Journal of Cognitive Psychology

Routledge
Taylor & Francis Group

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

The development of numerosity estimation: Evidence for a linear number representation early in life

Janny C. Stapel, Sabine Hunnius, Harold Bekkering & Oliver Lindemann

To cite this article: Janny C. Stapel, Sabine Hunnius, Harold Bekkering & Oliver Lindemann (2015) The development of numerosity estimation: Evidence for a linear number representation early in life, Journal of Cognitive Psychology, 27:4, 400-412, DOI: 10.1080/20445911.2014.995668

To link to this article: https://doi.org/10.1080/20445911.2014.995668

Published online: 13 Jan 2015.

Journal of Cognitive Psychology, 2015
Vol. 27, No. 4, 400–412, http://dx.doi.org/10.1080/20445911.2014.995668

The development of numerosity estimation: Evidence for a linear number representation early in life

Janny C. Stapel[1], Sabine Hunnius[1], Harold Bekkering[1], and Oliver Lindemann[2]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands
[2]Department Psychologie, Universität Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

Several studies investigating the development of approximate number representations used the number-to-position task and reported evidence for a shift from a logarithmic to a linear representation of numerical magnitude with increasing age. However, this interpretation as well as the number-to-position method itself has been questioned recently. The current study tested 5- and 8-year-old children on a newly established numerosity production task to examine developmental changes in number representations and to test the idea of a representational shift. Modelling of the children's numerical estimations revealed that responses of the 8-year-old children approximate a simple positive linear relation between estimated and actual numbers. Interestingly, however, the estimations of the 5-year-old children were best described by a bilinear model reflecting a relatively accurate linear representation of small numbers and no apparent magnitude knowledge for large numbers. Taken together, our findings provide no support for a shift of mental representations from a logarithmic to a linear metric but rather suggest that the range of number words which are appropriately conceptualised and represented by linear analogue magnitude codes expands during development.

Keywords: Bilinear models; Development; Number cognition; Number representation; Numerical estimation.

From early in life, infants and children are exposed to many forms of magnitude information, as the physical world consists of objects that differ in size and amount. A symbolic number system allows for reasoning about magnitudes beyond what can be immediately seen. Despite the abundant exposure to magnitude information, understanding and applying symbolic numbers develops relatively late, namely starting at preschool ages (Lipton & Spelke, 2006; Sarnecka & Gelman, 2004). In a symbolic number system, the distance between consecutive numbers is equal, regardless of the size of the numbers. Behavioural research has shown that the cognitive representation of numbers appears to follow this rule in adults, as the relation between estimated numbers and actual numerals is linear in many estimation tasks (e.g., Izard & Dehaene, 2008; Lindemann & Tira, 2011; Siegler & Opfer, 2003). Interestingly, however, children younger than 12 years of age seem not to display such a linear relationship in estimation tasks and tend to overestimate the distance between small

Correspondence should be addressed to Janny C. Stapel, Donders Institute for Brain, Cognition and Behaviour, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: j.c.stapel@donders.ru.nl

We thank the schools and the parents and their children who participated in the study. Furthermore, we are grateful to Angela Khadar, Rianne Haartsen and Elke de Groot for their assistance.

This article was originally published with errors. This version has been corrected. Please see Erratum (http://dx.doi.org/10.1080/20445911.2015.1044224).

numbers and underestimate the distance between large numbers (Booth & Siegler, 2006; Laski & Siegler, 2007; Siegler & Booth, 2004;). That is, in contrast to older children and adults, the relation between estimated and actual numbers for young children can best be described by a logarithmic function. Hence, Siegler and Opfer (2003) proposed that during development, children's number representations shift from a logarithmic to a linear coding of numerical magnitude (see also Opfer & Siegler, 2007; Siegler & Booth, 2004). This shift from log to linear does not necessarily take place at a certain age, but is thought to depend also on the task characteristics, such as the tested number range, and contextual influences (Opfer & Siegler, 2007). In this way, children of a particular age may in some situations rely on linear number representations and in others on the less mature logarithmic representations. Recent studies have criticised the idea of a representational shift on methodological (Barth & Paladino, 2011; Huber, Moeller, & Nuerk, 2014; see also Moeller & Nuerk, 2011; Slusser, Santiago, & Barth, 2013) and theoretical grounds (Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009). The current study therefore aims to further investigate the development of number representations by employing a recently established task of numerosity production (see Lindemann & Tira, 2011; Mejias, Mussolin, Rousselle, Grégoire, & Noël, 2012).

The mental representation of numbers is thought to at least partially rely on an approximate system for numbers (Ashcraft, 1992; Cordes, Gelman, Gallistel, & Whalen, 2001; Moyer & Landauer, 1967). This system is assumed to be operational from birth and enables children to develop a sense of numbers (Dehaene, 1997) that is for instance involved in the comparison of numerosities or the approximation of the outcome of arithmetic problems (Berch, 2005; Halberda & Feigenson, 2008; Locuniak & Jordan, 2008). A classic paradigm to investigate the acuity of the number sense in children is the number-to-position task (see e.g., Siegler & Opfer, 2003), in which participants are required to make numerical estimations by indicating the position of a target number on an horizontally aligned line segment that represents a number line. The interval of the number line (e.g., from 1 to 100) is indicated by start and end point anchor numbers printed at the left and right of the line segment.

Several studies have employed this number-to-position task to test whether the estimations form a linear pattern with a fixed ratio between distance in space and distance in numbers throughout the whole number range (Booth & Siegler, 2006; Laski & Siegler, 2007; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003). For instance, Siegler and Booth (2004) tested 6-, 7- and 8-year-old children on a number line representing the interval from 0 to 100. The authors contrasted a logarithmic model with a linear model of the relation between the given numerals and the median response for that numeral per age group. The relation between actual and estimated numbers in the 6-year-olds was best described by a logarithmic function, whereas the response pattern of the 8-year-olds was best described by a linear function. For the 7-year-olds, both models explained the data equally well. The authors interpreted their findings as evidence that with increasing age the format of analogue number representations shifts from a logarithmic to a linear metric.

