A longitudinal analysis of number estimation, counting skills and mathematical ability across the first school year.

Kevin Muldoon[a]

John Towse[b]

Victoria Simms[c]

Oliver Perra[d]

Victoria Menzies[e]

Author note:

[a]Kevin Muldoon, School of Life Sciences, Heriot Watt University, Edinburgh, Scotland; [b]John Towse, Department of Psychology, Lancaster University, Lancaster, England; [c]Victoria Simms, Department of Health Sciences, University of Leicester, Leicester, England; [d]Oliver Perra, School of Sociology, Social Policy and Social Work, Queen's University Belfast, Belfast, Northern Ireland; [e]Victoria Menzies, Institute for Effective Education, University of York, Heslington, England.

This research was support by a grant from the Economic and Social Research Council, UK, grant number RES-062-23-0970

Correspondence concerning this article should be addressed to Kevin Muldoon, School of Life Sciences, Heriot Watt University, Edinburgh, Scotland, EH14 4AS. Email: k.muldoon@hw.ac.uk

Abstract

In response to claims that the quality (and in particular linearity) of children's mental representation of number acts as a constraint on number development, we carried out a longitudinal assessment of the relationships between number line estimation, counting, and mathematical abilities. Ninety-nine five-year-olds were tested on four occasions at three-monthly intervals. Correlations between the three types of ability were evident, but while the quality of children's estimations changed over time, and performance on the mathematical tasks improved over the same period, changes in one were not associated with changes in the other. In contrast to the earlier claims that the linearity of number representation is potentially a unique contributor to children's mathematical development, the data suggest that this variable is not significantly privileged in its impact over and above simple procedural number skills. We propose that both early arithmetic success and estimating skill are bound closely to developments in counting ability.

Introduction

The importance of understanding the development of mathematical cognition has long been recognized, both for the appreciation of the emergence of numerical skills and what it can tell us about cognitive development more generally. It is commonly accepted that awareness of computational processes alone cannot account for individual differences in numeracy, with much to be gained from insights about the detailed nature of the underlying conceptual representations. In pursuit of these insights, the increasingly influential number-line estimation paradigm may offer crucial insights into the link between mental representations of number and numerical and mathematical development. In the present paper, we show that a longitudinal perspective on estimation and mathematics can offer valuable evidence that helps address a series of issues, including the developmental association between number line estimation and early arithmetic.

Age-related changes in children's mental representation of number

On number-line estimation tasks children are asked to indicate the position of a target number on a blank number line (Booth & Siegler, 2006; Siegler, Thompson & Opfer, 2009; Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010). By comparing the *actual* position of the target number to the position of the child's *estimate*, researchers have begun to develop detailed models of how numbers are represented in the mind and what developmental changes take place in numerical scaling of these representations.

Unsurprisingly, older children's estimates are more *accurate* than those made by younger children (Booth & Siegler, 2006, 2008; Siegler & Booth, 2004). More interesting is the finding that the *distribution* of estimates conforms to a particular pattern of change, where the resulting estimation plots appear to shift from a logarithmic to linear profile as children get older (Booth & Siegler, 2006, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003). The logarithmic-to-linear shift has been reported to take place among children who are between 5 and 8 years old (Booth & Siegler, 2008). Moreover, this shift is not only age-related but also scale-dependent; estimations of number initially become more accurate and linear on smaller number line scales (e.g., 0-10), only later becoming more accurate and linear on larger scales (e.g., 0-100). Thus, children may behave as if a particular number lies either on a logarithmic or linear scale as a function of their age and the reference points that they refer to when making a judgement.

Mental Representation of Number as a Constraint on Mathematical Ability

Research has demonstrated that the linearity of numerical estimations correlates with standardised arithmetical tasks (Booth & Siegler, 2006; Muldoon, Simms, Towse, Menzies & Yue, 2011; Siegler & Booth, 2004) and simple addition tasks (Siegler & Mu, 2008). Furthermore, a causal link is supported by the findings of Siegler and Ramani (2009); children who played a simple board game where tokens were moved along a line of evenly spaced squares following the roll of a dice not only began to produce more linear estimations but also displayed improved arithmetic problem solving. The data converge on the idea that the number-line estimation task has the potential to be linked with key developmental issues, and is not just an abstract, laboratory paradigm.

Evidence against close developmental ties between estimation and math ability has nonetheless recently emerged in a cross-cultural examination of estimation performance. Muldoon et al. (2011) matched UK and Chinese children on their math ability (as measured by scores on a standardized math test drawn from the British Ability Scales) by removing the lowest-scoring Chinese children. Crucially, removing the poorest math performers failed to produce significant changes in the linearity or accuracy of the remaining Chinese children's estimations, providing evidence that the two skills are independent. Of course, manipulating the sampling of children to identify developmental relationships is not the same as gathering longitudinal data, and so to address a number of key developmental issues we subsequently tested the children from the UK on a further three occasions across the school year and report our findings here.

