Abstract

Children’s ability to relate number to a continuous quantity abstraction visualised as a number line is widely accepted to be predictive of mathematics achievement. However, a debate has emerged with respect to how children’s placements are distributed on this number line across development. In the current study, different models were applied to children’s longitudinal number placement data to get more insight into the development of number line representations in kindergarten and early primary school years. Also, longitudinal developmental relations between number line placements and mathematical achievement, measured with a national test of mathematics, were investigated using cross-lagged panel modelling. A group of 442 children participated in a three-year longitudinal study (ages 5-8), in which they completed a number-to-position task every six months. Individual number line placements were fitted to various models, of which a one-anchor power model provided the best fit for many of the placements at a younger age (5-6 years), and a two-anchor power model provided better fit for many of the children at an older age (7-8 years). The number of children who made linear placements also grew with age. Cross-lagged panel analyses indicated that the best fit was provided with a model in which number line acuity and mathematics performance were mutually predictive of each other, rather than models in which one ability predicted the other in a non-reciprocal way. This indicates that number line acuity should not be seen as a predictor of math, but that both skills influence each other in the developmental process.

Keywords: Numerical abilities, Number line, Estimation, Mathematics, Children, Longitudinal
Longitudinal Development of Number Line Estimation and Mathematics Performance in Primary School Children

Will I need to run to be in time for school? If my brother gets three pieces of candy and I get two, is that fair? To answer these questions, one needs an understanding of number, often referred to as number sense, which is children’s ability to intuitively understand and relate numbers (Dehaene, 2001). Number sense is considered a precursor to formal understanding of mathematics (Dehaene, 2001; De Hevia & Spelke, 2009) and therefore of vital importance for later school success.

Recent insights into the development of number sense suggest that children develop an understanding of number, quantity, and relations between numbers at a young age. Although different studies may differ in their definition of number sense and involved skills or abilities, the cognitive tool most often associated with number sense is the mental number line (Dehaene, 1992; Dehaene, Bossini, & Giraux, 1993; Feigenson, Dehaene, & Spelke, 2004; Verguts & Fias, 2004). On this assumed mental number line, numbers are ordered in accordance to their magnitude, and comparisons between numbers can be made by mentally estimating the location of numbers on the number line (Laski & Siegler, 2007). Number line representations are typically investigated using the Number-to-Position task (Siegler & Opfer, 2003). In this task, children are shown a blank number line with the beginning- and endpoint marked with a number (for example, with 0 and 100), and are asked to indicate the position of a certain number on this line by drawing a hatch mark on the location or pointing to the intended location. Number line acuity is thought to be associated with number sense at an early age (e.g., Dehaene, 2001), but in this study assumed to be more dependent on strategy use and taught facts after the onset of formal education. In the current study, longitudinal development of number line placements and its relation to mathematics performance was investigated.
Changes in numerical abilities across developmental time can also be indexed with the Number-to-Position task. As children get older, their estimations of numbers on the number line become increasingly accurate (e.g., Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Friso-van den Bos, Kolkman, Kroesbergen, & Leseman, 2014; Laski & Siegler, 2007). Accuracy of number line placements increases because children learn to consistently place larger numbers on the right side of the number line (Friso-van den Bos et al., 2014), and because children’s ability to determine the spatial distance between placements improves, meaning that they learn to understand that the distance between 10 and 20 on the number line is equal to the distance between 80 and 90 (Laski & Siegler, 2007). These two forms of improvement result in more linear associations between the placements on the number line and the actual numerical value. Linear and accurate placements of numbers on a number line have been shown to be associated with higher mathematics achievement (Geary, 2011; Halberda, Mazzocco, & Feigenson, 2008; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Siegler & Booth, 2004). Therefore, the literature highlights the importance of linear and accurate placements for the development of mathematical achievement.

Models of number line placement

Whereas it has widely been acknowledged that young children’s number line placements do not yet follow a perfectly linear pattern (e.g., Geary, 2011; Halberda et al., 2008; Sasanguie et al., 2013; Siegler & Booth, 2004), different explanations have been given for this reduced linearity. In one of the first accounts of number line placements, Gallistel and Gelman (1992) reported that young children’s number line estimations did follow a linear shape but linear fit of their placements was reduced because of children’s difficulty with accurately placing larger numbers on the number line. More recent accounts, however, state that prior to becoming linear, children distribute numbers logarithmically across the number line.
line and shift towards linear distributions when they get older (Ashcraft & Moore, 2012; Dehaene, 2003; Opfer & DeVries, 2008; Opfer, Siegler, & Young, 2011; Rips, 2013; Siegler & Booth, 2004; Siegler & Opfer, 2003). Children that make logarithmic placements of numbers on a number line intuitively place the numbers on the lower end of the line far apart, and compress the numbers at the end of the scale, as in Figure 1.

Figure 1. Example of logarithmic and linear models, with numbers presented to the children on the x-axis, and placements made by the children on the y-axis.

Others have argued that the association between actual and estimated numbers on a number line can be better explained by a cyclic model, the shape of which results from the use of proportional reasoning to place numbers on a number line (Barth & Paladino, 2011; Hollands & Dyre, 2000; Slusser, Santiago, & Barth, 2013). In this cyclic model, number line placements are made based on a judgment of the magnitude of the target number in comparison to both the minimum and the maximum value on the number line. In other words,
it is suggested that children actively compare between a target number of 90 and a maximum
of 100 in a 0-100 number line, and therefore need to make an estimate of the magnitude of
both the whole number range and the part that needs to be inserted on that range (Barth &
Paladino, 2011; Holland & Dyre, 2000). Biased estimates of both the whole number range
and the proportion of the estimated number result in an overestimation of small numerals and
an underestimation of large numerals (Figure 2B). When the midpoint of a scale is added to
the reference points used to make a placement, this cycle of over- and underestimation
repeats itself past the midpoint, resulting in a two-cycle model (Figure 2C). Whereas the
extent to which children’s placements for a logarithmic curve can be indexed by a
logarithmic model, models of proportional reasoning can be indexed by a power
(exponential) function. Although the shape of these models can also be modelled using
logarithms (Rouder & Geary, 2014), as also done in the current study, they will be referred to
as power models from here on for the sake of consistency with other studies.