This interpretation has been recently criticised, because the observed changes in response patterns might be driven by other aspects in the cognitive development that are not necessarily related to numerical representations (Barth & Paladino, 2011; Huber et al., 2014). For instance, Barth and Paladino (2011) argue that the number-to-position task is not the best tool to investigate the potential representational shift, because this task inherently requires a judgement of the proportion of two line segments. Participants can use proportions that are familiar to them as anchors for their responses (e.g., 50 is halfway the scale, 100 is the end of the scale), which makes responses around anchor points more accurate. As children grow older and learn about proportions, they may start using more and more anchor points which might make their response pattern appear more linear. Consequently, improved performance on the number-to-position task does not necessarily reflect a shift from a logarithmic to a linear cognitive number representation with age, but might be caused by the improved abilities to judge proportions. To approach this methodological problem, Cohen and Blanc-Goldhammer (2011) suggested a modified version of the classic number line task. In their unbounded number line task, children are given a start number and a fixed unit (e.g., an indication of the distance from 0 to 1). Importantly, the end of the segment is not flagged with a number. Due to the absence of an end point anchor, proportion judgement strategies cannot be applied in this type of number line task. Recent developmental studies confirmed that whereas performance on the unbounded number line task did not reflect proportion judgements, the classic number-to-position task did (Link, Huber, Nuerk, & Moeller, 2013; Link, Nuerk, & Moeller, 2014). Taken together, it may be useful to adopt new tasks which do not include end point anchors to investigate the potential shift from log to linear in number representations.

Other scholars questioned the occurrence of a representational shift during development on theoretical grounds and proposed alternative models of the analogue number representations (Ebersbach et al., 2008; Moeller et al., 2009). For instance, Ebersbach et al. (2008) postulated that children start out with a bilinear representation, that is, with one linear representation for small numbers and a different, but also linear representation for large numbers. As the acuity of the representation of smaller numbers is higher than that of larger numbers (see e.g., Dehaene, 2009), the authors assumed that only for a small number range children are aware that a doubling of the numerical size requires a doubling of space on the number line. In line with this hypothesis, Ebersbach et al. (2008) reported the estimations in a number-to-position task of 5- to 9-year-old children to be best described by bilinear regression models and observed that the regression slopes for smaller numbers were substantially larger than the slopes for larger numbers.

Whereas this better fitting of bilinear models compared to logarithmic models suggests the existence of different cognitive codes for small and large numbers, the origin of these different types of number representations is still controversial. Ebersbach et al. (2008) argued, for instance, that the differences in the representations of small and large numbers are driven by the familiarity of the numbers: familiar numbers are represented more precisely, and with age, the range of familiar and thus well represented numbers increases. Alternatively, Moeller et al. (2009) proposed that the different representations for small and large numbers might be the result of the difficulty children have with two-digit numbers compared to single-digit numbers. As children grow older, they gain better insight into two-digit numbers, which leads to a linear representation for a larger number range. So far, the location and hence the underlying reason for having two distinct linear representations has not yet been directly investigated.

The current study therefore aims to further investigate the development of number representations. To control for the potential confound of developmental changes in proportion judgement abilities (Barth & Paladino, 2011), we employed a numerosity production task as an alternative method of approximate number estimations (Crollen, Castronovo, & Seron, 2011; Mejias & Schiltz, 2013; Mejias et al., 2012; Lindemann & Tira, 2011). That is, instead of using a number-to-position task, a numerosity production task was used to test children of 5 and 8 years of age. Participants were asked to produce a dot cloud of a specific numerosity on a screen by rotating a knob. Turning the knob clockwise increased the number of presented dots, whereas turning it counterclockwise reduced the numerosity. Studies on numerosity estimation (e.g., Izard & Dehaene, 2008) have shown that the number of dots can be estimated without referring to visual features such as covered area or contour length. Numerosity productions are therefore a promising tool to assess approximate numerosity representations. Two major advantages of this method over number-to-position judgements are that production tasks do not provide any visible anchor points at the upper end of the scale and do not require a transformation of numerical information into a spatial continuum (Lindemann & Tira, 2011).

Both aspects of the number-to-position task have been argued to be responsible for the observed developmental changes (cf. Barth & Paladino, 2011). We therefore argue that numerosity production provides a test of the metric properties of number representations not containing the biases discussed that will help to shed new light on the ongoing debate about the potential log-to-linear shift in number development.

Based on the controversy in the literature described above, we fitted four mathematical models to the numerosity production estimations of 5- and 8-year-old children to test the different accounts explaining the development of approximate number skills in children of these ages as reflected by the observed changes in the pattern of numerosity estimates (e.g., Ebersbach et al., 2008; Siegler & Booth, 2004): a logarithmic model, a simple linear model, a bilinear model with a free breakpoint and a bilinear model with a fixed breakpoint at 10 (see method section for more details about the bilinear models). Results of the modelling yield more insight in the development of number representations.

To be more specific, the logarithmic and the simple linear model were compared to test whether evidence for a change in the form of the representation between 5- and 8-year-old children can be observed in a numerosity production task. That is, based on the assumption of a representational shift (Opfer & Siegler, 2007; Siegler & Opfer, 2003), it should be expected that a logarithmic model provides a better model fit than a linear model for the 5-year-olds and vice versa for the 8-year-olds. The comparison between a logarithmic model on the one hand and bilinear models on the other hand was performed to shed light on the characteristics of the numerical representations in the younger children. That is, if a logarithmic function would be the best description for the estimations of the 5-year-olds, and a linear function the best description of the 8-year-olds' estimations, this would make a strong case for the notion that the underlying representations indeed shift from a logarithmic to a linear metric. However, if children have different linear representations for numbers in the different intervals (small vs. large numbers; Ebersbach et al., 2008), a bilinear model would provide a good description of the data. In the current study, we applied two different bilinear models to examine in greater detail what the potential underlying cause for distinct representations of small and large numbers might be. If two types of representations for small and large numbers are the result of the different complexities of single- to two-digit numbers, a bilinear model with a fixed breakpoint at 10 should be superior. In contrast, if the two representations stem from differences in familiarity between large and small numbers, a model with a free breakpoint, which varies depending on the child's experience with particular numbers, was expected to fit the data best.

