Cognitive Development 23 (2008) 409–422

Contents lists available at ScienceDirect

Cognitive Development

A validation of eye movements as a measure of elementary school children's developing number sense

Michael Schneider[a,*], Angela Heine[b], Verena Thaler[b], Joke Torbeyns[c], Bert De Smedt[d], Lieven Verschaffel[c], Arthur M. Jacobs[b], Elsbeth Stern[a]

[a] Institute for Behavioral Sciences, ETH Zürich, Switzerland
[b] Department of Psychology, Freie Universität Berlin, Germany
[c] Centre for Instructional Psychology and Technology, Katholieke Universiteit Leuven, Belgium
[d] Centre for Parenting, Child Welfare and Disabilities, Katholieke Universiteit Leuven, Belgium

ARTICLE INFO

Keywords:
Number sense
Eye-tracking
Number line estimation
Mental addition
Mathematical development

ABSTRACT

The number line estimation task captures central aspects of children's developing number sense, that is, their intuitions for numbers and their interrelations. Previous research used children's answer patterns and verbal reports as evidence of how they solve this task. In the present study we investigated to what extent eye movements recorded during task solution reflect children's use of the number line. By means of a cross-sectional design with 66 children from Grades 1, 2, and 3, we show that eye-tracking data (a) reflect grade-related increase in estimation competence, (b) are correlated with the accuracy of manual answers, (c) relate, in Grade 2, to children's addition competence, (d) are systematically distributed over the number line, and (e) replicate previous findings concerning children's use of counting strategies and orientation-point strategies. These findings demonstrate the validity and utility of eye-tracking data for investigating children's developing number sense and estimation competence.

Number sense is the "ability to quickly understand, approximate, and manipulate numerical quantities" (Dehaene, 2001, p. 16). The "mental number line" is regarded as the core neurocognitive system underlying number sense (Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Hubbard, Piazza, Pinel, & Dehaene, 2005; Pinel, Dehaene, Riviere, & LeBihan, 2001), which in turn underlies a variety of

* Corresponding author at: Institute for Behavioral Sciences, ETH Zürich, CAB G84.2, Universitaetsstrasse 6, 8092 Zürich, Switzerland.
E-mail address: schnmich@andrew.cmu.edu (M. Schneider).

0885-2014/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogdev.2008.07.002

behavioral competencies, like estimating, computing, and efficiently using notational systems to solve mathematical problems (Berch, 2005; Jordan, Kaplan, Locuniak, & Ramineni, 2007).

The mental number line represents the magnitudes of numbers in an analogous form, with the smaller numbers on the left and the larger numbers on the right (Dehaene, 1997). Arabic numerals, like 5 and 7, do not allow for any direct inference as to which of them is the one with the higher value. The same is true for number words, like 5 and 7. In contrast, when numerical magnitudes are represented by positions on a number line, one immediately grasps which of them is the higher value number. Therefore, Case and Okamoto (1996) suggested that children use the mental number line "to build models of the conceptual systems that their culture has evolved for measuring such dimensions as time, space. . . They use it to make sense of any direct instruction that they may receive regarding the particular systems that their culture has evolved for arranging numbers into groups and for conducting numerical computations" (pp. 8–9).

A large body of empirical evidence supports the important role of the mental number line with regard to the representation of numerical magnitudes in children and adults (Fias & Fischer, 2005). However, little is known about the influence the mental number line may have on the development of more complex mathematical competencies such as arithmetic or algebra (Campbell, 2005). An important precondition for conducting such studies is knowledge about how number sense and, more particularly, children's use of the mental number line can be measured with an acceptable degree of validity.

1. The number line estimation task as a measure of number sense

As a very broad construct, number sense can be measured in several ways (Berch, 2005; Jordan, Kaplan, Oláh, & Locuniak, 2006). One way, suggested by Siegler and Opfer (2003), is especially useful. Siegler and Opfer asked children to estimate the positions of given numbers on an external number line where only the starting and the end points were labeled. They interpreted the patterns of estimates as indicative of children's representation of magnitudes on their internal number line. These answer patterns relate not only to children's competence in performing other estimation tasks, such as numerosity estimation and computational estimation, but even to children's addition competence and general math achievement (Booth & Siegler, 2006, in press; Siegler & Booth, 2004). So, while the number line estimation task is by no means the only way to assess children's number sense, it is a practical and powerful tool applicable across a wide range of age groups. It has been hypothesized to reflect children's mental representation of numbers more directly than alternative assessments of number sense do.

2. The added value of eye-tracking

Despite its potential benefits, the number line estimation task has a major drawback. Although it is easy to measure such products of children's estimation processes as accuracy, solution times and estimate patterns, it is hard to investigate the processes themselves that children employ to construct their solutions.