There is an on-going discussion between proponents of the logarithmic model and
proponents of the cyclic power models. Various comparisons between these models, in which
children’s data have been fitted to the models in order to compare the adequacy of each
approach, have not yielded consistent results in favour of either model to explain young
children’s number line performance (Ashcraft & Moore, 2012; Bouwmeester & Verkoeijen,
2011; Opfer et al., 2011; Slusser et al., 2013). Rouder and Geary (2014) added a non-cyclic,
power model to the battery of cyclic power functions, which is computationally comparable
to the power functions as presented in other studies (e.g., Opfer et al., 2011), but similar in
shape to the logarithmic model (Figure 2A; Rouder & Geary, 2014; Stevens, 1957).

Importantly, in most accounts of the power model, only the cyclic models are considered, and
the non-cyclic power model is not taken into account (Barth & Paladino, 2011; Opfer et al.,
2011). In the current study, this model is taken into account next to the logarithmic model,
because the differences in computation may produce differences in fit. A third account of number line placements is the segmented linear model, in which the assumption is made that lower numbers are mapped onto the number line in a different way than higher numbers (e.g., Ebersbach et al., 2011; Moeller, Pixner, Kaufmann, & Nuerck, 2009). This model, however, is based on very different theoretical assumptions, and was not taken into account in the current study.

![Graph of non-cyclic power model, one cycle model, and two cycle model.](image)

Figure 2. A. Non-cyclic power model. B. One cycle model. C. Two cycle model. Adapted from “Children’s cognitive representations of the mathematical number line,” by J. N. Rouder and D. C. Geary, 2014, Developmental Science, 17, p. 526. Copyright 2014 by John Wiley & Sons Ltd. Adapted with permission.

It has been proposed that the shape of the number line shifts from logarithmic or non-cyclic power functions to cyclic representations due to practice or the development of other higher-order skills (Rouder & Geary, 2014). Rouder and Geary (2014) proposed that the non-cyclic power model (Figure 2A), which is similar to the logarithmic model in terms of shape, is a model in which a single reference point at 0 is used. On the other hand, the proportional reasoning models rely on two reference points at the beginning- and endpoint of the number line (one-cycle power model, Figure 2B) or three reference points with a reference added in
the middle of the line (two-cycle power model, Figure 2C; Rouder & Geary, 2014; Slusser et al., 2013), and is therefore developmentally more advanced than the non-cyclic power model, with more elements of the number line being used by the child. This means that older children should be more likely to generate cyclic power models than younger children, but studies in which such shifts are investigated are scarce. Support for a shift in the shape of the number line as an indicator of development of numerical reasoning comes from studies showing that older school-age children are more likely to make placements that fit with cyclic power models than younger children who just enrolled in formal education (Barth & Paladino, 2011; Rouder & Geary, 2014). However, support for older children producing estimates that fit a one-cycle model including two reference points, in comparison to younger children generating estimates that fit a two-cycle model including three reference points is also available (Slusser et al., 2013). This finding is contradictory to the notion of the two-cycle model being developmentally more advanced because of the use of three instead of two reference points. Mapping the developmental pathways of number line placements is important, because number line acuity has previously been associated with mathematical performance (Geary, 2011; Halberda et al., 2008; Sasanguie et al., 2013) and may serve as an early marker of difficulties in mathematical performance. However, the associations between number line placements and mathematics achievement are in need of clarification as well.

Number line acuity and mathematics achievement

There are mixed findings with respect to the role of the mental number line in the development of mathematical performance. Although various accounts have demonstrated that children’s number line acuity is predictive of later mathematical achievement (Halberda et al., 2008; Sasanguie, De Smedt et al., 2012; Sasanguie et al., 2013; Siegler & Booth, 2004) and that children with mathematical learning disability show delays in number line acuity (Van Viersen, Slot, Kroesbergen, Van ‘t Noordende, & Leseman, 2013), others have not been
able to demonstrate this relationship (Praet, Titeca, Ceulemans, & Desoete, 2013). Number line acuity may be involved in mathematics performance through calculation using a (mental or printed) number line (e.g. Xenidou-Dervou, Van der Schoot, & Van Lieshout, 2014), or through the use of a mental line in checking the likeliness of the answer to a problem (for example, a child may judge that the answer to 15+17 is unlikely to be 86 using evaluation on a number line). Associations between number line acuity and mathematical achievement have also been found to be bidirectional (LeFevre et al., 2013), which suggests that acuity on number line tasks should perhaps not only be seen as a precursor to mathematics performance, but repeated arithmetic practice might also enhance children’s insight in number relations and hence improve their number line acuity. For example, when a child learns to make an analogy between 3+2 and 93+2 through repeated calculation of the answer, insight into the numerical distance between 3 and 5 and that between 93 and 95 may be fostered through the analogy between the problems 3+2 = 5 and 93+2 = 95. However, LeFevre et al. (2013) used a relatively small and varied sample, and there was a year interval between measurements. Their results are thus in need of replication using a more homogeneous and larger sample of children, with measurements in smaller time intervals. The current study aimed to address these limitations. Moreover, although mathematical performance has been associated with number line acuity, little is known about differences in mathematical performance between children whose number lines adhere to different models of placement, as described above. Studies in which comparisons are made between children falling into different categories of number line placements often use a very limited number of models (Barth & Paladino, 2011; Opfer et al., 2011), making it difficult to observe developmental trends.

To conclude, although research concerning children’s number line estimations has expanded during the past few years, two controversies remain. In the current study, both the
debate regarding the shape of the number line in young school-aged children and the
discussion regarding the role of the number line acuity as a predictor of mathematical
achievement were addressed.

The current study

Three research questions were addressed in this study. Firstly, which model(s)
explains best children’s number line placements from kindergarten up to grade 2? This
research question adds to relevant previous literature (e.g., Ashcraft & Moore, 2012; Barth &
Palladino, 2011; Opfer et al., 2011; Rouder & Geary, 2014) by adopting models already used,
comparing models that have not yet been directly compared, and using a longitudinal design.
More specifically we included three of the models presented by Rouder and Geary (2014): a)
a non-cyclic power model, b) a one-cycle model in which two anchor points are used at the
beginning- and endpoint of the number line, and c) a two-cycle model in which three anchor
points are used: the beginning-, middle- and endpoint (see: Figure 2). Furthermore we
included logarithmic and linear models (e.g., Siegler & Booth, 2004; Siegler & Opfer, 2003)
and introduced a random model to identify children whose placements did not sufficiently
relate to the presented numbers to be reliably associated with one of the above models (see
Friso-van den Bos et al., 2014). Importantly, in the current study, no instruction was given to
the participants with respect to the midpoint of the number line, as this may serve as a
determinant of strategy selection (Ashcraft & Moore, 2012).