METHOD

Participants

Two groups of children took part in the study, namely a group of 5-year-olds and a group of 8-year-olds. The group of 5-year-olds consisted of 35 children (M = 5.54 years, SD = 0.20 years, 19 females), and the group of 8-year-olds consisted of 36 children (M = 7.92 years, SD = 0.34 years; 16 females). One additional 5-year-old child was tested but not included in the data analysis because the child produced on each trial the maximum number of dots.

A part of the group was recruited via and tested in their school (51 children), and another part of the group was recruited via and tested in a developmental lab. In the Netherlands, all children attend primary school from 4 years of age. All children participated voluntarily with informed consent of their parents and received marbles as a reward at the end of the experiment.

Children tested in the developmental lab additionally received a children's book or 10 Euros for their participation, as this was the standard for all test sessions in the lab.

Apparatus

For the experiment, a knob (height: 33.4 mm, diameter: 53.8 mm, type: PowerMate, Griffin Technology, Nashville, USA) was used which was connected via a USB cable to a laptop. It was positioned in front of the child on a table. The knob could be rotated infinitely clockwise as well as counterclockwise with its rotation axis parallel to the Cartesian z-axis. Additionally, a standard computer mouse was connected to the laptop and controlled by the experimenter sitting next to the child.

A custom-made programme, based on the open-source Python library *Expyriment* (Krause & Lindemann, 2014) processed the input from the knob and controlled the stimulus presentation. By rotating the knob clockwise, the number of green dots (diameter 19 mm) presented on the screen increased depending on the rotational angle (see Figure 1). Rotations in the counterclockwise direction reduced the number of presented dots proportionally. The location of appearing or disappearing dots on the screen was rendered unpredictable (for more details about this method see Lindemann & Tira, 2011), as dots were displayed at a randomly chosen free position within a virtual circular area with a diameter of 14.6 cm centred on the middle of the screen. Distance between the dots was at least 4 mm. The maximum number of dots that could be produced was restricted to 100. When the numerosity production was finished and the experimenter pressed a mouse key, the dots moved to the right side of the screen and disappeared.

Task

For motivational purposes, the task was framed to participants as a task of helping a cartoon figure. A physical version of the cartoon figure was positioned right next to the screen. Children were told that the figure worked at a marble factory that day and he had to put the right amount of marbles in a bag. The child participant was asked to help him by making use of the knob. The experimenter showed that the knob could be rotated both ways and explained that rotation would make dots ("marbles") appear on the screen (see Figure 1). The participants were instructed to produce not too many nor too little dots on the screen, but were also told that answers did not have to be exact. This instruction was repeated after every block of five trials. After every block, a colour picture of a cartoon figure appeared, accompanied by a cheering sound. Two coloured containers were placed on the table, one next to the participant and the other next to the experimenter. The experimenter's container held four marbles at the start of the experiment, which were transferred to the container of the child as a reward, one for each time the cartoon figure appeared on the screen.

Materials

All stimuli were presented on a black background. Trials started with an Arabic digit at the screen centre in light-grey. The study comprised 20 experimental trials in which the following numerosities were to be produced: *6*, *9*, *11*, *12*, *14*, *15*, *16*, *17*, *20*, *21*, *24*, *25*, *27*, *28*, *32*, *35*, *39*, *42*, *46* and *49*. The choice of numbers allowed for inspection around the potential breakpoint of 10, as well as a precise estimation at the expected turning point of the logarithmic curve, and still sampled relatively evenly across the scale from 0 to 50. No information was provided about what the highest requested number of dots would be or about what the maximum number of dots was that could be produced. This prevented participants from adopting a fractioning strategy. Experimental trials were presented in randomised order. Before the experimental phase started, there were three instruction trials with the numbers *5*, *51* and *26* (performed by the experimenter) followed by three training trials with the numbers *5*, *14* and *43*, in which the child was given the opportunity to try out the knob and the task. The order of instruction and training trials was fixed. Training trials could be repeated in the case that child did not yet understand the task properly.

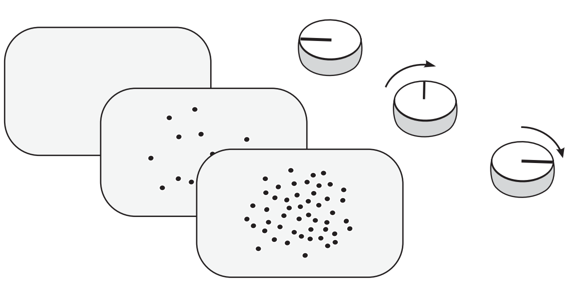

Figure 1. Schematic example of the task. Clockwise rotation of the knob led to an increased number of randomly presented dots on the screen. Counterclockwise rotation decreased the number of presented dots.

Procedure

Participants were seated in front of the laptop and received an oral explanation at the start of the experiment. Rotation of the knob and the result thereof was illustrated, namely the appearance and disappearance of "marbles" on the screen.

