

How feedback improves children's numerical estimation

Hilary Barth¹ · Emily Slusser^{1,2} · Shipra Kanjlia^{1,3} · Jennifer Garcia^{1,4} · Jessica Taggart^{1,5} · Elizabeth Chase^{1,6}

Published online: 14 December 2015
© Psychonomic Society, Inc. 2015

Abstract Developmental change in children's number-line estimation has been thought to reveal a categorical logarithmic-to-linear shift in mental representations of number. Some have claimed that the broad and rapid change in estimation patterns that occurs with corrective feedback provides strong evidence for this shift. However, quantitative models of proportion judgment may provide a better account of children's estimation patterns while also predicting broad and rapid change following feedback. Here we test the hypothesis that local corrective feedback provides children with additional reference points, rather than catalyzing a shift to a different mental representation of number. We tested 117 children from several second-grade classrooms in a number-line feedback study. Data indicate that the proportion-judgment framework accounts for individual differences in estimation patterns, and that the effects of feedback are consistent with the unique quantitative predictions of the framework. They do not provide evidence supporting the representational shift

hypothesis or, more broadly, for the proposal that cognitive change can occur rapidly at the level of entire mental representations.

Keywords Cognitive development · Mathematical cognition

Introduction

Experimental evidence from a simple number-line estimation task with children has been said to show that cognitive change can occur at the level of entire mental representations (Opfer & Siegler, 2007; Siegler et al., 2009), consistent with overlapping-waves theory (Siegler, 1996). In the context of this theory, children have access to multiple mental representations at once, and may rapidly substitute a more adaptive representation for another.

A typical number-line estimation task asks participants to mark the positions of Arabic numerals on a line with labeled endpoints. Generally speaking, younger children are less accurate than older children and there are systematic differences in their estimates: Younger children dramatically overestimate smaller numbers such that when estimates are plotted against given numbers, they are better fit by a logarithmic function than by a linear one. Older children's estimates are better fit by a linear function. Moreover, children whose estimates are more logarithmic with a larger, less familiar number range often produce more linear estimates with a smaller, more familiar range. As a result, children have been said to possess multiple coexisting mental number representations, drawing upon linear representations when dealing with more familiar ranges and logarithmic representations for less familiar ranges (Siegler & Opfer, 2003). With experience and over the course of development they rely more on linear representations, which support more accurate estimation, undergoing a

✉ Hilary Barth
hbarth@wesleyan.edu

¹ Department of Psychology, Wesleyan University, 207 High St, Middletown, CT 06459, USA

² Department of Child and Adolescent Development, San Jose State University, San Jose, CA, USA

³ Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA

⁴ MDRC, New York, NY, USA

⁵ Department of Psychology, University of Virginia, Charlottesville, VA, USA

⁶ California School of Professional Psychology, Oakland, CA, USA

“representational shift” from logarithmically organized to linear mental representations of number (e.g., Booth & Siegler, 2006; Booth & Siegler, 2008; Siegler & Opfer, 2003; Siegler & Ramani, 2008; Siegler, Thompson, & Opfer, 2009).

Particularly strong evidence for the representational shift, and more broadly the idea that cognitive change can occur rapidly at the level of entire mental representations, is said to come from findings that corrective feedback can turn children’s logarithmic-looking estimation patterns into more linear ones, purportedly by encouraging children to access linear rather than logarithmic mental number representations (Opfer & Siegler, 2007; Opfer et al., 2011). In a carefully designed study (Opfer & Siegler, 2007), second-grade students were first given a 0–1,000 number-line pretest that identified them as either logarithmic or linear responders. “Logarithmic” responders received feedback designed to promote linear responding. It was hypothesized that the best feedback for effecting this change would be in the region of the number line with the greatest discrepancy between logarithmic and linear representations (around 150; Opfer & Siegler, 2007). According to the authors’ “logarithmic-discrepancy hypothesis,” feedback near that region would direct learners’ attention to the area of greatest discrepancy while feedback about correct placements of other numbers, around 5 or 725 for example, would not.