These models were applied to data from a longitudinal study in which the
performance of a large sample of children was measured six times (twice a year) in a period
from kindergarten to grade 2. At each longitudinal measurement point children were
categorised on the basis of a strategy associated with one of the resulting six models using the
fit index $R^2$ (Opfer et al., 2011). Children were placed into the category that produced the
highest $R^2$ fit, regardless of the difference with fit of the next-best fitting category. Although
the analyses were in general exploratory, we expected to find models indicative of one reference point to be more prevalent in younger children, and models with multiple reference points to be more prevalent in older children, similar to the findings of Rouder and Geary (2014).

Secondly, we addressed the question: Do placement category groups at each time point differ with respect to mathematical achievement? This question targeted the hypothesised developmental account of number line placements. If children whose placements adhere to the more advanced cyclic models indeed score higher than children whose placements suggest a less advanced single reference point (non-cyclic power models or logarithmic models), and if children with linear placements score higher than both former groups on a mathematics test, this would confirm earlier suggestions that placements with more hypothesised reference points are indicative of more advanced number processing (Rouder & Geary, 2014; Slusser et al., 2013).

The third research question was: Is number line acuity a predictor of mathematics achievement, is mathematics achievement a predictor of number line acuity, or is the relationship bidirectional? We hereby aimed to address the discussion in the literature regarding the role of number line acuity as a predictor of mathematics achievement (e.g., Lefevre et al., 2013; Praet et al., 2013; Sasanguie et al., 2013). Only the children’s fits according to the linear model were used to address this research question, because this model is developmentally most advanced (Friso-van den Bos et al., 2014; Siegler & Booth, 2004; Slusser et al., 2013) and provides the best view on how accurately a child can place numbers.

Method

Participants

Data were from the longitudinal MathChild study\(^1\) in which children were followed from kindergarten to second grade of primary school, across a timespan of three academic
At the start of the study, 442 children were included with a mean age of 5 years and 7 months ($SD = 4.3$ months), and 198 ($44.8\%$) were girls. The children were recruited from a total of 25 schools in various municipalities in The Netherlands. Children completed a diverse battery of tasks twice per academic year, once in November/December, and once in May/June, resulting in six time points with six month intervals (from here on referred to as T1-T6). During the sixth and final round of data collection in grade two, 354 participants completed the tasks presented in the current study with a mean age of 8 years ($SD = 3.9$ months). Reasons for dropout varied, but the most common reasons were repeating a grade, which is very common in Dutch education, and moving to a different school or municipality. On average, children who dropped out showed less linearity in their placements ($R^2 = .19$) than children who did not drop out ($R^2 = .33$) during the first round of data collection, $t(440) = 4.19, p < .001$, and scored lower on Raven’s coloured matrices ($M = 17.76$) than children who did not drop out ($M = 21.60$), $t(434) = 5.47, p < .001$, which may be explained by the fact that the dropout group includes the children repeating a grade.

**Measures**

**Number-to-position task.** The number-to position task was a computerised version of the task initially designed by Siegler and Opfer (2003; Kolkman, Kroesbergen, & Leseman, 2013). In the task, children were presented with a horizontal line on the computer screen and were told that they would see numbers (Arabic numbers) that had to be placed in a line by the children, and that each number needed to get its own spot on the line. The numbers 1 and 100 were presented below the left and right ends of the line, respectively, and the target number was presented above the line (see Figure 3). In a first practice trial, the children were asked where the number 1 would go on the line, and in a second practice trial, they were asked where the number 100 would be located. Children pointed to this position with a finger on the computer screen. The correct placements were pointed out at both
practice trials, after which the test trials started. Note that the number 0 was deliberately omitted from the number line, both to circumvent problems with the integration of 0 in a numerical continuum (e.g., Merritt & Brannon, 2013), and to make the task analogue to a non-symbolic counterpart, which was also part of the test battery but not the focus of these analyses. During the testing phase, no feedback was given to the children, except for positive reinforcement. The numbers used in the test trials were 2, 4, 9, 11, 14, 17, 23, 26, 31, 38, 44, 45, 52, 59, 61, 66, 73, 78, 84, 86, 92, and 99. Numbers below 20 were slightly oversampled, consistent with other studies (Lasi & Siegler, 2007; Siegler & Opfer, 2003). These numbers were presented in a random order. Positions indicated by the child were entered into the computer by the experimenter, by dragging a digital hatch mark to the place the child had indicated. Children were instructed not to remove their finger from the target position until the experimenter had entered the response, for minimal error in data entry. Positions were saved digitally, ranging from 0,0 at the far left of the line to 100,0 at the far right of the line.

Figure 3. Number-to-position task as presented to the child, and a position as it might be indicated by a child.
CITO Mathematics Tests. The national Cito Mathematics Tests monitor the progress of primary school children. Each academic year starting in grade 1, two tests are administered: one in the middle (January) and one at the end (June) of the academic year. Each test consists of grade-appropriate mathematics problems, increasing in difficulty across grades, to be completed in full by all the children. The tests consist of primarily word problems that cover a wide range of mathematics domains, such as measurement, time, and proportions. Test scores are converted into normed ‘ability scores’ provided by the publisher that typically increase throughout primary school, making a comparison of results throughout the academic career possible (Janssen, Scheltens, & Kraemer, 2005). The Cito Mathematics Tests have been shown to be highly reliable; the reliability coefficients of different versions range from .91 to .97 (Janssen, Verhelst, Engelen, & Scheltens, 2010).

Procedure

Prior to the study, informed consent was obtained from all the parents or caretakers of children participating in the study. Children in the MathChild study participated during six rounds of data collection, each consisting of two or three sessions that lasted up to half an hour. In each academic year, a round of data collection was planned in November/December, and one in May/June.

Children were tested in a quiet room inside the school by trained research assistants at times convenient to both the teacher and child. All tests except the CITO Mathematics test were computerised and presented on HP 6550b notebooks. In the current study, only data from the Number-to-position task and the CITO Mathematics test were used. During testing, positive feedback was given to the child about effort, but not performance. After completing all the tasks planned for a session, children were rewarded with a colourful sticker.