The experiment started with instruction trials performed by the experimenter to demonstrate the method. To make sure that a child would see the correct amount of dots being produced in the demonstration phase, these trials were programmed such that the experimenter would thus automatically produce the exact required amount of dots by turning the knob. A trial started with an Arabic digit displayed on the screen. The experimenter read the number aloud and clicked the mouse button to initiate the trial for the child. Now, the knob was operational and time of the judgement start was registered. No dots were presented at onset; hence each dot displayed was a result of knob rotation. When the child indicated to be finished with the numerosity production, the experimenter pressed the mouse button to register the estimate and the trial duration. The dots moved out of the screen in the direction of the physical cartoon figure. Children were asked whether they understood the procedure, and if necessary the demonstration could be repeated. Then three practice trials for the child were conducted, in which the experimenter made sure that knob handling was clear to the participant by explicitly asking the child to increase and decrease the amount of dots if the child did not do so spontaneously. Furthermore, the experimenter checked for task comprehension by asking the child after producing the dots: "Do you think you produced approximately the requested amount of dots?" Then four blocks of five trials were presented. In-between blocks, the participants received a marble in their container and were praised for doing their best. After finishing the experiment, participants were thanked and received marbles (all children), and additionally 10 Euros or a children's book (only those tested in the lab) for participation.

Modelling

The applied models are summarised in Table 1 and comprised a logarithmic model, a simple linear model, a bilinear model with a fixed breakpoint at 10 (Model A) and a bilinear model with a free breakpoint (Model B).

The segmented regression models consisted of two line segments, one for the small and one for the large number range. The slopes of the two line segments were assumed two be independent. The point at which the lines met (breakpoint) was modelled in two different ways. The first segmented model (Model A) predicted different types of representations for single- and two-digit numbers and had thus a fixed breakpoint (ψ) of 10. That is, Model A comprised three parameters (see also Table 1): a constant (c), a regression slope for small numbers (β_1) and a modulation of the slope for large numbers (β_2). In line with the notion that smaller numbers are better represented due to their higher familiarity, the second segmented model (Model B) did not make any presumptions about the location of breakpoints and allowed for interindividual differences in the transition of the two types of number representations. It included therefore an additional free parameter for the breakpoint (ψ).

Data analysis was performed using statistical computation software R 2.14 (R Development Core Team, 2011) and the R-package "Segmented" (Muggeo, 2008). The bilinear regressions with unknown breakpoints were fitted using the iterative parameter estimation procedure suggested by Muggeo (2003).

Analysis

To investigate the cognitive representations underlying the numerosity productions, we analysed the data using two different procedures (see Moeller & Nuerk, 2011 for a similar approach). First, we calculated for each age group the mean estimates of each number and fitted the resulting data to different regression models (see below). Second, the same regression models were fitted on the data of each individual separately. The main criteria for evaluating and comparing models are parsimony and model fit. A good model explains much of the variance in the data, but is at the same time as simple as possible. As our models differ in their complexity, we not only calculated the adjusted R^2 (Kyllonen, Lohman, & Woltz, 1984), but also the corrected Akaike information criterion (AIC_c; Akaike, 1974) which takes into account both the goodness of fit and the complexity of the model. Based on the AIC_c of the four models, we calculated furthermore for each model the Akaike weight of evidence (w_a), reflecting an estimate of the likelihood that a particular model is the best model in the set (Burnham & Anderson, 2004).

The second statistical approach to analyse the data consisted of a statistical comparison of the individual model fits. The analysis was analogue to the linear regression analyses of repeated measures data as proposed by Lorch and Myers (1990). To do so, we calculated the adjusted R^2 of the individual data fitting of all four models. Since R^2 values cannot be assumed to be normally distributed, an arcsine transformation was applied prior to the inference statistics (Studebaker, 1985; Bartlett, 1947). The arcsine transformed adjusted R^2 values were then compared using paired samples t-tests.

RESULTS

Only judgements that were given within 30 s entered the analysis (1.6% of all responses were excluded). We observed that some children had a tendency to explore the device by turning the knob to produce as many dots as possible (this was obviously fun and rewarding). In order to avoid a bias in the measurements of numerosity estimations by this impulsive behaviour, we excluded trials in which the maximum number of 100 dots (6.0%) was produced. Furthermore, trials were excluded in which no dots (0.3%) were produced.

As instructed, numerosity judgements were given very quickly. The mean judgement time for the young children (8.19 s, SD = 2.22 s) did not differ significantly from the response latencies of the older children (7.97 s, SD = 3.67 s; $t(69) < 1$, p = n.s.). The short average response latency of about 8 s in both age groups demonstrates that the estimations were given quickly and this renders it unlikely that the children have counted the dots to indicate their estimation.

Figure 2 depicts the mean estimation to each number. It becomes evident from this figure that the participants had in general a strong tendency to produce too many dots, even for small numbers. This is especially true for the 5-year-olds, who for instance produced as a median 15 dots when 6 were requested and around 30 dots when 10 were requested.

Model fittings for the 8-year-old participants

The results of the model fitting are summarised in Tables 1 and 2. For the 8-year-olds, all models explained the data very well. The across-subjects analysis showed that the relationship for this age group was almost linear (see Figure 2). Already the simplest linear model almost perfectly fitted the data and explained 98% of the variance. Comparing the goodness-of-fit indices of the linear, logarithmic and two segmented models revealed that the logarithmic model did not fit the data as well as the other three models (see Table 2). The lower AIC_c and the higher weight of evidence of the segmented models show that they describe the data even better than the linear model. However, improvement in explained variance is only minimal, despite the clear increase in parameters used (see Table 1). Both segmented models can also model a pure linear function, as then the beta weights become equal in size. Something similar seems to occur in the data of the 8-year-olds. Their estimations are thus approximately linear.

The analysis of the individual model fittings provides more indications that their estimations are approximately linear. The average explained variance of the individual linear models (mean adj. R^2 = .73) was larger than the variance explained by the individual logarithmic models (mean adj. R^2 = .70; $t(35)$ = 3.05, p < .001). Importantly, neither the segmented models A (mean adj. R^2 = .73) nor the segmented models B (mean adj. R^2 = .74) differed significantly from the linear models in the amount of variability they explained on the level of the individual (both t's(33) < 1).