Both estimating and mathematical ability are yoked to counting proficiency

Previous longitudinal studies have found developmental relationships between math ability and procedural counting (e.g., Jordan, Kaplan, Locuniak & Ramineni, 2007). Thus, progress in math ability appears linked to both estimating and counting proficiency. Two hypotheses emerge from this conclusion. The first is that children whose estimations display the greater linearity will outperform other children of the same age on mathematical tasks. The second is that the scale-dependency observed in estimating studies results from children drawing upon a linear mental representation when the number to be estimated lies within their counting range but referring to an intuitive, logarithmic representation for numbers beyond this range (Berteletti et al., 2010). This latter hypothesis has already been investigated by some (Ebersbach , Luwel, Frick, Onghena, & Verschaffel ,2008; Nuerk, Kaufman, Zoppoth, & Willmes, 2004) who maintain that the logarithmic profile observed in younger children's estimations is really two separate linear functions, one (with a steep slope) for smaller numbers and another (with a slope close to zero) for larger numbers. According to this 'bi-linear' model of estimating, the position of the 'break-point' (i.e., the point at which the two slopes intersect) is related to the number range with which children are familiar (Moeller, Pixner, Kaufmann, & Nuerk, 2009).

However, whilst there are important differences in theoretical accounts of the function(s) explaining estimation profiles, the picture to date is that (a) estimations become more accurate and linear with age; (b) estimation performance is linked in some way to level of mathematical attainment; and (c) both estimating and math performance are yoked to counting ability. We use a longitudinal analysis to address the relationships between these three number domains.

Aims of the Present Study

The present longitudinal study set out to address three key issues. First, we measured changes in performance in estimation and math performance to identify the extent to which these skills change over the period of one school year and to examine the hypothesis that *changes* in one are associated with *changes* in the other. Thus far, only cross-sectional data have been used to consider math performance and age-related shifts from logarithmic to linear representations of number (e.g., Booth & Siegler, 2006; 2008). In using a longitudinal design, the data have the potential to reveal any similarities or differences in the trajectories of these two skills, and we use growth curve modelling to test the hypothesis that the two rates of change are linked.

Secondly, we investigate the degree to which number line estimations provide a unique predictor of math ability, contrasting this with an alternative account that focuses on counting knowledge as a reliable predictor of arithmetical development. In a longitudinal analysis, kindergartner's counting proficiency predicted the acquisition of basic arithmetical skills the following year (Aunio & Niemivirta, 2010). It is possible, of course, that counting ability and number line estimation skill both draw on a common representation of number. Using regression analysis we will assess whether counting and estimation skills make independent contributions to the variance in math ability.

Thirdly, we test the claim that only numbers within a child's counting range generates linear estimations on a number line (Ebersbach et al., 2008). If children's counting proficiency influences the way they space numbers on a mental number line, then children who can, for example, count to 10 (or 20) should produce better estimations than children who cannot count to 10 (or 20); at least on 0-10 and 0-20 scales. Although preschoolers are likely to be familiar with numbers up to 10, it is rarely assumed that they will be competent with numbers above 10 when they begin school. Whilst some familiarity with the decade structure up to 100 is typically a goal for children in Year 1, more effort is put into introducing children to the numbers 11 to 20. Consequently, the inclusion of a 0-20 scale has the potential to capture changes in accuracy and linearity when estimating that might be 'hidden' on the 0-100 scale due to the compression effect evident on larger scales. We also address the question of whether logarithmic estimation profiles are really bi-linear, with the two separate linear functions associated with numbers within and beyond children's counting range.

Method

Participants

Ninety-nine children were recruited through four primary schools in Edinburgh, Scotland for this and another study that compared children from the UK with children from China (see Muldoon et al., 2011). The mean age at Time 1 was 64 months (range 57-70 months, 49 males, 50 females). All children spoke English as their first language and were taught under the same guidelines (the Scottish 'Curriculum for Excellence').

Procedure

Two female researchers visited schools on four occasions throughout the school year. The first testing session was in September (Time 1) with others following at 10-week intervals. Number line estimation measures were taken on all four visits, while counting and math ability measures were taken only at Times 1, 2 and 4 due to constraints imposed by the school calendar.

Children were tested individually in a quiet part of the school building, or at the back of the classroom and completed the following tasks in random order:

[1] Number estimation tasks. Three number lines were used: 0-10, 0-20 and 0-100. Each line was 25 cm long, with 0 on the left end and either, 10, 20, or 100 on the right end, depending on the scale. Children were shown where the middle number on each scale (i.e., 5, 10 or 50) was positioned by the experimenter. Children were told that their task was to mark where they thought a specific number should be positioned on the number line with their pencil (these target numbers were positioned centrally, approximately 10 cm above the line, and verbalised by the researcher). The target numbers to be estimated were 1, 2, 3, 4, 6, 7, 8, and 9 (on the 0-10 line); on the 0-20 line were 3, 4, 6, 8, 12, 14 and 17 (on the 0-20 line) and 3, 4, 6, 8, 12, 14, 17, 18, 21, 24, 25, 29, 33, 39, 42, 48, 52, 57, 61, 72, 79, 81, 84, 90 and 96 (on the 0-100 line). The presentation of the number lines was in the same order for all children, with the 0-10 line presented first, followed by the 0-20 line, and finally the 0-100 number line; however the presentation of specific trials was randomised.