Analytical strategy
For each child at each time point, number line placements of each item were recorded. Using various formulas, for each individual child and at each longitudinal measurement point a fit of the data with the various models of number line placements was computed using the fit index $R^2$ (for the logarithmic model, see: Siegler & Opfer, 2003; for the non-cyclic power, one-cycle, and two-cycle model, see: Rouder & Geary, 2014; linear fit was indexed by the squared correlation of untransformed values). If the correlation between presented items and placements by the child did not exceed $r = .30$, placements were coded as random, because effect sizes below .30 are considered small (Cohen, 1992). In all other cases, the model producing the highest fit with the data was selected as the model best fitting the child at that time point to address the first research question (which model(s) can best explain children’s number line placements from kindergarten to second grade of primary school?). Transitions between these models were recorded for each child longitudinally.

To address the second research question (do placement category groups at each time point differ with respect to mathematical achievement?), analyses of variance were applied to test for potential differences in mathematical performance between children placed in different categories of number line placement (based on their best fit scores) at different ages. In case of a significant main group effect, Tukey’s post-hoc tests were performed to test contrasts between specific groups of children.

To address the third research question (is number line acuity a predictor of mathematics achievement, is mathematics achievement a predictor of number line acuity, or is the relationship bidirectional?), a series of cross-lagged panel models (Kenny, 2005) was built using Mplus software (Muthén & Muthén, 1998-2011). Although cross-lagged panel analysis cannot prove causality between variables, a strong claim for causal relations can be made because of the prediction of scores across time, controlling for autoregressive effects, and because causality in both directions can be investigated. Only data from first and second
grade were used for this, because mathematics scores were available only from the start of first grade, since these tests cannot be completed by kindergartners. First, correlations between the estimated position indicated by the child and the actual position of the number values were computed for each child at each longitudinal time point. Correlations can be interpreted in a similar manner as the linear model of number line placements, reported in, for example, Siegler and Opfer (2003). A starting model included these linear correlations as an indicator of number line acuity, and scores on the mathematics test at each longitudinal measurement point. To answer the research question about mutual interdependencies between mathematical achievement and number line acuity, five different models were tested:

a. an empty model, containing only autoregressive effects and co-variances between number line acuity and mathematics achievement at the first and last time point,

b. a model in which paths from number line performance at each time point to mathematics achievement at the next time point were added,

c. a model in which paths from mathematics achievement at each time point to number line achievement at the next time point were added, but no paths from number line to mathematics were included,

d. a model in which both paths from number line to mathematics and from mathematics to number line performance were included,

e. a model in which the best-fitting model of the former was adjusted to achieve the best possible fit.

Model fit for each model was evaluated using various cut-off criteria commonly accepted for statistics of model fit, reported in Table 4 (Hu & Bentler, 2009; Schermelleh-Engel & Moosbrugger, 2003). Reported fit statistics are the Root Mean Square Error of Approximation (RMSEA, smaller values are indicative of better fit), the Comparative Fit Index (CFI, higher values are indicative of better fit), the Tucker-Lewis Index (TLI, higher
values are indicative of better fit), and the Standardised Root Mean Residual (SRMR, smaller values are indicative of better fit). Moreover, the ratio $\chi^2$ to degrees of freedom was evaluated, with smaller values indicative of better fit, as an alternative for the $\chi^2$ test, which has drawbacks when large samples are being examined (Schermelleh-Engel & Moosbrugger, 2003).

Comparisons between fit indices addressed the research question: Of models a-d, the model with the best fit best described the relationship between number line acuity and mathematical achievement, allowing us to conclude whether associations are unidirectional and in which direction, or bidirectional. Comparisons between the Satorra-Bentler scaled $\Delta \chi^2$ test (Satorra & Bentler, 2010) provided information about the significance of differences in fit between nested models. The final model(s) (i.e., chosen best fitting model) was used to explore an optimal model. In this model, added paths were maintained only if they made a significant contribution to model fit as indexed by the Satorra-Bentler scaled $\chi^2$ (Satorra & Bentler, 2010), which is superior to $\chi^2$ difference testing to compare models.

**Results**

**Number line models**

First, each child was placed into a category of number line placements based on his/her best fit across various models. The model with the highest $R^2$ value was considered to be the best-fitting model. The number of children showing the best fit for each of the models of number line placements can be found in Table 1. The most dominant category of number line placements was the non-cyclic power model for kindergarten and grade 1 (T1-T4), and the one-cycle power model for grade 2 (T5 and T6). Across time, an increasing number of children were placed into the category of linear placements. A graphical representation of transitions between categories of number line placements can be found in Figure 4. This graph shows both stability within categories and transitions in all directions, but the most
obvious pattern was the stability in categories in which one reference point is used (logarithmic or non-cyclic power models) in kindergarten and first grade (e.g., 232 children showed stability between T1 and T2, fitting best into one reference point models at both time points). Other very frequent patterns were transitions to cyclic models (one- or two-cycle model) or change to a linear model in second grade (e.g., 57 children went from a two reference point model to a linear model from T5 to T6).

In a next step, the non-cyclic power category was removed, and children fitting best into the non-cyclic power model were placed in the next best fitting category of number line placements (the model with the highest fit of all models, but explicitly not the non-cyclic power model). This was done to gain insight into placements into models when the non-cyclic power model, as the most prevalent model, is disregarded, similarly to comparable studies (Barth & Paladino, 2011; Opfer et al., 2011). The number of children placed in each category after removal of the non-cyclic power model can be found in Table 2, and by subtracting the original number of children in each category as presented in Table 1 from the number of children placed in the same category in Table 2, one can compute the number of children moving to that category when the non-cyclic power model is disregarded. In kindergarten, most children whose number line placements fit a non-cyclic power model show a logarithmic model as next best fitting category, while a one-cycle power model would fit their data better after the start of formal education (T3 and later time points). The number of children whose next best fitting category was a linear model increased across time, $\chi^2(5, N = 96) = 55.00, p < .001$. 
Table 1

Number of Children Fitting into Categories of Number Line Placements for All Time Points

<table>
<thead>
<tr>
<th>Category</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>$R^2$</td>
<td>$n$</td>
<td>$R^2$</td>
<td>$n$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Random</td>
<td>95</td>
<td>-</td>
<td>58</td>
<td>-</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>50</td>
<td>.44</td>
<td>48</td>
<td>.53</td>
<td>55</td>
<td>.64</td>
</tr>
<tr>
<td>Non-cyclic</td>
<td>252</td>
<td>.53</td>
<td>257</td>
<td>.62</td>
<td>243</td>
<td>.74</td>
</tr>
<tr>
<td>One-cycle</td>
<td>34</td>
<td>.38</td>
<td>55</td>
<td>.48</td>
<td>74</td>
<td>.65</td>
</tr>
<tr>
<td>Two-cycle</td>
<td>8</td>
<td>.18</td>
<td>5</td>
<td>.22</td>
<td>3</td>
<td>.33</td>
</tr>
<tr>
<td>Linear</td>
<td>3</td>
<td>.31</td>
<td>7</td>
<td>.42</td>
<td>14</td>
<td>.57</td>
</tr>
<tr>
<td>Mean age (y;m)</td>
<td>5;7</td>
<td>6;0</td>
<td>6;6</td>
<td>7;0</td>
<td>7;6</td>
<td>8;0</td>
</tr>
</tbody>
</table>