Model fittings for the 5-year-old participants

As summarised in Table 1, there is a descriptive advantage of logarithmic fitting compared to the linear fitting for the group-level data of the 5-year-old participants. That is, whereas the logarithmic model explains 70% of the variance in the data, the linear model explains only 45%. Bilinear model A (R^2 = .69) performed comparable to the logarithmic model. In contrast, the bilinear model with a free breakpoint, Model B, explained considerably more variance in the data (R^2 = .91) than the other three models. Bilinear Model B proved

TABLE 1
Regression model equations and estimated parameters resulting from the fitting of the average produced numerosities in children from Age 5 and 8

			Model estimations	
	Model equation	*Parameter*	*Age 5*	*Age 8*
Linear model	$y = c + \beta x$	2	$\beta = 0.60$, $c = 29.17$	$\beta = 0.87$, $c = 4.09$
Logarithmic model	$y = c + \beta \ln(x)$	2	$\beta = 16.38$, $c = -6.09$	$\beta = 18.87$, $c = -32.19$
Segmented Model A	$y = \begin{cases} c+\beta_1 x, & \text{if } x < 10 \\ c+\beta_1 x+\beta_2(x-10), & \text{if } x > 10 \end{cases}$	3	$\beta_1 = 7.04$, $\beta_2 = 0.41$, $c = -30.74$	$\beta_1 = 1.89$, $\beta_2 = 0.84$, $c = -5.33$
Segmented Model B	$y = \begin{cases} c+\beta_1 x, & \text{if } x < \psi \\ c+\beta_1 x+\beta_2(x-\psi), & \text{if } x > \psi \end{cases}$	4	$\beta_1 = 3.73$, $\beta_2 = 0.14$, $\psi = 15.43$, $c = -10.34$	$\beta_1 = 1.54$, $\beta_2 = 0.83$, $\psi = 12.36$, $c = -3.18$

TABLE 2

Model fittings from four different regression models of the average produced numerosities. For each model, the Akaike information criterion (AIC_c), the Akaike weight of evidence (w_a), and the explained variance (R^2) are displayed

	Age 5			*Age 8*		
	AIC_c	w_a	R^2	AIC_c	w_a	R^2
Linear model	86.87	<.001	.45	22.00	.06	.98
Logarithmic model	74.43	<.001	.70	44.66	<.001	.93
Segmented Model A	76.78	<.001	.69	17.60	.58	.98
Segmented Model B	54.68	.99	.91	18.63	.34	.99

to be superior in describing the data pattern of the 5-year-olds when comparing the goodness-of-fit. The segmented model with a free breakpoint yielded clearly the best model fit, as the weight of evidence was by far the highest of the four models (see Table 2).

Similar results were found in the analysis of the individual model fittings. The logarithmic models explained more variance (mean adj. $R^2 = .28$) than the individual linear models (mean adj. $R^2 = .20$; $t(34) = 7.45, p < .001$). The variances in the data of the individual 5-year-olds were equally well accounted for by the segmented models with fixed breakpoint of 10 (mean adj. $R^2 = .28$) as the logarithmic models, $t(33) < 1$. Interestingly, however, segmented models with a free breakpoint (mean adj. $R^2 = .42$) explained the variability in the individuals best compared to the linear models, $t(34) = 5.31$, $p < .001$, the logarithmic models $t(34) = 3.10$, $p < .001$, as well as the segmented models of type A, $t(34) = 4.47$, $p < .001$.

Comparison of the model estimates of the 5- and the 8-year-old participants

When applying a linear model on the group averaged data, the estimated slope of the 5-year-olds was