In a cross-sectional study with children in Grades 1–3, Petitto (1990) identified two types of solution strategies. To find the magnitude represented by a position on the line, most first graders start at one end of the line and count up or down in whole units or in decades until they reach the target position (*counting-up strategy* and *counting-down strategy*, see also Newman & Berger, 1984). Older children use a second strategy, the *midpoint strategy*, more often. If, for example, the target position is closer to the midpoint of the line than to one of its ends, children start at the middle of the line and count on from there. However, findings regarding strategy use were based only on online observations of children's task performance. Therefore, more detailed data are needed to achieve a more fine-grained characterization of children's interactions with the number line.

In the present study, we analyze the use of eye movements for investigating children's use of the number line when solving number line estimation tasks. Eye-movement data can be collected with high temporal and spatial resolution (e.g., several-hundred measures per second with a spatial precision of 0.01°) notwithstanding the fact that the reliability of the resulting data can suffer from technically caused measurement error and task-irrelevant fixations. Compared to accuracy and speed measures, eye-tracking data potentially provide more direct evidence of the process of problem solving. Moreover, these data are more objective than self-reports or behavioral observations of strategy use.

By means of eye-tracking data, Rehder and Hoffman (2005) demonstrated that in adults, increasing competence in object categorization goes along with an increasing tendency to focus attention on task-relevant characteristics of a problem situation. To guide attention to task-relevant characteristics of a problem situation and to ignore task-irrelevant features is an important part of mathematical competence (The Cognition and Technology Group at Vanderbilt, 1992). Therefore, eye movements might reflect individual differences in mathematical competence. However, surprisingly, eye-tracking has been used mainly to investigate perceptual or reading processes, and there are only a very small number of eye-tracking studies of mathematical cognition. Thus, little is known about the validity or utility of eye movements as a measure of mathematical thinking processes, especially in children.

Two studies (Green, Lemaire, & Dufau, 2007; Verschaffel, De Corte, Gielen, & Struyf, 1994) have demonstrated that eye movements validly reflect various strategies chosen by elementary school children and adults to solve mental addition problems. Other studies have successfully used eye movements to investigate how the structure of word problems affects the reading processes of university students (Verschaffel, De Corte, & Pauwels, 1992). In these studies, either arithmetic expressions or texts were used as stimuli. However, the external number line, as a diagram, is an analogous and more holistic diagrammatic knowledge representation (Larkin & Simon, 1987). The validity of eye movements as an indicator of children's competence concerning this type of external knowledge representation has not been investigated.

3. The current study

Research on the development of children's number sense and its relations to other competencies, we propose, may benefit from the inclusion of eye-movement data because of their objectivity and their potential to reveal underlying processes. However, the validity of eye-movement data for the number line estimation task is as yet unclear and must be established. Furthermore, as noted earlier, little is known about the relation between number sense and the addition skills of typically developing children. Finding such plausible relations would establish the criterion validity of eye-movement data.

In the present study, we tested the validity of eye-movement data as measure of children's developing number sense. In a cross-sectional design with children from Grades 1 to 3, we assessed (a) the accuracy of manual solutions of number line estimation tasks, (b) the accuracy of the positions fixated by gaze while solving a second set of number line estimation tasks, and (c) the accuracy of responses to mental addition tasks.

We addressed the following five questions. First, is grade-related increase in children's estimation competence only reflected by manual answers or also by eye-tracking data? Second, are individual differences in the accuracies of the estimated positions and in the accuracies of their eye movements during the estimation process correlated? Third, are the accuracy of the estimated positions and the accuracy of the positions fixated by gaze during solution production correlated with children's addition competence? Fourth, to what extent does the criterion validity of the eye-movement measure increase with age? Finally, does our eye-movement measure indicate that older children increasingly use the midpoint of the number line as an orientation point?

Assuming that eye-movement measures are sufficiently valid, we expected to find that both manual answers and eye movements show improvement of the number sense with increasing grade level. We further expect the two variables to be correlated at each grade level. We agree with Case and Okamoto's (1996) hypothesis that number sense helps children to understand and carry out mathematical operations. If so, our eye-movement measure should be related to children's addition competence. We additionally expect Petitto's (1990) finding of an increasingly frequent use of the midpoint strategy from Grades 1 to 3 to be replicated. Children's increasing ability to focus their attention on the task-relevant features of a problem situation are likely to reduce the number of task-irrelevant eye movements and, thus, to increase the validity of our eye-movement measure over the three grade levels.