When initially “logarithmic” second graders received feedback (e.g., “You told me that 159 would go here. Actually, this is where 159 goes (pointing). The line you marked is where 430 goes.”), children who received feedback for target numbers around 150 produced more linear estimates at post-test than children who received feedback around 5 or 725. Thus, feedback about estimates in regions of the number line targeted to highlight the inadequacy of a logarithmic representation was more effective than feedback in other regions (or no feedback). In theory, simply demonstrating that the immature (logarithmic) representation was inappropriate for the task led children to adopt a different representation. Moreover, improvements were broad and rapid: they were not limited to the region in which feedback was concentrated. Instead, changes occurred over the entire line. These results were interpreted as powerful evidence for a categorical shift in children’s numerical magnitude representations (Opfer & Siegler, 2007; Opfer et al., 2011) and for the accompanying broader idea of rapid, cognitive change at the level of mental representations.

There are at least two reasons to interpret these results differently. First, the claims rest on the idea that local, targeted feedback caused broad changes. However, it is important to note that the feedback procedure gave children information not only about the targeted focal regions, but also about numbers corresponding to the locations of their erroneous estimates. Thus the range of feedback was actually rather broad, especially for children receiving 150-feedback (given that pretest estimates were maximally inaccurate around 150). In fact, differential effects following differential feedback are

compatible with many interpretations, and are not unique to representational change. (We return briefly to this point in the “Results and discussion” section.)

The main focus of this paper is on a second, more fundamental reason for a different interpretation: An ongoing debate exists over the idea that a representational shift underlies developmental change in numerical estimation. This debate has stemmed, in part, from recent evidence providing support for a theoretical explanation of developmental change in number-line estimation that involves judgments of proportion (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Peeters, Degrande, Ebersbach, Verschaffel, & Luwel, 2015; Rouder & Geary, 2014; Slusser et al., 2013; Slusser & Barth, under review; Sullivan, Juhasz, Slattery, & Barth, 2011; see also Cantlon, Cordes, Libertus, & Brannon, 2009; Hurst, Monahan, Heller, & Cordes, 2014). This explanation is based on a psychophysical model of proportion estimation (Hollands & Dyre, 2000; Hollands, Tanaka, & Dyre, 2002; Spence, 1990), originally developed to account for judgments of perceptual (not numerical) magnitude. It makes sense here because number-line tasks require the estimation of a smaller magnitude (the value presented) relative to a larger one (the value given at the upper endpoint). Models of proportion estimation provide good quantitative explanations of performance on a wide variety of tasks that involve magnitude judgments; they afford clear ways of tracking developmental change by exploring change in the model parameters; and they account for cyclical biases in estimation data that remain unexplained by the logarithmic-to-linear shift account (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Cohen & Sarnecka, 2014; Rouder & Geary, 2014; Slusser et al., 2013; Slusser & Barth, under review; Sullivan et al., 2011).

Our interpretation of the observed feedback effects is that post-feedback estimates change in part because of the use of additional reference points (see also Barth, Slusser, Cohen, & Paladino, 2011), rather than because of a shift to a linear mental number representation. A deep representational change need not be invoked – broad and rapid change should come about if feedback simply supplies children with additional reference points. To show how this explanation would account for the data, we first provide a brief review of the proportion estimation framework (see Slusser et al., 2013, for details).

A basic psychophysical model of proportion estimation (Spence, 1990) was derived from Stevens’ Law, which describes the relationship between the estimated or perceived magnitude of a stimulus and its actual magnitude as a power function $y = \alpha x^\beta$ (Fig. 1a). The exponent β has traditionally been thought to quantify the bias associated with an estimate of a particular type of stimulus magnitude (but see Teghtsoonian, 2012), and α is a scaling parameter. Spence (1990) showed that estimates of proportions (rather than individual magnitudes) take the form of S-shaped or inverse S-

shaped curves, depending on the value of β in question. For a typical 0–1,000 number-line task, estimates are predicted by $y = x^\beta / (x^\beta + (1000 - x)^\beta)$ (a one-cycle model; Fig. 1b). The basic model assumes that observers judge relative to two reference points (i.e., the endpoints of the stimulus range), and was generalized to account for the use of additional reference points, such as an inferred midpoint (Hollands & Dyre, 2000). The use of additional reference points leads to cyclic over- and underestimation patterns (Fig. 1c).

