisboek [student monitoring system: Mathematics tests 3—manual]*. Leuven, Belgium: Garant.

Feigenson, L., Carey, S., & Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33–66.

Gebuis, T., & Gevers, W. (2011). Numerosities and space: Indeed a cognitive illusion! A reply to de Hevia and Spelke (2009). *Cognition, 121*, 248–252.

Gebuis, T., & Reynvoet, B. (2011a). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*, 981–986.

Gebuis, T., & Reynvoet, B. (2011b). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*, 642–648.

Gebuis, T., & Reynvoet, B. (in press). Number representations and their relation with mathematical ability. In R. C. Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition*. Oxford, UK: Oxford University Press.

Gebuis, T., & Van der Smagt, M. J. (2011). False approximations of the approximate number system? *PLoS One, 6*. http://dx.doi.org/10.1371/journal.pone.0025405.

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology, 64*, 2099–2109.

Ginsburg, N., & Nicholls, A. (1988). Perceived numerosity as a function of item size. *Perceptual and Motor Skills, 67*, 656–658.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 11116–11120.

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

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17–29.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 19599–19604.

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*, 1222–1229.

Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*, 669–680.

Jordan, K. E., & Baker, J. (2011). Multisensory information boosts numerical matching abilities in young children. *Developmental Science, 14*, 205–213.

Mazzocco, M. M. M., Devlin, K. T., & McKenney, S. J. (2008). Is it a fact? Timed arithmetic performance of children with mathematical learning disabilities (MLD) varies as a function of how MLD is defined. *Developmental Neuropsychology, 33*, 318–344.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*, 1224–1237.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One, 6*, e23749.

Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., & Noël, M.-P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychology, 18*, 550–575.

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition, 115*, 10–25.

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

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*, 547–555.

Raven, J. C., Court, J. H., & Raven, J. (1992). *Standard Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.

Reigosa-Crespo, V., Valdes-Sosa, M., Butterworth, B., Estevez, N., Rodriguez, M., Santos, E., et al. (2012). Basic Numerical Capacities and Prevalence of Developmental Dyscalculia: The Havana Survey. Developmental Psychology, 48.

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

Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology, 44*, 544–560.

Rousselle, L., Palmers, E., & Noël, M.-P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology, 87*, 57–84.

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

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

Scholte, E. M., & Van der Ploeg, J. D. (1998). *ADHD Vragenlijst (AVL): Handleiding [ADHD rating scale: Manual]*. Lisse, The Netherlands: Swets & Zeitlinger.

Soltész, F., Szucs, D., & Szucs, L. (2010). Relationships between magnitude representation, counting, and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions, 6*, 1–13.

Sophian, C. (2007). Measuring spatial factors in comparative judgments about large numerosities. In D. D. Schmorrow & L. M. Reeves (Eds.). *Foundations of augmented cognition* (Vol. 4565, pp. 157–165). Berlin: Springer.

Sophian, C., & Chu, Y. (2008). How do people apprehend large numerosities? *Cognition, 107*, 460–478.

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision, 12*, 8.

Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception, & Psychophysics, 72*, 1839–1853.

vanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science, 9*, F41–F49.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition, 89*, B15–B25.

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*, 88–101.