4. Method

4.1. Participants

Sixty-six children from two public primary schools in Berlin participated in the study. Both schools are attended by mostly Caucasian middle-class to upper middle-class children. At each grade level, half of the children were recruited from one school, the other half from the other. The schools did not differ with respect to instructional approaches. All children were selected by teachers on the basis of average or above-average mathematical achievement. This was done to reduce error variance caused by children with very little relevant knowledge guessing the majority of their answers. Parental informed consent was obtained for all participants. The sample included 22 first graders (9 females) with a mean age of 6.8 years, S.D. = 0.8, 20 second graders (12 females) with a mean age of 8.1 years, S.D. = 0.5, and 24 third graders (12 females) with a mean age of 8.9 years, S.D. = 0.5.

4.2. Materials and procedure

4.2.1. Addition accuracy

We assessed children's addition accuracy using 20 addition trials with two-digit addends. The addends were chosen randomly with the only constraint that their sum also is a two-digit number. The tasks were presented on a computer screen and the children had to enter their answers on a keyboard. They could correct their answers until they pressed a confirmation button, in which case the computer scored their answer as either correct or incorrect before presenting the next task. The percentage of correct answers was computed for each child.

4.2.2. Estimation accuracy

To measure estimation accuracy, the children had to solve 30 trials of the number line estimation task. For this task a horizontal number line ranging from 0 to 100 with a number shown above it was presented on a computer screen. The only hatch marks and labels were those at the starting point (0) and the end point (100) of the number line. The length of the line was approximately 16 cm. The children were asked to use the computer mouse to click at the position on the number line which was indicated by the number on top of it. The mouse pointer on the screen was confined to horizontal movement along the number line. Following Rittle-Johnson, Siegler, and Alibali's (2001) example, we coded an answer as correct if it was within an error margin of ±10% of the number line around the actual position of the stimulus on the line, and as incorrect in all other cases. For example, an answer would still be counted as correct if the stimulus 23 was estimated to be located at the position of 13 on the line, but as incorrect if it was estimated to be located at the position of 12. The program automatically coded the correctness of the answers and computed the percentages of correct answers for each child. The stimuli were selected by a pseudo-random algorithm from the natural numbers between 0 and 100.

4.2.3. Fixation accuracy

In the eye-tracking version of the number line estimation task, the same number line as in the behavioral task was used, and children were instructed to actively search for and focus their gaze on the correct position for each number. After 4000 ms, a marker appeared. The children were asked to decide as fast as possible whether the marker position was correct or not and to give their answer by clicking a respective button. Button clicks and reaction times were recorded automatically.

In both settings (i.e. behavioral and eye-tracking), children were seated in front of a computer within a 60 cm distance from the screen (1024 × 768 pixels). Children were familiarized with the tasks during 10 practice trials before each experiment. In these practice trials, children were asked to explain their responses to the experimenter to make sure that their understanding of the tasks was correct. After the practice trials, the main experiment started. No feedback was given to the children during the experimental trials. The order of the two number line estimation tasks was counter-balanced at each grade level.

Eye-movement data were collected using a stationary eye-tracking system with a temporal resolution of 1000 Hz and a spatial resolution of 0.01° (Eyelink 1000®, SR Research Ltd., Mississauga/Ontario, Canada). The data were analyzed using customized software scripts written in Perl (open source, http://www.perl.com/). For each trial, fixation durations and average X, Y positions of gaze were computed. Only fixations within the first 4000 ms of each trial (i.e. before the marker appeared) entered the analyses. Fixations located outside the number line and fixations lasting less than 50 ms were excluded. Trials with more than three blinks during the first 4000 ms and more than one button click were excluded. After excluding invalid trials, an average of 94.4% (S.D. = 7.2) of the data were included in the statistical analyses for each participant.

For each fixation, the position on the number line was computed. Fixations within an error margin of ±10% of the length of the number line around the correct stimulus position were scored as correct, all others as incorrect. For each task and participant, the percentage of correct fixations was computed and, finally, averaged per participant. Although more elaborate ways of coding spatial and temporal patterns in eye movements are available, we decided to use the most basic method in this initial study of eye-tracking in a number line estimation task.

This measure is based on the assumption that children generally have time to make several fixations on the line before the arrow appears and they manually enter their answer. Children who have no knowledge about the location of the given number fixate positions which are randomly distributed on the line. Children who know that the given number lies in a certain segment of the line without knowing the exact positions fixate positions within this segment. The more exactly children know the position of the given number on the line, the smaller the segment they monitor before the arrow appears. Finally, children who know the exact position of the given number on the line will fixate this position almost exclusively. Therefore, the fixation accuracy is expected to increase with increasing knowledge of the organization of whole numbers on the number line.

Of course, children will not only fixate the positions on the line where they think the given number might lie; they will also fixate elements of the task surface necessary to find these positions. For example, children might look at the labels at the beginning and end of the line, at another orientation, such as the middle of the line, and so forth. Therefore, even children with perfect knowledge about the position of all numbers on the line cannot be expected to have fixation accuracies of 100%.