[2] Counting tasks. Children's counting range can be assessed in at least two ways. One approach is to ask them to simply 'count as high' as they can (i.e., by reciting the number string). Accordingly, children were asked to count as high they could, unaided, in increments of one (hereafter referred to as 'reciting'). Once the child made a mistake in their counting sequence the experimenter asked the child to stop and noted the number that they had reached. A second index of counting is a child's ability to enumerate sets of items. Children were asked to count a row of 10 dots and a separate row of 20 dots. Children were recorded

as being successful or unsuccessful in each case (i.e., they had to count all 10/20 dots without making an error; hereafter referred to as ‘enumerating’).

[3] Math ability tasks. Math ability was measured using the Early Number Concepts component of the British Ability Scales. This set of tests is designed for children between 2:6 and 7:11 years of age. For reference purposes, the maths test comprises 24 items that assess various basic numerical competencies:

1. *Recognizing number names and numerals* (e.g., “Point to the person who has 3 boxes”) – 4 items.
2. *Identifying quantitative relationships (less than/more than/same as)* (e.g., “Two of these people have the same number of boxes. Point to those two people”) – 5 items.
3. *Matching corresponding magnitudes* (e.g., [Gesturing to a mixed array of small and large spades] “Show me the buckets that go with the little spades”) – 5 items.
4. *Matching sets of discontinuous quantity* (e.g., [Pointing to the yellow ladybird with four dots] “Find the red ladybirds that go with this one.”) – 4 items.
5. *Solving basic addition and subtraction word problems* (e.g., “John [shown clutching balloons in each hand] is going to give all his balloons to Lisa. How many will she have then?”) – 6 items.

A raw score represents the number of correctly answered questions, which can then be converted into an ability score (this also accounts for question difficulty). For all analyses we used the ability score.

Due to constraints in access to schools and the children attending them, it was not feasible to collect data on math ability at all four time points (only time points 1, 2, and 4). Since there was nothing in the estimation data to suggest that time point 3 shows unusual performance, we do not feel this places a major constraint over the ability to analyse the relationship between number estimation and other numerical tasks.

Derivation of performance

We analysed the estimation data for changes in both linearity and accuracy (denoted by error). To establish the trend in children's estimations both linear and logarithmic functions (R^2_{lin} and R^2_{log}) were fitted using the equations '$y = slopex + b$' (linear function) and '$y = clnx + b$' (logarithmic function). Error was calculated following the method used by Siegler and Booth (2004):

Error = [(child's estimate- number to be estimated)/number line scale] x 100

For example, if the number to be estimated on the 0-10 number line was 6, but the child made a mark at the position that corresponds to 8, error would be calculated at 20 % for this number [i.e., (8-6)/10 x 100 = 20%].

Results

The study yielded a very large set of data and it is not feasible to describe in detail all possible analytic outcomes without compromising clarity. Accordingly, we focus on the three key issues in order.

Key issue 1: Developmental associations between changes in children's estimating and math ability

A 4 x 3 repeated measures ANOVA with Time (Time 1 to Time 4) and scale (Number Line; 0-10, 0-20 and 0-100) as factors and the fit of the linear function (R^2_{lin}) as the dependent measure shows that linear fits improved over time, $F(3, 276) = 16.87$, $p < .001$, $\eta_p^2 = .16$. There was also a main effect of Number Line, $F(2, 184) = 454.28$, $p < .001$, $\eta_p^2 = .83$, with a less linear performance on the 0-10 relative to the 0-100 scale, F(1, 92) = 508.87, p<.001, η_p^2 = .85, and also a less linear performance on the 0-20 than the 0-100 scales, F(1, 92) = 696.79, p < .001, η_p^2 = .88. There was also a significant interaction between Time and Number Line, F(6, 588) = 4.19, p < .001, η_p^2 = .04. Whilst there were no significant changes in R^2_{lin} values on the 0-10 scale over testing epoch, F(3, 294) = 2.24, p > .05, η_p^2 = .02, estimations became increasingly linear on the 0-20 line, F(3, 294) = 8.20, p < .001, η_p^2 = .08. However, there were no significant improvements between consecutive testing sessions (i.e., improvement was only evident across a span of roughly 20 weeks). There was also a significant improvement on the 0-100 scale, F(3, 276) = 24.36, p < .001, η_p^2 = .21, with a 10 week gain evident between Time 3 and Time 4, F(1, 94) = 21.85, p < .001, η_p^2 = .19.

We next examined changes in the fit of the log function (R^2_{log}), again using a 4 x 3 repeated measures ANOVA. It is important to note that changes to the log function are not, contrary to intuition, necessarily the inverse of changes in the fit of the linear function (i.e., a better linear fit does not necessarily imply a poorer logarithmic fit; see Muldoon et al., 2011). This second series of ANOVAs revealed significant improvement in logarithmic fits over time, F(3, 276) = 8.17, p<.001, η_p^2 = .08. There was also a significant main effect of Number Line, F(2, 184) = 154.88, p<.001, η_p^2 = .62, with a more logarithmic performance on the 0-20 relative to the 0-10 scale, F(1,98) = 131.49, p<.001, η_p^2 = .57, and more logarithmic performances on the 0-10, F(1, 92) = 46.02, p<.001, η_p^2 = .33, and 0-20 scales, F(1,92) = 353.22, p<.001, η_p^2 = .79, relative to the 0-100 scale. There was also a significant interaction between Time and Number Line, F(6, 552) = 6.06, p<.001, η_p^2 = .06. The interaction resulted from the same pattern found when analysing changes in the linear function fits; although there was no significant effect of Time on the 0-10 number line, F(3, 294) = 0.76, p > .05, η_p^2 = .01, Time did exert an effect on the 0-20, F(3, 294) = 4.56, p < .01, η_p^2 = .04, and 0-100 number line, F(3, 276) = 15.90, p < .001, η_p^2 = .15.