In number-line tasks, relatively older children and adults are more likely than younger children to produce data consistent with the multi-cycle model than with the one-cycle model (e.g., Barth & Paladino, 2011; Rouder & Geary, 2014; Slusser et al., 2013), suggesting that they make estimates relative to the endpoints of the number line and a midpoint as well (e.g., Slusser et al., 2013; see also Ashcraft & Moore, 2012). Relatively younger children, especially when faced with a less familiar number range, may produce estimates consistent with an “unbounded” model (i.e., a standard power model, Fig. 1a) as they seem not to use an upper reference point at all (though there are multiple strategies that could result in roughly this pattern of estimation; see Slusser et al., 2013; see also Barth & Paladino, 2011; Cohen & Sarnecka, 2014).

The use of available reference points is strategic, and reference points need not be centrally located. A similar pattern of over- and underestimation should appear between any pair of reference points (Hollands & Dyre, 2000). For example, the predicted estimation pattern in Fig. 1b (two endpoint reference points) is the same pattern repeated between each pair of reference points in Fig. 1c (two endpoints plus a midpoint). We propose that local corrective feedback provides children with a new reference point that can be applied immediately, consistent with the rapid change seen after feedback. The use of additional reference points also predicts broad change in estimation accuracy¹: compare, for example, Fig. 1b and c, in which the simple addition of a middle reference point has brought estimates closer to $y=x$ throughout the entire range.

Here we report the findings of a number-line estimation study with second graders. The main goal was to test our hypothesis that feedback serves to provide reference points that promote changes in estimations patterns, as predicted by the framework. To test a key prediction, we screened a large number of participants for a particular response pattern at pretest. While we report general results for the full sample, this work focuses on the evaluation of post-test data from this particular subset. A second goal was to determine whether the proportion judgment framework successfully models individual differences in estimation patterns.

¹ The value of β can also vary independently of the number of reference points, with $\beta=1$ falling on $y=x$ and deviations from 1 reflecting lower accuracy. Thus an observer could produce a one- or two-cycle pattern even if her estimates were not very close to perfect accuracy at $y=x$.

Method

This study was modeled closely on Opfer and Siegler (2007), with three modifications. First, rather than sampling heavily from the lower portion of the number range, we sampled evenly from the entire range because the proportion judgment account makes specific quantitative predictions across the entire range, and samples concentrated at the low end are inadequate for testing these predictions. Second, we did not include a condition with feedback clustered around 5 because this was not necessary to test our hypotheses. Third, the current study includes two (instead of three) blocks of feedback trials between pretest and post-test, because effects of feedback were found to appear quickly, often after just one or two trial blocks (Fig. 4 in Opfer & Siegler, 2007).

Participants

One hundred and seventeen children participated during October or November of second grade ($M = 7$ years and 6 months, range 6 years and 10 months to 7 years and 10 months). Children were tested at four elementary schools in central Connecticut. Two schools primarily served families of lower socioeconomic status (SES) and two served families of mid-range SES (indicated by the percentage of students receiving free/reduced lunch).

Materials and design

The number-line task was administered in a booklet, with one trial per page. Each sheet displayed a 23-cm line with “0” at the left end and “1,000” at the right end. The target number to be estimated was 2 cm above the center of the line. Children responded by drawing a vertical mark through the line.

Children were randomly assigned to one of three conditions: 150-feedback, 725-feedback, or no-feedback. All participants completed four blocks of number-line estimation trials (pretest, two trial blocks, and post-test). The pretest included 23 trials in a random order (7, 13, 22, 52, 111, 157, 240, 285, 365, 429, 464, 518, 558, 596, 643, 691, 752, 840, 887, 932, 975, 988, and 995). Each trial block began with three feedback trials, followed by 23 test trials with no feedback (9, 15, 24, 60, 108, 142, 244, 289, 348, 420, 466, 511, 563, 590, 645, 692, 748, 844, 933, 960, 978, and 996). The post-test contained the same 23 target numbers as the pretest.