In addition to individual fixation accuracies, we derived the frequency distribution of all fixations over the number line for each grade level by computing the position of each fixation on the number line and by rounding to the nearest whole value. We counted per individual and task how often each of the 101 positions of whole numbers on the line had been fixated. The resulting value for each position was averaged per grade level. To enhance readability, these values were multiplied by 10,000.

5. Results

Results by grade level are shown in Table 1. Each of the three variables shows significant increases across grade levels. Addition accuracy increases most strongly from 15 to 91% and has the highest proportion of explained variance, as indicated by η^2 values. The effect sizes also show that grade-related increases in knowledge are more clearly reflected by estimation accuracy than by fixation accuracy. Of the three variables, fixation accuracy shows the least change and the smallest, albeit still high, proportion of explained variance. The results of planned comparisons show that significant change in estimation accuracy and fixation accuracy occurs only between Grades 1 and 2, but not between Grades 2 and 3. Differences in addition accuracy are highly significant between each of the grade levels.

Our fixation accuracy measure is based on the assumption that not only the last fixation but all fixations during a trial indicate children's knowledge. Exploratory comparisons of (a) the last fixation in each trial with (b) all previous fixations in that trial confirm our expectations. The average percentage of correct fixations in the sample is 42.1% for only the last fixations and 34.0% for all fixations. Both accuracies lie well above chance level (i.e. 20%) and, thus, indicate knowledge. The accuracies for both types of fixations correlate across individuals with $r = .53$, $p = .017$, for first graders, $r = .73$, $p < .001$, for second graders, and $r = .70$, $p < .001$, for third graders, suggesting that the two groups of fixations

Table 1
Grade-related knowledge changes (means in %, standard deviations, ANOVA main effects, Cohen's *d* and error probabilities *p* of planned comparisons)

	Grade 1		Grade 2		Grade 3		Main effect for grade		Grade 2–Grade 1		Grade 3–Grade 2	
	M	S.D.	*M*	S.D.	*M*	S.D.	*p*	η^2	*d*	*p*	*d*	*p*
Estimation accuracy	37.0	20.3	68.8	20.0	82.8	17.3	<.001	.519	1.58	<.001	0.75	.058
Fixation accuracy	24.7	8.0	37.1	9.8	43.0	13.1	<.001	.359	1.39	.001	0.51	.207
Addition accuracy	15.0	5.0	60.0	5.2	91.0	4.8	<.001	.662	1.57	<.001	1.31	<.001

Table 2
Standardized regression weights, proportions of explained variance, and error probabilities from regressions of addition accuracy on estimation accuracy, fixation accuracy, or both

Predictors	Grade 1			Grade 2			Grade 3		
	β	R^2	p	β	R^2	p	β	R^2	p
Estimation accuracy	.545	.262	.009	.573	.290	.008	−.116	.000	.590
Fixation accuracy	.214	.000	.338	.504	.212	.023	−.311	.056	.139
Estimation accuracy and fixation accuracy	–	.228	.033	–	.280	.024	–	.022	.305

measure similar aspects of children's knowledge. Therefore, in all subsequent analyses we use the measure fixation accuracy computed on the basis of all fixations.

Guided by our second research question, we computed the correlations between estimation accuracy and fixation accuracy. These are $r = .28$, $p = .212$, for first graders, $r = .66$, $p = .002$, for second graders, and $r = .63$, $p = .001$, for third graders.

To address our third research question, we tested to what extent children's performance on the number line estimation task is related to their addition competence. The first row of Table 2 shows results for estimation accuracy as a single predictor of addition accuracy. The second row shows results for fixation accuracy as a single predictor of addition accuracy. The last row shows how much of the variance of addition competence is explained if both predictors are simultaneously included in the regression.

In first grade, results for the two predictors differ at a qualitative level. Estimation accuracy, but not fixation accuracy, significantly predicts addition accuracy. In second grade, both predictors are significantly related to addition accuracy and explain about equal proportions of variance. In third grade, neither of the two predictors is significantly related to the criterion. At all three grade levels both predictors entered simultaneously do not explain a larger variance proportion than the best single predictor does. This suggests that both predictors explain the same part of the criterion variance and, thus, assess the same construct.

The frequency distributions for the fixations on the number line are plotted separately for the three grade levels in Fig. 1. The x-axis shows the positions on the number line, the y-axis the absolute number of fixations of each position averaged over children and tasks measured in 1/10,000.

The distributions indicate that children in all three grades fixate positions near the starting point of the line, near the end of the line, and near the number 50 in the middle of the line more frequently than any other position. Thus children fixate different segments of the line with different frequencies.