A third 4 x 3 repeated measures ANOVA using error as the dependent measure revealed that, on the 0-100 scale in particular, children became more accurate across the four waves of the study, $F(3, 276) = 12.24$, $p<.001$, $\eta_p^2 = .12$. There was, in contrast, no overall change in accuracy on the 0-20 scale, $F(3, 276) = <1$, $p > .05$, $\eta_p^2 = .01$. It is interesting that children's estimations became increasingly linear on the 0-20 scale even though they did not become more accurate. Surprisingly, accuracy on the 0-10 scale actually declined over time, $F(3, 294) = 3.88$, $p = .01$, $\eta_p^2 = .04$. When all three number scales were combined and entered into a 3 x 4 ANOVA, there was no overall effect of Time, $F(3, 276) = 1.30$, $p > .05$, $\eta_p^2 = .01$. However, this omnibus test revealed a significant main effect of Number Line, $F(2, 184) = 175.12$, $p < .001$, $\eta_p^2 = .66$, where the highest error rates were observed on the 0-100 scale. There was also a significant interaction between Number Line and Time, $F(6,552) = 8.18$, $p < .001$, $\eta_p^2 = .08$, reflecting the scale-dependent changes already outlined. As longitudinal data on number line estimation is relatively scarce, and bearing in mind the fact that it offers the potential for crucial contextual information to inform theories and research into the assessment of number line estimation, we have included a correlation matrix for measures of error taken on all three number lines at all four time points as supplementary material.

Having identified changes in children's estimating, we turn to the question of whether these changes are related to changes in math ability. To provide most focus on the relationship between these two skills, Time 1 was compared with Time 4 to offer maximum contrast between phases of the study. A statistically significant increase was observed in math scores between Time 1 (M = 133.9, SD = 16.0) and Time 4 (M = 160.1, SD = 13.9); $t(98) = 17.16$ $p < .001$, $\eta_p^2 = .75$. Significant correlations were noted between math ability and R^2_{lin} values on the 0-20 ($r(99) = .28$, $p = .01$) and 0-100 scales ($r(99) = .35$, $p < .001$) at Time 1 and again at Time 4; 0-20 ($r(99) = .26$, $p = .01$) and 0-100 ($r(96) = .53$, $p < .001$). There were no significant correlations between math ability and R^2_{lin} values on the 0-10 number lines probably due to ceiling effects on that scale. A similar pattern was observed for the log function fits; there were significant correlations between math ability and R^2_{log} values on the 0-20 (*r(99)* = .25, *p* = .014) and 0-100 number lines (*r(99)* = .36, *p* < .001) at Time 1 and the 0-100 number line at Time 4 (*r(96)* = .42, *p* < .001). There was also a significant correlation between math ability and estimation error on the 0-20, *r(97)* = -.34, *p* = .001, and 0-100, *r(97)* = -.26, *p* = .01, scales at Time 1, and between estimation error (0-100) and math ability at Time 4, *r(94)* = -.30, *p* = .003.

Having identified correlations between estimating and math ability, latent growth curve modelling was carried out to test the hypothesis that the rate of changes in one variable (in this case linearity of estimations) was developmentally associated with the rate of change in the other variable. Growth curve modelling allows for the growth factor(s) of one process to be predicted by the growth factor(s) of another process. This was achieved by specifying regression relations between various parameters of growth using Mplus statistical software (Muthén & Muthén, 2006). We conducted separate growth curve models for the linearity values for each of the separate number line scales in order to assess their individual predictive quality. One growth curve was created for math ability (3 time points) and three separate growth curves for the linearity data (4 time points; one curve for each of the three scales) and used ‘Means and Variance Adjusted Weighted Least Square’ (WLSMV) as an appropriate and robust estimator for such distributions. In these models we controlled for the influence of the initial math ability scores on the rate of growth of the two slopes. These initial curves failed to produce a satisfactory model due to co-linearity between estimation and math ability.

However, after ‘banding’ the math ability scores into four ordinal categories (0-25%; 26-50%; 51-75%; 76-100%), re-analysis yielded one satisfactory model, χ^2 (df = 10, N = 99)

= 17.06, p = .06, using the linearity values produced in response to the 0-20 scale. Initial analyses had shown non-significant residuals associated with the slope of the math scores (banded). To allow identification of the model, we fixed both the variance and the residual variance associated with the slope to zero. Similarly, the growth model for linearity on the 0-20 scale revealed significant residuals associated with the intercept but no significant residual variance associated with the slope. To allow identification of the model the latter was fixed to zero. Regressing the preliminary math ability score on the value for linearity on the 0-20 scale revealed that maths ability at the start of the study (i.e., the intercept for math ability) predicts estimation linearity values at Time 1, .02, z = 3.02, p < .01. There was no evidence that the rate of change in math ability was related to the rate of change in linearity. However, it should be noted that as this model is a relatively poor fit for the data these conclusions must be viewed with caution. Furthermore, as successful models could not be produced using the linearity values associated with the 0-10 and 0-100 number line scales we could not definitively establish the nature of the relationship between changes in math and changes in linearity of estimation on these scales.