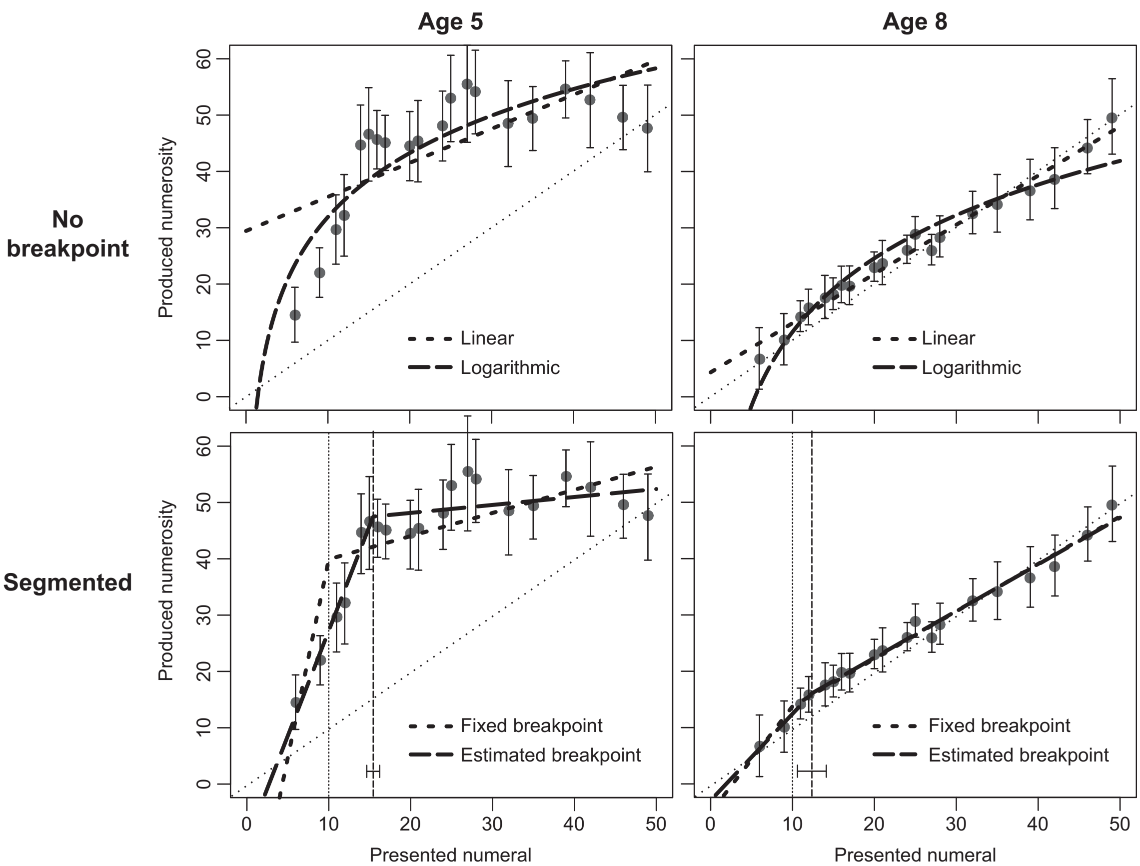

Figure 2. Mean produced numerosities plotted against the presented numerals. Panels on the left represent the data of the 5-year-olds, panels on the right represent data of the 8-year-olds. The upper panels display the logarithmic and linear curves fitted on the group-averaged data. The lower panels display the segmented models fitted on the group-averaged data. Error bars were constructed based on the method suggested by Cousineau (2005) accommodating a within-subjects design.

0.60 and of the 8-year-olds 0.87, both indicating an underproduction. However, for the bilinear models, the model estimates pointed to overproduction in the low numbers, and more so for the 5-year-olds (Model A: 7.04, Model B: 3.73) than the 8-years-olds (Model A: 1.89, Model B: 1.54). Overproduction has been found to be larger in numerosity production tasks than in numerosity perception tasks (Ebersbach & Erz, 2014; Izard & Dehaene, 2008), which may explain the relatively high overproduction in the current study. Naturally, model estimates for the larger numbers showed the opposite pattern (5-year-olds: Model A: 0.41, Model B: 0.14; 8-year-olds: Model A: 0.84, Model B: 0.83). As the breakpoint is higher for the 5-year-olds (15.43) than for the 8-year-olds (12.36), it may be more informative to look at the modelling results of the individual data.

The slope estimates of Model B fitting the individual data are plotted in Figures 3 and 4. From these figures, it can be seen that the estimated slopes of the 8-year-old children varied closely around 1 for the small (median slope 0.9) and large number range (median slope 1.01). This shows that estimations were similarly precise and linear for the whole number range. The estimated slopes for the small numerosities of the 5-year-olds were however clearly larger than 1 (median slope 4.01). This indicates that they overproduced in response to the small numbers, and

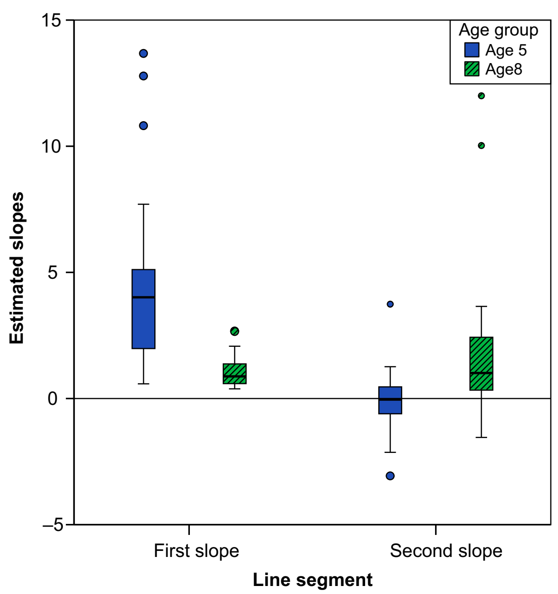

Figure 3. Boxplots displaying the distribution of estimated slopes of the first and second line segment for the 5- and 8-year-olds.

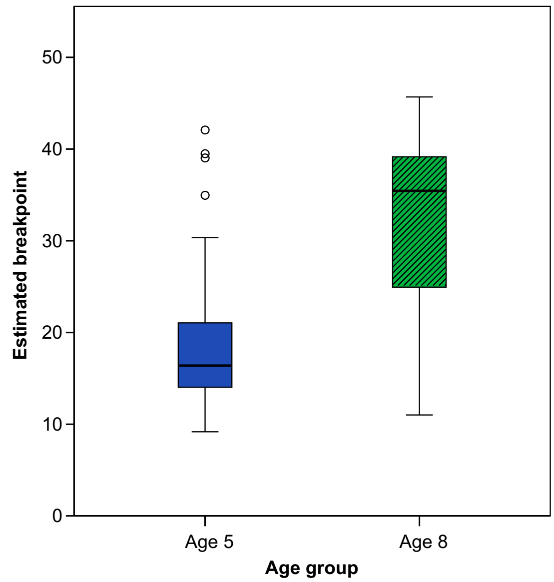

Figure 4. Boxplots displaying the distribution of estimated breakpoints for the 5- and 8-year-olds.

that they were less precise compared to the 8-year-olds. Interestingly, the analysis of the second slopes of the 5-year-olds revealed that these slopes for large numerosities were on average almost zero (median slope −0.03), indicating that these estimations were fully uncorrelated with the larger target numbers. This pattern suggests that the 5-year-old children merely guessed the to-be produced numerosities in the larger number range.

The breakpoints estimated from the individual models are displayed in Figure 4. As can be seen from the figure, the breakpoints of the 5-year-olds (median breakpoint = 16.40) are in general lower than the breakpoints of the 8-year-olds (median breakpoint = 35.45).

DISCUSSION

The present study examined the development of the cognitive representation of numerical magnitudes between 5 and 8 years of age by applying a recently developed numerosity production task. We compared four regression models reflecting different theoretical ideas about the nature of the development of number representations. The results yielded three main findings. First, in line with previous studies (Booth & Siegler, 2006; Laski & Siegler, 2007; Siegler & Booth, 2004), a linear model described the estimations of the 8-year-olds most accurately, whereas a bilinear model best explained the data of the 5-year-olds.