Fig. 1 suggests that these between-segment differences in fixation frequency differ slightly between the three grade levels. To test whether this interaction of segment and grade on fixation frequency is significant, we transformed the continuous variable position on the line into the categorical variable segment on the line. This allowed us to use segment and grade level as factors in an ANOVA and test for an interaction effect. Thus, we computed the mean number of fixations of five different segments of the line: numbers near the starting point of the line (Segment 1; 0–9), numbers between the starting point and the midpoint (Segment 2; 10–39), numbers near the midpoint (Segment 3; 40–59), numbers between the midpoint and the end point (Segment 4; 60–89), and numbers near the end point of the line (Segment 5; 90–100). The relative stimulus frequency (i.e. the number of stimuli divided by the

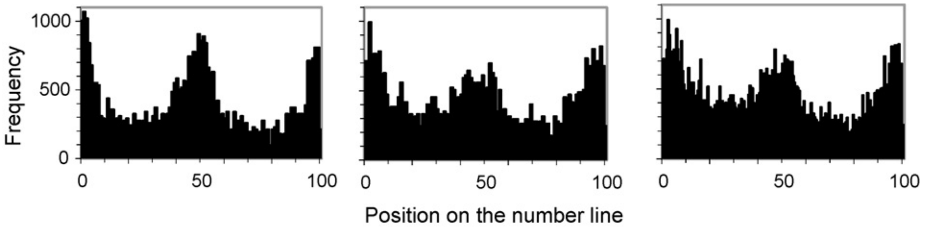

Fig. 1. Distribution of fixations on the number line (left: first grade; middle: second grade; right: third grade).

Table 3
Frequencies with which each number in a segment is on average fixated by each person during a trial (measured in 1/10,000)

	Segment 1 (0–10)		Segment 2 (10–40)		Segment 3 (40–60)		Segment 4 (60–90)		Segment 5 (90–100)		Total line (0–100)	
	M	S.D.	*M*	S.D.	*M*	S.D.	*M*	S.D.	*M*	S.D.	*M*	S.D.
Grade 1	679	327	303	159	613	256	236	141	497	284	403	114
Grade 2	698	159	366	134	526	182	289	133	616	254	435	77
Grade 3	710	303	413	164	551	145	257	117	518	254	435	105
Sample mean	696	309	362	158	564	199	260	130	541	265	424	100

Table 4
Cohen's *d* and error probability *p* of planned comparisons between the fixation frequencies of the number line segments

	Segment 2 (10–39)		Segment 3 (40–59)		Segment 4 (60–89)		Segment 5 (90–100)	
	d	*p*	*d*	*p*	*d*	*p*	*d*	*p*
Segment 1 (0–9)	−1.36	<.001	−0.51	.005	−1.84	<.001	−0.54	.002
Segment 2 (10–39)	–	–	1.12	<.001	−0.71	<.001	0.82	<.001
Segment 3 (40–59)	–	–	–	–	−1.81	<.001	−0.10	.649
Segment 4 (60–89)	–	–	–	–	–	–	1.35	<.001

Table 5
Beta-weights and error probabilities of regressions with the criterion fixation frequency

Predictor	Grade 1		Grade 2		Grade 3	
	β	p	β	p	β	p
Distance from the nearest orientation point	−0.792	<.001	−0.715	<.001	−0.662	<.001
Distance from 0	−0.190	.002	−0.165	.019	−0.372	<.001
Distance from the nearest stimulus	−0.049	.405	−0.027	.696	−0.167	.010

number of all whole numbers) in the five segments is 0.3, 0.3, 0.3, 0.3, and 0.27, respectively. Therefore, differences between fixation frequencies for each of the segments cannot be attributed to different stimulus frequencies.

We computed a repeated-measures ANOVA with the within-subjects factor Segment (1–5) and the between-subjects factor Grade Level (1–3). Means and standard deviations are shown in Table 3. The total sample mean is 424, indicating that each child fixated each number on the line 0.0424 times, on average, during each trial. Since there are 101 numbers on the line, this results in a mean number of about 4.28 fixations per child and trial.

A significant within-subjects effect appeared for Segment, $F(4, 252) = 42.049$, $p < .001$, partial $\eta^2 = .400$[1], but not for Grade Level. Nor was there a significant interaction. Thus, the differences between the three frequency distributions in Fig. 1 are not due to systematic differences in the fixation of orientation points across the three grade levels.

Table 4 shows the results of planned comparisons among the five levels of the factor Segment. All differences except for the difference between Segment 3 (40–59) and Segment 5 (90–100) are highly significant. Cohen's *d*s indicate quite strong effects ranging up to mean differences of 1.84 standard deviations.