Notwithstanding the limited success of the growth curve analyses, it is still worthwhile looking at the trends in changes as the literature on children's estimation skills would benefit from such scrutiny. Also, it is necessary to consider why only performance on the 0-20 scale showed any developmental relationship with maths ability. Following the approach of Bertelleti et al., (2010) children's performance was categorized as being either *linear* (when the linear function represented a better fit than logarithmic function), *logarithmic* (when the logarithmic function represented a better fit than linear), or *neither* (when neither of the functions significantly fitted the data). Table 2 underlines several core messages. Firstly, most children produced estimations best fitted by the linear function on the 0-10 scale by the age of 5 (i.e., Time 1), with the superiority of the linear fit being maintained over subsequent testing periods. Secondly, as might be expected, the logarithmic function was a superior fit on the much larger 0-100 scale for a majority of children at Time 1. Perhaps counter-intuitively (but in line with the findings of Bertelleti et al.) not only does the logarithmic function provide a superior fit to performance at Time 1, there is a positive developmental trend in the numbers of children whose estimations are best fitted by this function. Thirdly, estimation profiles on the intermediate 0-20 scale reveal the pattern of change that might be expected for children just starting to engage more with numbers between 10 and 20; the percentage of children classified as possessing linear representations increases steadily from Time 1 to Time 4 with a concurrent reduction in the number of children classified as ‘logarithmic’.

Key issue 2: The quality of number line estimations as a unique predictor of math ability

The second issue was whether counting ability mediates the relationships between that estimation skill and other variables. Although the two counting measures (reciting and enumerating) were taken at Time 1, 2 and 4 the following analyses focus on Time 1 and Time 4. Between Time 1 and Time 4, the highest number children could recite increased significantly from a mean of 34 (SD = 30) at Time 1 to 69 (SD = 36) at Time 4, $F(1, 95) = 117.48$, $p < .001$, $\eta_p^2 = .55$. In line with expectation, most children (88%) could enumerate 10 dots at Time 1, with 94% being successful at Time 2 and 98% at Time 4. As anticipated with this age group, development was seen on the more challenging task of counting through the teens; 47% could enumerate 20 dots at Time 1, rising to 73% at Time 2 and 85% at Time 4.

There was a significant correlation between Reciting and math ability at both Time 1, $r(98) = .40$, $p<.001$, and Time 4, $r(97) = .58$, $p < .001$. Children who could enumerate all 20 dots performed significantly better on the math tasks than those who failed to enumerate all

20 dots at Time 1, $t(97) = 3.2$, $p = .002$, $\eta_p^2 = .09$, and again at Time 4, $t(95) = 5.70$, $p < .001$, $\eta_p^2 = .25$. This difference was not evident on the enumerating 10 dots task, probably due to a ceiling effect.

Having established links between counting skills, quality of estimations and math ability it is important to consider the extent to which these facets of numerical cognition overlap. Although enumerating scores were not, as nominal data, entered into these analyses, stepwise linear regression revealed that at Time 1, Reciting was the largest contributor to the explained variance in math ability, $R^2 = .16$, F(1, 96) = 18.41, $p < .001$, with only R^2_{lin} on the 0-20 scale making a significant contribution thereafter, $\Delta R^2 = .040$, F(1, 95) = 4.98, $p = .03$. The additions of R^2_{lin} on the 0-10 and 0-100 number lines failed to make significant contributions to the model ($p = .90$ and .17 respectively). Reciting was similarly the largest contributor to the explained variance in math ability at Time 4, $R^2 = .34$, F(1, 92) = 47.14, $p < .001$. As for the Time 1 analysis, only one of the three R^2_{lin} measures made an additional contribution to the model once Reciting had been entered; linearity on the 0-100 scale, $\Delta R^2 = .07$, F(1, 91) = 10.93, $p = .001$.

As it had already established that development in estimating can entail a shift towards a better logarithmic fit as well as a linear fit, the stepwise linear regressions were run again, this time substituting R^2_{log} measures for R^2_{lin} measures. At both Time 1 and Time 4, the inclusion of the R^2_{log} measure on the 0-100 scale made a significant contribution to the relevant models once Reciting had been entered; $\Delta R^2 = .07$, F(1, 95) = 7.98, $p = .006$ (Time 1) and $\Delta R^2 = .06$, F(1, 91) = 8.91, $p = .004$ (Time 4).

Key issue 3: Are number line estimations more linear for numbers within children's counting range?