Note. $R^2$ values are the average model fits within each time point, for all participants.
Table 2

Number of Children Fitting into Categories of Number Line Placements for All Time Points, Excluding the Non-cyclic Power Model

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N = 442)</td>
<td>(N = 430)</td>
<td>(N = 398)</td>
<td>(N = 394)</td>
<td>(N = 363)</td>
<td>(N = 354)</td>
</tr>
<tr>
<td>Random</td>
<td>95 (21%)</td>
<td>58 (13%)</td>
<td>9 (2%)</td>
<td>2 (1%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>226 (51%)</td>
<td>215 (50%)</td>
<td>177 (44%)</td>
<td>73 (19%)</td>
<td>14 (4%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Non-cyclic power</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>One-cycle model</td>
<td>103 (23%)</td>
<td>136 (32%)</td>
<td>186 (47%)</td>
<td>248 (63%)</td>
<td>233 (64%)</td>
<td>189 (53%)</td>
</tr>
<tr>
<td>Two-cycle model</td>
<td>11 (2%)</td>
<td>8 (2%)</td>
<td>7 (2%)</td>
<td>5 (1%)</td>
<td>5 (1%)</td>
<td>2 (1%)</td>
</tr>
<tr>
<td>Linear</td>
<td>7 (2%)</td>
<td>13 (3%)</td>
<td>19 (5%)</td>
<td>66 (17%)</td>
<td>111 (31%)</td>
<td>161 (45%)</td>
</tr>
</tbody>
</table>
Figure 4. Transitions between models fitting each child’s data, with logarithmic and non-cyclic power model grouped under “1 reference model”, and 1-cycle and 2-cycle models grouped under “2 reference model”. Arrow sizes represent the number of children making a transition.
Mathematical achievement differences between children in number line categories

As a next step, at each longitudinal measurement point we tested for potential differences in mathematics proficiency between the categories of number line placement in which children were divided based on their best fit. This was done with a series of one-way ANOVAs with number-line acuity category (e.g., random, one-or two-cyclic, etc.) as between subjects factor. Since scores of mathematics proficiency were only available starting from T3 (from first grade onwards), four different analyses were performed (T3, T4, T5, T6). Mean mathematics scores per group, as well as number of children in each category can be found in Table 3. Analyses of homogeneity of variances, an assumption of the ANOVA, yielded no problematic results. In case of significant main group effects, Tukey’s post-hoc tests were used to test for differences between the number line acuity categories.

Table 3

<table>
<thead>
<tr>
<th>Category</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>N</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Random</td>
<td>27.00</td>
<td>8</td>
<td>15.00</td>
<td>1</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>30.85</td>
<td>53</td>
<td>38.95</td>
<td>20</td>
</tr>
<tr>
<td>Non-cyclic power</td>
<td>35.25</td>
<td>224</td>
<td>42.44</td>
<td>176</td>
</tr>
<tr>
<td>One-cycle</td>
<td>42.17</td>
<td>69</td>
<td>50.18</td>
<td>122</td>
</tr>
<tr>
<td>Two-cycle</td>
<td>34.00</td>
<td>3</td>
<td>80.00</td>
<td>2</td>
</tr>
<tr>
<td>Linear</td>
<td>50.23</td>
<td>13</td>
<td>57.58</td>
<td>45</td>
</tr>
</tbody>
</table>
At T3, there was a significant difference between groups of number line acuity with respect to mathematical achievement, $F(5, 364) = 6.87, p < .001$. Post-hoc analyses indicated that children in the linear group scored significantly higher than children in the random, logarithmic, and non-cyclic power groups ($p < .001$), and that children in the one-cycle group scored significantly higher than children in the logarithmic and non-cyclic power groups ($p < .01$). No other contrasts produced significant differences.

At T4, there was also a significant difference between groups of number line placement with respect to mathematical achievement, $F(4, 360) = 13.59, p < .001$. Post-hoc analyses indicated that children in the linear and one-cycle power group scored higher with respect to mathematical achievement than children in the logarithmic and non-cyclic power group ($p < .05$). Contrasts with the two-cycle power group could not be interpreted because of the low number of children in this group. No post hoc contrasts were computed for the random group, because only one child was available for whom both number line and mathematics data were available.

At T5, there was also a significant difference between groups of number line placement with respect to mathematical achievement, $F(4, 347) = 7.46, p < .001$. No children were placed in the random group at this time point. Post-hoc analyses indicated that both the linear and the one-cycle group scored higher on mathematics than children in the non-cyclic power group ($p < .001$). No other contrasts were indicative of significant differences ($p > .05$).

Finally, at T6, there was a significant difference between groups of number line placement with respect to scores of mathematics, $F(2, 344) = 4.16, p = .01$. Post hoc analyses indicated that children in the linear group scored significantly higher than children in the non-cyclic power group ($p = .03$) and marginally higher than children in the one-cycle power group ($p = .05$). The difference between the non-cyclic power and one-cycle group was not
significant \((p = .86)\), and contrasts with the logarithmic group or two-cycle group were not computed because only one child in these groups had a mathematics score available.

**Longitudinal associations between number line acuity and mathematics**

To address the third research question, regarding the longitudinal associations between mathematics achievement and number line performance, a series of path analyses was conducted. An empty model, found in Figure 5A, contained no cross-lagged paths, but only paths between measures at each time point and the same measure at previous time points, or autoregressive associations. Covariances between number line performance and mathematics achievement at T3, and number line performance and mathematics achievement at T6 were also added. Moreover, direct paths were added between number line acuity at T3 and number line acuity at T5, and between mathematics achievement at T3 and mathematics achievement at T5 and T6, because these paths improved the \(\chi^2\) fit of the models greatly without affecting the associations between number line acuity and mathematics achievement. The latter associations were used for hypothesis testing.