Second, the bilinear model outperformed the logarithmic model in accounting for the numerosity production responses of the 5-year-olds. This finding is not in line with the assumption of a representational shift from log-to-linear (e.g., Siegler & Opfer, 2003). The results rather indicate an all-or-nothing pattern in which small numbers are linearly represented in contrast to a poor or almost absent approximate representation of large numbers. Third, we demonstrated that the modelled breakpoint reflects the point up to which the children represented the distances between small numbers accurately. Our analyses showed that this point is not necessarily at the transition of single- to two-digit numbers but instead provided evidence that the range of accurately represented numbers increases with age.

A shift in magnitude representations?

The central question of the current study was whether children's magnitude representations undergo a shift from logarithmic to linear throughout development. Thus far, this potential representational shift has mainly been tested using the number-to-position task (Ebersbach et al., 2008; Moeller et al., 2009; Opfer & Siegler, 2007; Siegler & Opfer, 2003). Crucially, the number-to-position task involves the presentation of a visible start and end point anchor, and with increasing age, children have increasing knowledge of proportions and may apply this knowledge to provide additional anchors. The more anchors are used, the more linear the data pattern. As a result, increased knowledge of proportions will lead to a more linear response pattern (Link et al., 2013, 2014), obscuring whether representations indeed shift from log-to-linear or not (Barth & Paladino, 2011). If the previously observed changes in the response patterns are indeed a reflection of a log-to-linear shift in the underlying cognitive number representations, we expect to find a similar change in the response pattern in alternative methods of approximate number estimations. The current results from the numerosity production task indeed replicated previous findings, as the estimations for the younger age group displayed a clearly different pattern compared to the one of the older age group. The statistical modelling, however, revealed that the shape of the estimation curve for the younger participants can more accurately be described with a bilinear model rather than with a logarithmic model. Model fits were best for the bilinear model with a free breakpoint, which provides support for this segmented-line model from a statistical perspective. At a descriptive level, the data provided even more evidence for the notion that the 5-year-olds represented large numbers differently than small numbers. The regression slopes for the small numbers were substantially greater than the slopes for the large numbers, indicating an abrupt switch in the type of representation for small and large numbers in the group of younger children. The relation between verbally presented numbers and estimations was not only *different* for the small and large range, for large numbers the relation was even *absent*. This response pattern indicates that children had an approximate sense for numbers only up to a certain point. Beyond that number range, children seemed to fail to conceptualise the number words and merely guessed the response by producing an arbitrary numerosity that exceeded their number knowledge.

Our finding of an all-or-nothing pattern of representations is in line with previous findings reported by Le Corre and Carey (2007). These authors showed that 3- to 5-year-old children produced increasing estimations for rising numbers up until the point for which they understood the meaning of the number as assessed by another task. For numbers outranging their number sense, no correlation between estimated and actual numerals was found. Also the results of Ebersbach et al. (2008) are compatible with our observations, in the sense that larger slopes were found for the lower numbers and smaller slopes for the higher numbers when fitting bilinear models to the estimation data. However, in contrast to our results, the slopes for the higher numbers were different from zero. The potential reason for this positive correlation between estimations and numbers in the large range can be found in the demands of the number-to-position task. In contrast to numerosity productions, this task allows for anchoring not only at the lower end, but also at the upper end of the number range (cf. Cohen & Blanc-Goldhammer, 2011; Moeller & Nuerk, 2011). As a result, the presence of a high end anchor artificially affects the precision of the estimations of larger numbers and might in this way account for the positive correlations in the larger number range in this task. Interestingly, the analysis of Ebersbach et al. (2008) on the variability of the judgements in the number-to-position task across the number range provided direct empirical support for the explanation that children made use of high-end anchors. If no anchors are present as in the numerosity production task reported here, estimations of large numbers in young children may fail to show any systematic relation with the numerical size, supporting the notion of the complete absence of approximate knowledge for numbers exceeding a certain point.

Different representations for small and large numbers

Given that a bilinear model proved to be a better description of the data at hand, the question arises what might be the underlying cause for a difference in the estimations of small and large numbers. To answer this question looking at the location of the breakpoint can be informative. Moeller et al. (2009) proposed to use a breakpoint fixed at 10. They argued that the differences in representations for large and small numbers probably stem from difficulties integrating knowledge of single- and two-digit numbers. If the difference in representations between small and large numbers originates in the conversion from single- to two-digit numbers, one would expect that all participants have 10 as breakpoint. However, the outcome of the breakpoint modelling does not provide support for this notion of Moeller et al. (2009). Participants differed in their breakpoints, and also the breakpoint distribution of the young children revealed clearly an average estimated breakpoint larger than10 (median breakpoint of the 5-year-olds: 15.43). Our results are thus in conflict with the interpretation that the different response pattern along the number range originated from the difficulty to integrate multiple digits into one holistic number concept. Interestingly, Ebersbach et al. (2008), who also described higher breakpoints for children older than 5 years (starting at 20), found a strong correlation between the modelled breakpoints of the individuals and the decade to which participants could count. These findings support our notion that modelled breakpoints in bilinear regressions represent a diagnostic measure of the point to which numbers are well represented by the child.

Modelled breakpoints should however be interpreted with some care. As the model restrictions can force a breakpoint on the data of every participant, one may question the validity of the modelled breakpoint in some cases. In the tested 8-year-olds, there was a large spread in the distribution of the breakpoints. At the same time, the difference in slopes for small and large numerosities was small, and no difference was found in explained variance for bilinear versus linear models in the individual data. Together, this indicates that the modelled breakpoints for the 8-year-olds might have been arbitrary. Possibly, the "true" breakpoint of the 8-year-olds lay outside the range tested in the study, or even stronger, potentially the 8-year-olds had a linear representation of even precision throughout any given range.