The final key issue focuses on the impact that number familiarity has on estimation. At Time 1 there were significant correlations between the highest number words recited and R^2_{lin} values on the 0-20 number line, *r(96)* = .20, *p* = .05, and the 0-100 number line, *r(96)* = .44, *p* < .001, a finding replicated at Time 4; (0-20) *r(95)* = .26, *p* = .01, (0-100) *r(92)* = .501, *p* < .001. However, significant correlations between Reciting and R^2_{log} values were only evident on the 0-100 number line; at Time 1, *r(98)* = .26, *p* = .01, and Time 4, *r(94)* = .37, *p* = .001. As most children (n = 87) could enumerate 10 items at the start of the study (a task achieved by all children by Time 4), it was not possible to produce any meaningful comparison statistics. However, at Time 1, children who could enumerate all 20 items (n = 46) produced estimations that were more linear on the 0-20, (*t(97)* = 3.10, *p* < .01, Cohen's *d* = .64, and 0-100, *t(97)* = 2.40, *p* < .05, Cohen's *d* = .48, scales than children who failed to enumerate all 20 items (n = 53). Furthermore, enumerators of all 20 items were more accurate on the 0-20 scale than other children, *t(97)* = 3.30, *p* < .001, Cohen's *d* = .67. Although analysis at Time 4 was problematic given the near- ceiling effect on the 'Enumerate 20 items' task (only 13 children were unable to carry this out successfully), non-significant trends in linearity were evident, with those children who could enumerate all 20 items producing higher R^2_{lin} values on all three scales.

In response to the claims that logarithmic estimation profiles should more accurately be described as two separate (bi-) linear profiles, and actually reflect separable processes (Ebersbach et al., 2008), we subjected the estimation data to bi-linear regression analysis using SegReg software (Oosterbaan, 2002). SegReg carries out a segmented linear regression by finding a breakpoint (if possible) in order to fit two separate linear functions that maximise the explanatory power of the resulting statistical coefficients.

One argument for bi-linearity in estimation profiles is that young children in particular hold linear mental representations for numbers they are familiar with, but they guess when estimating numbers above this range (e.g., Berteletti et al., 2010) thereby producing the more-or-less flat profile that has been taken by others to be part of a continuous logarithmic curve. The analysis revealed no break-point on either the 0-10 or 0-20 number lines at any of the time points. However, there was a mean break-point at approximately 15 on the 0-100 number line at all four time points, even though 66% of the group could count beyond 15 at Time 1, and over 95% counted beyond 15 at Time 4. Therefore, the stability of the break point at such a low value is difficult to explain using the ‘familiarity’ model of number line estimation performance, when one asks children directly about number knowledge in counting.

Discussion

Our longitudinal study provided insights regarding the three key issues we set out to address. First, and as anticipated, across the duration of the study variance in children’s estimation profiles was evident. Overall, estimations were best explained by a linear function on the smallest of the three scales (0-10) and by a logarithmic function on the largest (0-100) scale. In line with this developmental trajectory, linear and logarithmic function fits tended to be similar (i.e., equally good fits) on the intermediate (i.e., 0-20) scale. The same general patterns were observed when children were categorized as having a specific estimation profile (linear, logarithmic or neither), results that are consistent with cross-sectional research (Bertelleti et al., 2010). In both Bertelleti et al. and the present study there was an increase in both the proportion of children being categorized as linear and logarithmic, and a decrease in the proportion of children have no best fitting profile. Also, like Berteletti et al. it was observed that not only did the amount of variance explained by R^2_{lin} values significantly increase over time (at least on the 0-20 and 0-100 number lines), there was also a significant increase in the amount of variance explained by the R^2_{log} values on the same lines.

Of interest, however, was the question concerning the relationship between estimating skill and mathematical progress. One way of assessing the importance of estimation within the domain of number development is to ask whether longitudinal advances in estimating ability are associated with wider numerical development. Growth curve analysis revealed that while math performance was correlated with linearity of estimations on the 0-20 scale at the start of the study, and both indices of numerical ability improved over subsequent phases of testing, the rate of change in one was not associated with the rate of change in the other. Thus we could not find support for the hypothesis that changes in math ability are constrained by the linearity of children's mental representation of number. Although estimating and math ability do correlate at a given time point, the trajectories of the underlying processes are not tied.

There are at least two possible reasons why number line estimation might only correlate with mathematics skill when cross-sectional data are analysed. First, children with inaccurate internal representations (as manifested on the estimation task) may lack important conceptualisations about number that are required for some number tasks but not others. For example, linearity of 5-year-old's estimations is more closely associated with problems where the tasks are to match quantities with symbols than it is with problems where the task is to match equivalent sets of items (Muldoon, et al, 2011). Accordingly, when performance on a battery of different math tasks is taken as the dependent measure, changes in estimation quality will potentially constrain only a subset of those tasks. Second, the two domains – estimating and math - may each be a proxy index for one or more other skills, where those skills exert an influence differentially at different stages of number development.

The relationship between math, estimating and another related ability was addressed in the second of our key issues. An obvious candidate for a skill that potentially influences both the accuracy of children's mental representations of number and their math ability is counting. At an age when children are still grappling with learning how to count and being introduced new number terms, it seems reasonable to infer that knowing the number terms 1 to 10, and being able to enumerate a set of 10 items, will strengthen the mental representations of those numbers and lead to linear estimations. However, at the same time as they master the first decade, children are still learning about the numbers 11 to 20, and numbers up to 100.