Next, three hypothesis-testing models were explored, which were all extensions of the empty model, meaning that all paths in the empty model were nested in all consecutive models: A model containing only paths from number line acuity to mathematics achievement at the next time point (Figure 5B), a model containing only paths from mathematics achievement to number line acuity at the next time point (Figure 5C), and a full cross-lagged panel model with bidirectional associations (Figure 5D). Fit indices of these models can be found in Table 4. Of these models, only the fit indices of the full cross-lagged model were acceptable.

The full cross-lagged model (Figure 5D) demonstrated a better fit than both the Number line to maths Model (Figure 5B), \(\Delta \chi^2 = 72.45, \Delta df = 3, p < .001\), and the Maths to
number line model (Figure 5C), $\Delta \chi^2 = 40.97, \Delta df = 3, p < .001$. This confirms that the Full cross-lagged model described the data better than the other models.

In a final step, the Full cross-lagged model was adjusted to determine whether a more optimal fit could be found. First, the non-significant path from number line performance at T5 to mathematics achievement at T6 was removed, leading to a non-significant decrease in fit and thus a better and more parsimonious model, $\Delta \chi^2 = 1.21, \Delta df = 1, p = .21$. Then, additions to the model were explored in which mathematical achievement and number line were predicted from two time points earlier, being the same month of the year, one year previous, and the only additional path that made a significant contribution to the model was the path from number line acuity at T3 to mathematics achievement at T5, $\Delta \chi^2 = 11.87, \Delta df = 1, p < .001$. The final best-fitting model is presented in Figure 5E, and fit statistics of this model can be found in Table 4 in the row Improved cross-lagged model. In this model, approximately 16% of variance in number lines at T6 is explained by predictor variables, and 65% of mathematics scores at T6. Please note that the high explained variance in mathematics scores is mostly based on stability within the construct, as indicated by the standardised weights reported in Figure 5E. All fit statistics were indicative of acceptable to good fit.
### Table 4

Fit Indices of Path Models A-E.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>RMSEA</th>
<th>CFI</th>
<th>TLI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Empty model</td>
<td>166.38</td>
<td>17</td>
<td>9.79</td>
<td>.15</td>
<td>.86</td>
<td>.77</td>
<td>.19</td>
</tr>
<tr>
<td>B Number line to maths model</td>
<td>120.88</td>
<td>14</td>
<td>8.63</td>
<td>.14</td>
<td>.90</td>
<td>.80</td>
<td>.15</td>
</tr>
<tr>
<td>C Maths to number line model</td>
<td>94.35</td>
<td>14</td>
<td>6.74</td>
<td>.12</td>
<td>.92</td>
<td>.85</td>
<td>.13</td>
</tr>
<tr>
<td>D Full cross-lagged model</td>
<td>48.25</td>
<td>11</td>
<td>4.39</td>
<td>.09</td>
<td>.96</td>
<td>.91</td>
<td>.08</td>
</tr>
<tr>
<td>E Improved cross-lagged model</td>
<td>35.43</td>
<td>11</td>
<td>3.22</td>
<td>.07</td>
<td>.98</td>
<td>.94</td>
<td>.07</td>
</tr>
</tbody>
</table>

**Fit criteria**

- **Acceptable fit**
  - $\chi^2 \leq 5.0$
  - $< .08$
  - $\geq .90$
  - $\geq .90$
  - $\leq .10$

- **Good fit**
  - $0 \leq \chi^2$/df $\leq 2$
  - $< .05$
  - $\geq .95$
  - $\geq .95$
  - $.00 \leq SRMR \leq .05$

Note. $\chi^2$ = chi square statistic; df = degrees of freedom; $\chi^2$/df = chi square and degrees of freedom ratio; RMSEA = Root Mean Square Error Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardised Root Mean Square Residual.
Figure 5. A. Empty model with no cross-lagged paths. B. Number line to mathematics model. C. Mathematics to number line model. D. Full cross-lagged model. E. Improved cross lagged model. All estimates are standardised coefficients. * $p < .05$. ** $p < .01$. *** $p < .001$. 
Discussion

In the current study, various models of number line placements were compared across a series of longitudinal measurements from kindergarten until grade 2 – a period during which number line acuity grows considerably. We found the non-cyclic power model demonstrating the best fit for a large number of children’s data up to grade 1, and the one-cycle power model in grade 2. The logarithmic model was less frequently found to be the best-fitting model. The non-cyclic power model is similar in shape to the logarithmic model, but is ignored in many studies in favour of the one- and two-cyclic models whose cyclic shape is thought to result from the use of multiple reference points when making number line estimations (Barth & Paladino, 2011; Opfer et al., 2011). Although we can conclude that a power model (either cyclic or non-cyclic) indeed produces a better fit for most children’s number line placements, the interpretation of these data is closer to that of the studies in which a logarithmic model is proposed (Ashcraft & Moore, 2012; Dehaene, 2003; Opfer & DeVries, 2008; Opfer et al., 2011; Rips, 2013; Siegler & Booth, 2004; Siegler & Opfer, 2003): One dominant reference point is used to obtain data fitting both the power model and the logarithmic model.

It should be noted that the logarithmic model and the non-cyclic power model are very similar in shape and mathematical properties. Both models imply no difference in strategy taken by the child, nor do they differ with respect to assumptions regarding reference points used. Their difference is purely computational, although very relevant, as evidenced by the differences in best-fitting model, outlined in the Results section. The logarithmic model, therefore, remains suitable to compare between logarithmic fit and linear fit, as is done in various studies (e.g., Ashcraft & Moore, 2012; Opfer & DeVries, 2008; Siegler & Booth, 2004), and results of these studies can be interpreted in a meaningful way, despite the fact that the non-cyclic power model provided a better fit in the current study. The power models
presented in Rouder and Geary (2014) are theoretically less suitable to make the comparison between linear and pre-linear placements: In the unlikely case of perfect placements, this model is not statistically distinguishable from a linear model. Note, however, that any deviation from perfect placements makes models statistically distinguishable.