Development of the range of linearly represented numbers

The numerosity production judgements revealed that the increase of number acuity with age was characterised by an upwards shift of the range of numbers represented by approximate number codes. That is, the linear representation of numbers of 8-year-old children appeared to cover a larger range of numbers than the numerical system of 5-year-old children. Possibly, numerical development is a continuous learning process and the range of adequately represented numbers increases gradually. Alternately, the individual number development might be abrupt and driven by the child's insight in the consistency of the number system, which allows her to generalise numerical knowledge to a large range of numbers. The data of the present study cannot distinguish between these two accounts and we believe that future longitudinal studies are required to address the developmental changes.

Conclusion

The present study substantially extends previous work on numerosity productions (Crollen et al., 2011; Lindemann & Tira, 2011; Mejias & Schiltz, 2013; Mejias et al., 2012) and demonstrates that this method provides a promising new approach to examine the acuity for small and large numbers in early childhood. The statistical modelling of the data of the 8-year-olds not only showed considerable overlap with previous studies (Ebersbach et al., 2008; Opfer & Siegler, 2007; Siegler & Opfer, 2003) but revealed furthermore new evidence for developmental differences in the representation of large numbers. We propose that the numerosity production task might be considered as a method for an assessment of the interface between the approximate number system and number symbols and an alternative method that complements the well-established number-to-position task to trace the numerical development.

The results of the present study revealed no support for a shift of mental number representations from a logarithmic to a linear metric and rather suggest that numerical representations of children follow an all-or-nothing pattern in which only number words in a certain range are appropriately conceptualised and represented by analogue linear magnitude codes.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. doi:10.1109/TAC.1974.1100705

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*(1–2), 75–106. doi:10.1016/0010-0277(92)90051-I

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*(1), 125–135. doi:10.1111/j.1467-7687.2010.00962.x

Bartlett, M. S. (1947). The use of transformations. *Biometrics*, *3*(1), 39–52.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, *38*, 333–339. doi:10.1177/00222194050380040901

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*, 189–201. doi:10.1037/0012-1649.41.6.189

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261–304. doi:10.1177/0049124104268644

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, *18*, 331–338. doi:10.3758/s13423-011-0059-z

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707. doi:10.3758/BF03196206

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*, 4–45.

Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation. *Experimental Psychology*, *58*(1), 39–49. doi:10.1027/1618-3169/a000064

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

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, *1156*, 232–259. doi:10.1111/j.1749-6632.2009.04469.x

Ebersbach, M., & Erz, P. (2014). Symbolic versus non-symbolic magnitude estimations among children and adults. *Journal of Experimental Child Psychology*, *128*, 52–68. doi:10.1016/j.jecp.2014.06.005

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *Journal of Experimental Child Psychology*, *99*(1), 1–17. doi:10.1016/j.jecp.2007.08.006

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465.

Huber, S., Moeller, K., & Nuerk, H.-C. (2014). Dissociating number line estimations from underlying numerical representations. *The Quarterly Journal of Experimental Psychology*, *67*, 991–1003. doi:10.1080/17470218.2013.838974

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247. doi:10.1016/j.cognition.2007.06.004

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, *76*, 1325–1345. doi:10.1037/0022-0663.76.6.1325

Krause, F., & Lindemann, O. (2014). Expyriment: A Python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*, 416–428. doi:10.3758/s13428-013-0390-6

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*, 1723–1743. doi:10.1111/j.1467-8624.2007.01087.x

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*, 395–438. doi:10.1016/j.cognition.2006.10.005

Lindemann, O., & Tira, M. D. (2011). Operational momentum in numerosity production judgments of multi-digit number problems. *Zeitschrift für Psychologie/Journal of Psychology*, *219*, 50–57.

Link, T., Huber, S., Nuerk, H.-C., & Moeller, K. (2013). Unbounding the mental number line – new evidence on children's spatial representation of numbers. *Frontiers in Psychology*, *4*.

Link, T., Nuerk, H.-C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology*, *67*, 1597–1613. doi:10.1080/17470218.2014.892517

Lipton, J. S., & Spelke, E. S. (2006). Preschool children master the logic of number word meanings. *Cognition*, *98*, B57–B66. doi:10.1016/j.cognition.2004.09.013

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, *41*, 451–459. doi:10.1177/0022219408321126

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*(1), 149–157. doi:10.1037/0278-7393.16.1.149

Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., & Noël, M.-P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychology*, *18*, 550–575. doi:10.1080/09297049.2011.625355

Mejias, S., & Schiltz, C. (2013). Estimation abilities of large numerosities in Kindergartners. *Frontiers in Psychology*, *4*, 1–12.

Moeller, K., & Nuerk, H.-C. (2011). Psychophysics of numerical representation: Why seemingly logarithmic representations may rather be multi-linear. *Zeitschrift für Psychologie/Journal of Psychology*, *219*, 64–70.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, *103*, 503–515. doi:10.1016/j.jecp.2009.02.006

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520. doi:10.1038/2151519a0

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*, 3055–3071. doi:10.1002/sim.1545

Muggeo, V. M. R. (2008). Segmented: An R package to fit regression models with broken-line relationships. *R News*, *8*(1), 20–25.

Opfer, J., & Siegler, R. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169–195. doi:10.1016/j.cogpsych.2006.09.002

R Development Core Team. (2011). R version 2.14. Vienna: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Sarnecka, B. W., & Gelman, S. A. (2004). Six does not just mean a lot: Preschoolers see number words as specific. *Cognition*, *92*, 329–352. doi:10.1016/j.cognition.2003.10.001

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–250. doi:10.1111/1467-9280.02438

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*, 193–208. doi:10.1037/a0028560

Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research*, *28*, 455–462.