Thus, an important question was whether estimation performance is a unique predictor of mathematical ability compared to counting. Although our range of counting tasks were not as extensive as those used by other researchers using counting skills as one facet of early numeracy (e.g., Jordan et al., 2007), we nonetheless observed overall improvements both in children's reciting and enumerating, and performance on the battery of math tasks. Therefore, the previously established link between counting range, math ability and the quality (linearity and accuracy) of children's internal representation of number is supported by the present findings. However, results from regression analysis indicate that after accounting for children's counting ability, performance on estimation tasks (and by extension, the linearity of children's internal representation of number) made, in most analyses, only a modest (i.e., non-significant) contribution to variance on standardised measures of mathematical attainment. Whilst previous studies have suggested that the linearity of number representation is both an important and potentially unique contributor to children's mathematical development (e.g., Booth and Siegler, 2006; Siegler & Booth, 2004; Siegler & Mu, 2008) they have not always controlled for many other potential mediating variables. By including just one additional number task, our data challenge at least the latter conclusion. Although performance on the number line task was related to mathematical ability throughout the course of our longitudinal assessment, math performance was equally associated with the most basic of counting tasks, at least for the range of number tasks and age group in our study.

This finding has implications for two other theoretical positions that were addressed in the last of our key issues. One possibility is that children's estimations shift towards patterns explained by logarithmic functions and *then* progress to patterns explained by linear functions. Why might this be so? We can speculate that when children who lack a meaningful understanding of the number line task are presented with estimation problems, they are likely to produce a more-or-less flat (i.e., horizontal) profile of estimations. This yields linear and logarithmic functions close to zero that also have slope values close to zero. Development therefore entails a shift from a straight line at 0 degrees to one at 45 degrees, with - chronologically - a logarithmic profile in between, which would concur with the model proposed by Bertelleti et al. (2010) concerning the effects of contextual familiarity.

In relation to the relatively new research that has put forward a case for a bi-linear model of mental number representations our data failed to provide evidence for this type of function on two of the three scales we used. On the smallest (0-10) scale we can infer that the absence of any bi-linearity in the estimation profiles is inevitable given the goodness of fit of a single linear function. More interesting is the absence of bi-linearity on the intermediate (0-20) scale. Here, the data yielded significant fits for the logarithmic function but there was no evidence for two separate linear functions providing a good fit (i.e., the logarithmic curve could not be broken down into two linear sections). In contrast, a bi-linear function was observed for the 0-100 number line, similar to the results of Ebersbach et al. (2008). The identified break-point was inconsistent with Ebersbach et al.'s suggestion that the bi-linear function would show a break-point at the extent of children's familiarity of number, (Ebersbach et al. explained the lack of this relationship in their data due to the nature of their familiarity task - counting across decades). However, we asked children to count as high as they could, but still did not find an association between reciting and the identified break-point.

Conclusions

Why should we find evidence of a close developmental association between counting, estimation and math ability? Counting items is a behavioural analogue of the mental parsing of space. Importantly, for very young children, counting is most often conducted on sets of items distributed linearly, where the task demands procedure coordinating verbal recitation of the number string with pointing to, or touching, items from left-to-right. In terms of current thinking on the development of numeracy, counting provides input to a cognitive system attuned to the concept of 'one' and the successor function (see Carey, 2004). It is possible that the action of counting items – particularly physical objects – supports an emerging understanding that the use of one more/one less number word means one more/one less item, and that adding or subtracting one item changes the cardinal number by one.

Thus, advances in counting ability (both reciting the number string and enumerating sets of items) have the potential to generate mental representations that conform to the set of natural integers. Moreover, better quality representations should aid performance on simple number operations like addition, subtraction and judgments of numerical equivalence, and be reflected in an increasingly accurate and linear mental number line (Muldoon, Lewis, & Freeman, 2009). Siegler and Ramani (2009) found that preschoolers who played board games (where counters were moved linearly across equally spaced squares) showed gains in the quality of their estimations on a 0-10 scale and outperformed children in control groups on a battery of arithmetic tasks (See also Whyte & Bull [2008] for similar results). Siegler and Ramani concluded that playing such board games helped children form a retrieval structure for encoding, storing, and accessing single-digit numbers. Our data are consistent with this hypothesis. Children who could count all 20 items in a linear array (i.e., those children who were able to maintain the coordination required to adhere to 1-1 correspondence between words and items) were better on the math tasks and displayed more linear profiles on the estimation tasks at both the start and end of the study.

The conclusion is that children with less accurate internal representations of number are probably disadvantaged on some math tasks when compared to children with better quality representations. Most importantly, the domains of both arithmetic and estimating appear to be tied closely to counting. However, notwithstanding the importance of procedural counting to both estimation performance and math attainment, we propose that future work in this area might benefit from considering conceptual understanding of counting. Gains in procedural counting are typically followed by gains in the ability to reflect on how the accuracy of the counting procedure identifies both ordinal relations and cardinal representations (Muldoon, Lewis, & Freeman, 2003). It is possible that conceptual knowledge of the counting routine is an additional predictor of individual differences in both estimation and math performance, and future studies testing this hypothesis would make a useful contribution to this line of research. In considering the inter-relationships between math ability, counting and estimating, the present findings make it possible to conclude that neither procedural counting nor estimation are uniquely or especially privileged correlates with mathematical development in the early years.