When the non-cyclic power model is disregarded, children in kindergarten grades are more likely to make placements best fitting the logarithmic model, and there is a gradual developmental shift towards the one-cycle power model as the statistically next best-fitting model across time points. Two inferences can be made from these data: First, the logarithmic model, despite it being inferior in fit to the non-cyclic power model as evidenced by the smaller number of children fitting the model best, quite adequately described number line placements of children before the start of formal education. Second, the fact that children best fitting the non-cyclic power model did not all have the same statistically next best-fitting model suggest that the shift from a model in which one reference point is used towards a model in which multiple reference points are used is not sudden and paradigmatic, with children shifting directly from one model to another across time, as suggested in previous work (Opfer et al., 2011). More gradual shifts between models may better describe the development of number line placements, with phases in between during which more reference points are used, or even phases during which a mixture of reference point strategies can be used: It remains possible that children use different sets of reference points to place various numbers on a number line, making none of the models perfectly suited to their data. Previous discussions of a gradual versus an abrupt shift in representation have so far been inconclusive, and microgenetic studies are needed to address this issue in more detail (Barth & Paladino, 2011; Opfer et al., 2011). Item-based analyses could reveal item-specific differences in strategy use within and between children that cannot be investigated using only placements on the number line.
Shifts in the use of reference points, however, were prevalent in our data, confirming the hypothesis that children started using more reference points with increasing age and experience with numbers (Ashcraft & Moore, 2012; Rouder & Geary, 2014; Slusser et al., 2013). The frequent occurrence of logarithmic and non-cyclic power models in kindergarten suggests that kindergartners, although most scaled their responses to fit on the line, did not often use the endpoint of the number line as a reference point. Rather, kindergarteners seemed to scale their responses based on the beginning of the number line. A shift towards an increasing use of the endpoint as a reference point in making number line placements is suggested by the increasing number of children who were placed in the one-cycle power model throughout the study, indicating the use of two reference points (Rouder & Geary, 2014). The number of children whose number line placements were best fit by a linear model was also growing steadily until the end of grade 2. By the final measurement occasion (T6), the linear model was nearly just as prevalent as a best-fitting model as the one-cycle power model. These findings suggest that after the second year of primary school, the number of children whose number line estimates best fit a linear model at this scale will still increase until (almost) all children have achieved linear estimates.

The current data do not provide information on what underlies the shift between models in which various reference points are used. Shifts in number line placements may be the result of growing expertise in domain-specific numerical abilities, as suggested by the longitudinal associations between number line acuity and mathematics performance (also see: Siegler & Lortie-Forgues, 2014). This implies that children who use more reference points to make number line placements are more aware of the magnitude of numbers, the relations between numbers, and part-whole relations associated with numerical proportions displayed on a number line, in comparison to their peers who use fewer reference points. Alternatively, the use of more reference points may be the result of an increase in measurement skills.
(Cohen & Sarnecka, 2014). However, it is also possible that these shifts are the result of increasing domain-general capacities such as working memory (Friso-van den Bos et al., 2014; Geary, Hoard, Nugent, & Byrd-Craven, 2008). Integrative longitudinal studies are needed to compare the validity of the various predictors that have been proposed to underlie number line placements and identify the processes through which children shift between sets of reference points over time.

The observation that only very few children made number line placements that fit best with the two-cycle model is striking, because this model best fit the number line placements of children of a similar age and older children in a number of previous studies (Barth & Paladino, 2011; Rouder & Geary, 2014; Slusser et al., 2013). This difference in outcomes may be attributable to the fact that in previous studies during the practice phase children were explicitly instructed to place 50 in the middle of the number line, and not to place any other numbers exactly on that spot (Barth & Paladino, 2011; Rouder & Geary, 2014; Slusser et al., 2013). This may have motivated children to place values that should be placed close to the midpoint a bit further from the midpoint in these studies, while the lack of constraints with respect to placement on the midpoint may have elicited much closer placements to this specific point on the number line. This hypothesis is supported by the fact that in a study by Ashcraft and Moore (2012), in which the midpoint was also not stressed in the instructions, the two-cycle model was also the least representative of children’s number line placements. Perhaps this model is not of use when no instruction is given with respect to a reference point in the middle. This observation is consistent with the finding that number line acuity can be trained through number line-directed practice (e.g., Kucian et al., 2011; Siegler & Ramani, 2009).

An alternative explanation for the deviation in findings with previous studies (e.g., Barth & Paladino, 2011; Rouder & Geary, 2014; Slusser et al., 2013) and the current study is
that in all previous studies, children were taught in English, in which the number system is assumed to be more transparent than the Dutch number system. Dutch number words include the ones before the tens, instead of tens before ones (e.g., instead of saying thirty-five, one would say five-and-thirty), which is inconsistent with the order of written numerals. This may make it more difficult for young children to gain insight into the number system, and might explain the large number of children being placed in the random group during kindergarten, leading children to prevail in using less mature placement strategies and skipping the strategy with three reference points to inform number line placements in favour of the most advanced strategy, which is making linear placements. This hypothesis, however, rests on the assumption that children make placements through interpretation of verbal number words, either by transcoding the written number or by listening closely to the experimenter reading the numbers out loud. A study by Helmreich et al. (2011) indeed suggested that inversion errors, which are errors such as reading ‘53’ as ‘thirty-five’, may be of influence on number line placements in primary school children. More experimental studies are needed to investigate similar differences in findings and manipulate strategy use through variations in instruction in various groups.

Across time points, children generally moved from models with fewer reference points towards models with more reference points or linear models, as evidenced by the model transitions. This is consistent with the notion that models with more reference points are more advanced than models with fewer reference points (Ashcraft & Moore, 2012; Barth & Paladino, 2011; Rouder & Geary, 2014), and adds to the body of research by providing a more extensive set of models to index number line placements (Barth & Paladino, 2011; Rouder & Geary, 2014) using a longitudinal approach (Ashcraft & Moore, 2012). Children did not only maintain the same model or move towards more advanced models across time points, but small numbers of children regressed towards less advanced models from one time
point to the next. According to Siegler’s overlapping waves model, children do not abandon a
strategy entirely in favour of more advanced strategies, but have multiple strategies available
for solving any kind of problem. Gradually more advanced approaches become more
prevalent in children’s behaviour (Siegler, 1996). Regression towards less advanced models,
in this framework, may be considered adaptive, or at the very least, can be expected. It can
also not be ruled out that children use different strategies simultaneously, specific for each
item, and that this reduced the fit of certain models to index children’s placements.