In a final step, we used regression analyses to investigate the proportion of children's fixation frequencies (as shown in Fig. 1) due to systematic influences. Three variables were used to predict the fixation frequency for each number on the line ($N = 101$). Since children are assumed to use 0, 50, and 100 as orientation points, the fixation frequency should decrease as the distance to these landmarks increases. The distance to the nearest landmark was thus computed for each position on the line. In the mental representation of numbers, smaller numbers are on the left and larger numbers are on the right (Dehaene, Bossini, & Giraux, 1993). Therefore, we expect children's eye movements to follow the number line from left to right, rather than from right to left, until they find the target position. To test this assumption, we used the distance from zero of each position as a predictor of the fixation frequency, expecting a negative influence. Finally, it can be expected that fixation frequencies for positions used as stimuli in our study are higher than those for other positions, especially in Grades 2 and 3 where estimation accuracies are as high as 68.9 and 82.8%. Therefore, the absolute value of the distance from the nearest stimulus position was used as a third predictor. The three predictors are uncorrelated (*r*s between −.10 and .10; *p*s > .35) and were, thus, simultaneously included in the regression. Results are shown in Table 5. The adjusted R^2 values are .657, .525, and .613 (all *p*s < .001) for the three grade levels, respectively.

The results confirm that the distance from the nearest orientation point is the most important and highly significant predictor of the frequency with which a position on the number line is fixated. This effect is strong at all three grade levels, but decreases from Grades 1 to 3. Similarly, distance from 0 has a much smaller but still significant influence at all three grade levels, while distance from the nearest stimulus has a significant influence only in Grade 3, but not in Grade 1 or 2. In accordance with the high estimation accuracy rate in Grade 3, this shows that children tend to increasingly focus on the correct positions on the number line while solving the estimation tasks.

[1] The partial η^2 is computed by dividing the square sum of a factor by this square sum plus the error square sum. It, thus, gives the explained variance proportion under the assumption that all other factors in the design are controlled for (Pierce, Block, & Aguinis, 2004).

6. Discussion

The present results suggest that eye-tracking data collected with the number line estimation task are a valid, detailed, and sensitive indicator of children's developing number sense. We presented five pieces of empirical evidence in support of this claim, which correspond to our five research questions. Children's fixations (a) validly reflect grade-related competence increases, (b) are closely related, in Grades 2 and 3, to manual solutions of estimation tasks, (c) are related, in Grade 2, to addition competence, (d) are very systematically distributed over the number line, and (e) replicate Petitto's (1990) findings with respect to the use of the midpoint strategy and the counting-up strategy by students in Grades 1–3.

The fact that grade level explains more than one third of the interindividual variance of children's fixation accuracies establishes that eye movements reflect children's increasing knowledge about natural numbers, their interrelations, and ways of their spatial representation. This understanding of natural numbers lies at the very core of children's number sense and is a prerequisite for their future acquisition of more advanced mathematical concepts (Dehaene, 1997).

The finding that eye movements reflect children's increasing number sense is further supported by our second finding. In Grades 2 and 3, there is a highly significant correlation between the accuracies of children's eye movements and their behavioral data. These correlations are notable, especially because in contrast to the manual answers, fixations reflect what is going on during the process of task solution. Children have to orientate themselves on the number line and to use the information given by the labels at its starting point and its end point to locate a specific position between these points. The fact that, for second and third graders, the correlations between the accuracy of fixations and the accuracy of behavioral responses are above .60 suggests that children use relatively direct ways to find the position of the number. For first graders, absence of a significant relation between the manual data and the eye-tracking data may be due to the fact that these children fixate so many different correct as well as incorrect positions during answer production that there is no detectable relation between the accuracy of the final answer and the fixations during answer production. Hence, the criterion validity of eye-tracking data notably increases from Grades 1 to 2.

Our third finding supporting the validity of eye-tracking data is especially interesting since it emphasizes the relation between number line estimation and another important mathematical competence, addition. In Grade 2, ability to solve number line estimation tasks is significantly related to addition competence, both at the level of manual responses and at the level of eye movements. Our behavioral data replicate and extend results by Booth and Siegler (2008), who found that first graders' manual response patterns in the number line estimation task predict their addition competence. We replicate this relation for first graders and demonstrate it also with second graders' manual answers. The first graders scored on the lower end of the scale addition accuracy, while the task was very easy for third graders. Mild ceiling or floor effects in these age groups may explain why the relation between estimation and addition is strongest for the second graders. However, the age groups do not differ in the variances of addition accuracy, leaving it open whether this explanation is correct. In addition to these findings at the level of manually given answers, we show a moderating effect of grade level on the relation between fixations during number line estimation and addition competence.