References

Aunio, P. & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20,* 427-435.

Barth, H. C. & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14,* 125-135.

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology, 46* (2), 545-551.

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

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

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus. Winter,* 59-68.

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 line. *Journal of Experimental Child Psychology, 99*, 1-17.

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

Moeller, K., Pixner, S., Kaufmann, L. & Nuerk, H-C. (2009). Children's mental number line: Logarithmic or decomposed linear. *Journal of Experimental Child Psychology, 103*, 505-515.

Muldoon, K., Lewis, C., & Freeman, N. H. (2003). Putting counting to work: preschoolers' understanding of cardinal extension. *International Journal of Educational Research, 39*, 695 – 718.

Muldoon, K., Lewis, C. & Freeman, N. (2009). Why set-comparison is vital in early number learning. *Trends in Cognitive Sciences, 13*, 203-208.

Muldoon, K., Simms, V., Towse, J., Menzies, V. & Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *Journal of Cross Cultural Psychology, 42*, 669-681.

Muthén, L.K., & Muthén, B.O. (2006). *MPLUS: Statistical Analysis with Latent Variables: User's Guide (4th ed.)*. Los Angeles, CA: Muthén and Muthén.

Nuerk, H. C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Developmental Psychology, 40*, 1199-1211.

Oosterbaan, R. J. (2002). http://www.waterlog.info/segreg.htm

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

Siegler, R. S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science, 19*, 759-763.

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

Siegler, R.S., & Ramani , G. B.(2009). Playing linear board games- but not circular ones- improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545-560.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind*, *Brain*, *and Education*, *3*, 143-150.

Whyte, J. C. & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology, 44*, 588-96.

Table 1: Estimation error, slope, R^2_{lin}, R^2_{log} for 0-10, 0-20 and 0-100 number lines at four time points

	0-10				0-20				0-100			
Time												
	1	2	3	4	1	2	3	4	1	2	3	4
(a) Analysis of individual children's estimates (mean values)												
Error(%)	17	19	18	19	15	14	13	15	31	29	28	26
R^2_{lin}	.78	.77	.80	.83	.71	.76	.81	.83	.27	.30	.34	.45
R^2_{log}	.66	.63	.65	.66	.75	.79	.83	.83	.46	.49	.52	.62
(b) Analysis by group (median values)												
R^2_{lin}	.94	.94	.93	.92	.94	.97	.98	.99	.45	.48	.56	.69
R^2_{log}	.76	.76	.74	.73	.99	.99	.98	.95	.80	.82	.87	.92

Table 2: Participants' profile (%) for 0-10, 0-20 and 0-100 number lines at four time points

	0-10				0-20				0-100			
Time	1	2	3	4	1	2	3	4	1	2	3	4
Linear	85	87	93	97	28	33	35	55	4	3	7	8
Logarithmic	8	4	2	2	56	55	59	43	80	85	87	91
Neither	7	9	5	1	16	12	6	2	16	12	6	1

Supplementary material.

As we administered the task with three scales on four separate occasions, we were able to investigate the stability / consistency in performance longitudinally. These data offer the first assessment that we are aware of, relating to the reliability of number estimation measures. This is important in part because it potentially provides an upper bound on the ability of number line estimation variables to correlate with others (via attenuation, for example). The data also address consistency across different number scales, which can help illuminate issues of process overlap. We use the error measure from the main section as an exemplar (the inter- correlations across scales and time points are shown in the table below).

Thus, with both the 0-10 and 0-100 scales, children's estimations at each time point correlated with all three other time points. Performance on the 0-20 scale was less stable, yet there was still a correlation between Time 1 and 2, and Time 3 and 4. In contrast, only 6 of the 12 correlations between scales at any particular time point were significant. In other words, accuracy for a given scale was often correlated across a full school year even when it failed to correlate with a neighbouring scale administered at the same time. This leads us to believe that there were coherent scale-dependent processes operating over an extended period, but that different ranges elicited potentially different representations.

Correlation matrix for error rates on all three number line scales (10, 20,100) at all four time points (T1 – T4).

	T1-10	T2-10	T3-10	T4-10	T1-20	T2-20	T3-20	T4-20	T1-100	T2-100	T3-100	T4-100
T1-10	1	.479*	.263*	.462*	.020	.105	-.002	.126	-.107	-.036	-.131	-.101
T2-10		1	.428*	.440*	.207+	.409*	.101	.074	.083	-.100	-.047	-.239+
T3-10			1	.496*	.050	.046	.261*	. 169	.246+	-.032	-.048	.034
T4-10				1	-.139	.014	.190	. 263*	.070	-.056	-.065	-.010
T1-20					1	.421*	.109	. 173	.300*	.385*	.250+	.211+
T2-20						1	.164	. 054	.158	.138	-.066	-.122
T3-20							1	. 436*	.147	.102	.238+	.159
T4-20								1	.208+	.069	.086	.463*
T1-100									1	.559*	.346*	.412*
T2-100										1	.432*	.572*
T3-100											1	.369*

Note. + = p<.05; * = p<.01