More support for the notion that models with more reference points are indicative of
more advanced development of numerical abilities comes from the contrasts in mathematics
scores between children in different groups of number line placements: Although not all
contrasts were significant (some presumably due to a lack of power), a clear trend can be
seen in the pattern of children whose data fit more advanced models scoring higher on
mathematical performance. Importantly, children in the one-cycle group and linear group
scored higher than children in groups that were associated with the use of fewer reference
points, confirming that children who made placements in accordance with these models
indeed were more advanced with respect to number line placements, indicative of numerical
abilities associated with mathematical achievement (Dehaene, 2001; De Hevia & Spelke,
2009). This finding replicates earlier reports that children whose placements conform to
linear models score higher on mathematical tests (Ashcraft & Moore, 2012; Geary, 2011;
Halberda et al., 2008; Sasanguie et al., 2013; Siegler & Booth, 2004), and adds to the
understanding of this association by including multiple number line models. These findings
show that a more specific number of reference points can be associated with mathematics
performance, and not only the contrast between linear and pre-linear models.

The cross-lagged panel analyses addressing the interrelations between number line
acuity and mathematics performance yielded similar conclusions to those in the study by
LeFevre et al. (2013): The authors concluded that arithmetic performance predicted consecutive number line performance as much as number line performance predicted arithmetic performance. The current analyses were more extensive than the model presented by LeFevre and colleagues (2013), comparing a number of different models with twice as many occurrences and a more adequate sample size for this type of analysis. Word problems (as measured by the Cito Mathematics Test) and number line acuity showed bidirectional relationships, and a bidirectional model showed better fit than both the model with number lines predicting mathematics and the model with mathematics predicting number lines.

Compared to LeFevre et al. (2013), the current study included a more uniform group of children (all from the same grade), smaller intervals between time points (six months rather than a year), and a larger sample, making the data better suitable for path analysis, and included a direct comparison of various models with different theoretical implications. Therefore, the present study made a stronger case for the interplay between number line development and mathematical reasoning. Moreover, the current study compared mathematics scores between children placed in various categories of number line placement.

The rationalisation in the mutual interdependencies reported in the cross-lagged model lies not only in the notion that knowledge of the number system is needed for both tasks, but also in the current model’s implication that for a large part, mathematics performance enhances young school-aged children’s understanding of number. In other words: By performing calculations and reasoning about additions, subtractions, and other calculations, children gain insight into the ordinality of the exact number system and the relations between numbers, in addition to insights into number relations fostering insights into calculation processes.

The bidirectional relationship between mathematics and number line acuity may also be directly responsible for the sudden drop in random placements after the start of grade 1 (T3): Although the number of children showing random placements already decreased during
kindergarten, random placements were rare at the start of first grade. This may be a direct result of the structured mathematics education that is given from the start of first grade.

Although bidirectional relations between number line acuity and mathematics performance could be found throughout most of the first two years of formal education, number line acuity at T5 (middle of grade 2) or any other time point did not predict mathematics performance at T6 (end of grade 2). This apparent drop in predictive power may carry two explanations, which are not mutually exclusive. A first possible explanation is that mathematics performance at the end of grade 2 becomes more advanced, and requires the use of algorithms in which evaluation of mathematics problems on a number line is not required, making acuity on a number line task for a large part irrelevant for future – more advanced - mathematics performance. A second explanation may be that there is too little variation in number line acuity: explained variance of a linear slope approached 90% at the beginning of grade 2, and exceeds 90% at the end of grade 2 on this scale. Although this does not imply that variation between scores is irrelevant, it may not yield different outcomes for children, for example when they compare the likeliness of an obtained answer using number line estimation.

Also, mathematics at the start of grade 1 was directly predictive of mathematics performance at the start and end of grade 2. This may indicate that efficacious development of mathematical achievement at an early age is not only predictive of skills that are taught successively, but also have a direct impact on the more advanced skills that are taught later in education, for example through the use of retrieval strategies that are less cognitively demanding (Hecht, 2002). This would open up mental workspace, now no longer needed for basic calculations, to address a larger part of a more complex problem, and directly foster mathematical performance at a later age (Siegler, 1996; Van der Ven, Boom, Kroesbergen, &
Leseman, 2012). This issue, however, requires more thorough longitudinal investigation of the exact skills involved in making calculations and interpreting number and quantity.

**Conclusion and future directions**

Summing up, the current study provides deeper insight into the development and impact of number line acuity of children at the start of formal education. This study showed that children’s number line placements fit various power models, with the non-cyclic power model being more dominant in the lower grades and the one-cycle power model becoming more dominant over time, and that the group of children making linear placements becomes larger when children grow older. Also, mathematics performance is a predictor of number line acuity as well as vice versa. This may indicate that children do not only use their numerical abilities in learning to understand and solve mathematics problems (Xenidou-Dervou et al., 2014), but that they also, and maybe more importantly, develop more exact representations of number due to the practise with mathematical problems. This finding is not only of theoretical importance to knowledge development concerning numerical abilities, but can also be a motive for a more thorough investigation of how different types of mathematical problems best foster numerical abilities.

Future studies are needed to gain insight into the various aspects of number line placements. First, studies are needed to investigate the influence of instruction type on number line placements, and in particular, to what extent instruction with respect to number line placements around the midpoint influences the shape of the number lines produced by the children (Ashcraft & Moore, 2012). Second, although various studies have investigated transitions of number line shapes using number line tasks of various scales (Ashcraft & Moore, 2012; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Laski & Yu, 2014; Slusser et al., 2013), a broader range of models to describe the shape of number lines of various scales should be used, in order to gain insight into the development of number line
placements in other scales than in the present study. Third, although a developmental account has been made in the current study, no strong causal inferences can be drawn from the longitudinal analyses. Experimental studies are needed to investigate the causal associations between number line placements and mathematics performance, and to investigate development of number line shape in the face of quasi-experimental practice and feedback conditions, at the same time targeting shifts in number line placements (Barth & Paladino, 2011; Opfer et al., 2011). Moreover, strategy assessment in mathematics problem solving may provide insight into the various strategy choices made by children whose number line placements are indicative of different reference points and expand upon the current study by linking reference points to strategy choices in problems that centre around those reference points in numerical magnitudes. Finally, the models used in the current study carry the assumption that each pattern of placements can be associated with a varying set of reference points to make number line placements. Online assessment of number line placement strategies, such as eye tracking studies, could be used to confirm the use of reference points assumed to be associated with the various models of placement (Van Viersen et al., 2013). Nevertheless, the current study contributes to the body of knowledge concerning number line development by comparing different models of number line placements in a large-scale longitudinal study including children that were making important steps in their mathematical development throughout the course of the study. The results confirmed previously posed hypotheses as well as yielded new questions with respect to the role of number line acuity in mathematical achievement, and made important steps in uncovering the intriguing interplay between related, yet distinct skills.

References


1 Research project supported by NWO grant 411-07-110.