In Grade 1, only the accuracy of manual estimates, but not eye-tracking data, is a predictor of addition competence. This is additional evidence for the low criterion validity of eye-movement data collected in Grade 1. In Grade 3, neither manual estimation competence nor eye movements predict addition competence. This might indicate that children's addition competence becomes increasingly automatized with practice and, therefore, more and more independent of the number sense. In Grade 2, the criterion validity of eye-tracking data is already high and the number sense still influences children's blooming addition competencies. Therefore, this may be why eye-tracking data predict addition accuracy only in this age group, but not in younger or older children.

Findings on children's individual accuracy data are complemented by comparisons of how fixations are distributed over the 101 numbers of the number line. At all three grade levels, the distance from the nearest orientation point – 0, 50, or 100 – had the strongest influence on how often a position was fixated with the eyes. This confirms Petitto's (1990) finding that even elementary school children use the midpoint strategy to locate numbers on the line, but contradicts Siegler and Opfer's

(2003) hypothesis that second graders, in contrast to sixth graders and adults, generally do not use orientation-point strategies. A further predictor of the fixation frequency at all three grade levels is provided by the position of a number with respect to the right end of the number line. The farther a number was to the right, the less frequently it was fixated, indicating that the counting-up strategy reported by Petitto (1990) was used frequently. Children count up from zero or from the midpoint until they reach the position of the number to be estimated and then stop moving their eyes further to the right. This effect was strongest in third graders, which suggests that younger children's eye movements are somewhat less directed. Finally, the position of our stimuli on the number line predicted the frequency of fixations, but only for third graders. This again shows that third graders' fixations are more direct in targeting the positions of the numbers to be estimated than those of younger children. For all three grade levels, these three predictors explained more than half of the variance of the fixation frequencies, proving that eye movements on the number line reflect children's systematic behavior to a higher degree than random behavior and measurement error taken together.

Elementary school children's eye-tracking data are noisier and hence less reliable than the data of older children and adults (Whiteside, 1974). Yet we have shown that even the eye movements of elementary school children solving the number line estimation task are valid and reliable. With older age groups, our measures can be expected to show even higher validity and reliability.

An important next step in further research is the use of eye-tracking data to identify strategy choices on a trial-by-trial basis, which we did not assess in the present study. The validation of such data requires the use of a second trial-by-trial measure of strategy use, for example, verbal self-reports. A high congruence between the two measures would indicate that eye-tracking data validly reflect which strategy an individual uses in each trial. The development of a valid verbal or manual measure of trial-by-trial strategy use is a challenge on its own, especially with children, who have a limited attention span and cannot always verbalize accurately what they are doing. Therefore, we decided to concentrate on the validation of an eye-movement measure in the current study and to leave in-depth analyses of trial-by-trial strategy use for later studies.

As explained by Verschaffel et al. (1994) and Green et al. (2007), identifying trial-by-trial strategy use by means of eye-tracking data can considerably improve research on the development of mathematical strategies. For example, parallel use of a verbal and a nonverbal measure of strategy use allows researchers to investigate the interaction of explicit and implicit knowledge during strategy development (Siegler & Stern, 1998). However, distinguishing strategy use on a trial-by-trial basis by means of eye-tracking may be as difficult as it is desirable. In most eye-tracking studies the data are aggregated over trials – and often over persons – because 'eye tracking data is never perfect. The system may lose track of the pupil or the corneal reflection, or the observation may be simply incorrect (e.g., beyond the screen limits even though the subject is clearly looking at the screen)' (Aaltonen, Hyrskykari, & Räihä, 1998, p. 135). "Eye-movement data are inherently noisy" (Hornof & Halverson, 2002, p. 593), due to task-irrelevant fixations (e.g., when the individual is distracted by sounds from the surroundings). Individuals may make data-distorting head-movements (children more so than adults) which might even require a re-calibration of the scanning system. Finally, individuals may use peripheral vision to detect information from regions on the screen that they do not fixate directly. All these influences increase measurement error, thus, making it harder to find effects on the trial level than on the level of aggregated data (Rayner, 1998; Verschaffel et al., 1994).

The less than perfect reliability of eye-tracking data is reflected in our data. At all three grade levels, the accuracy of the fixations is lower than the accuracy of the manually given answers to the number line estimation task. This indicates that the eye-movement data indeed reflect additional factors, that is, measurement error, in addition to actual competence.

Eye-tracking data assessing developing number sense may nonetheless contribute to research on the development of mathematical learning disabilities or developmental dyscalculia. The ability to represent and assess numerical magnitude information has been suggested to be a major cause of developmental dyscalculia and related learning deficits (Butterworth, 2005). However, direct empirical evidence on this question remains sparse (Geary & Hoard, 2005). In contrast to, for example, accuracy and speed measures, eye-tracking data may contribute to research on how children with dyscalculia differ from others in their orientation processes on the number line. Ideally, this might lead to insights into the qualitative differences between typical and atypical developments of children's mental magnitude representations.

Finally, and most important, eye movements allow for a direct investigation of how children orient themselves in problem situations and how they direct their attention to specific features. Although previous theories tended to conceptualize mathematical problem solving as an abstract symbol-manipulation process, more recent approaches emphasize the interaction with problem situations as a highly important part of mathematical competencies (Collins, Greeno, & Resnick, 2001). For example, advocates of the situated-cognition view have argued that the ability to pick up action-relevant information from the environment is the most important foundation of competent problem solving and knowledge transfer (Greeno, 1994; Greeno, Moore, & Smith, 1993). Eye movements offer a means of investigating the dynamic and selective search for information in problem situations. Such search invokes higher-level cognitive processes, including what one knows about, and expects of, a situation (Rehder & Hoffman, 2005). More advanced analysis techniques for eye-tracking data, such as fixation density maps (Ouerhani, von Wartburg, Hügli, & Müri, 2004), may become useful for deepening our understanding of these phenomena.

Given these multiple benefits of the eye-tracking method, it is surprising that it is widely and productively used in research on the development of reading processes and yet has been largely ignored in research on the development of mathematical understanding. We hope that the encouraging evidence presented here will help to change this.

Acknowledgements

This work was supported by a grant from the German Federal Ministry of Education and Research (BMBF), funded under the interdisciplinary research initiative Neuroscience, Instruction, Learning (NIL). It was additionally funded by the GOA grant 2006/1 from the Research Fund K.U.Leuven, Belgium.

We thank Katharina Tempel for her assistance with data collection. Bert De Smedt and Joke Torbeyns are Postdoctoral Fellows of the Fund for Scientific Research-Flanders (Belgium).

References

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*(4), 333–339.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *41*, 189–201.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude estimations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031.

Butterworth, B. (2005). Developmental dyscalculia. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 455–467). New York: Psychology Press.

Campbell, J. I. D. (Ed.). (2005). *Handbook of mathematical cognition*. New York: Psychology Press.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, *61*(1/2), 1–295.

Collins, A., Greeno, J. G., & Resnick, L. B. (2001). Educational learning theory. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 4276–4279). Oxford: Elsevier Science.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S. (2001). Précis of "The number sense". *Mind and Language*, *16*, 16–32.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

Fias, W., & Fischer, M. H. (2005). Spatial representations of numbers. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 43–54). New York: Psychology Press.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and non-symbolic magnitude. *Journal of Cognitive Neuroscience*, *15*(1), 47–56.

Geary, D. C., & Hoard, M. K. (2005). Learning disabilities in arithmetic and mathematics: Theoretical and empirical perspectives. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 253–267). New York: Psychology Press.

Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychologica*, *125*, 257–278.

Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, *101*(2), 336–342.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*. Norwood, NJ: Ablex.

Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye tracking data by using required fixation locations. *Behavioral Research Methods, Instruments, and Computers*, *34*(4), 592–604.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*(1), 36–46.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematical difficulties. *Child Development*, *77*, 153–175.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.

Newman, R. S., & Berger, C. F. (1984). Children's numerical estimation: Flexibility in the use of counting. *Journal of Educational Psychology*, *76*(1), 55–64.

Ouerhani, N., von Wartburg, R., Hügli, H., & Müri, R. M. (2004). Empirical validation of the salience-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, *3*, 13–24.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction*, *7*(1), 55–78.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor anova designs. *Educational and Psychological Measurement*, *64*(6), 916–924.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation if parietal activation by semantic distance in a number comparison task. *NeuroImage*, *14*, 1013–1026.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362.

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

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations. *Psychological Science*, *14*, 237–243.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*(4), 377–397.

The Cognition and Technology Group at Vanderbilt. (1992). The Jasper series as an example of anchored instruction: Theory, program description and assessment data. *Educational Psychologist*, *27*, 291–315.

Verschaffel, L., De Corte, E., Gielen, I., & Struyf, E. (1994). Clever rearrangement strategies in children's mental arithmetic: A confrontation of eye-movement data and verbal protocols. In J. E. H. V. Luit (Ed.), *Research on learning and instruction of mathematics in kindergarten and primary school* (pp. 153–180). Doetinchem, The Netherlands: Graviant.

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer's consistency hypothesis. *Journal of Educational Psychology*, *84*(1), 85–94.

Whiteside, J. A. (1974). Eye movements of children, adults, and elderly persons during inspection of dot patterns. *Journal of Experimental Child Psychology*, *18*(2), 313–332.