Target numbers on feedback trials clustered around 150 (147–173) for children in the 150-feedback condition and around 725 (722–758) for children in the 725-feedback condition. Half the children in the no-feedback condition received target numbers matching the 150-feedback condition; the other half received target numbers matching the 725-feedback condition (but no feedback was actually given in this condition).

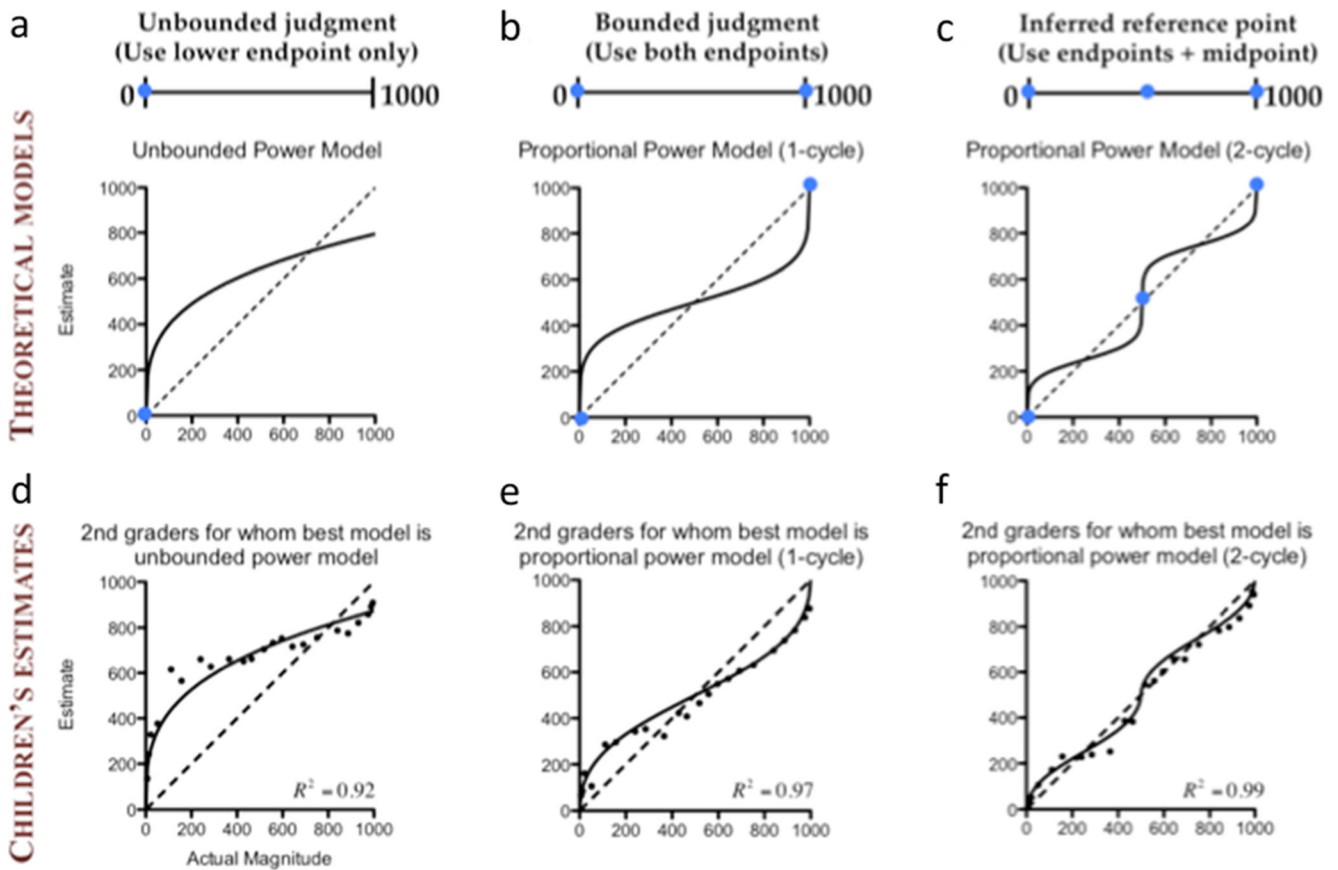


Fig. 1 Predictions of the model on a 0-1000 number line task (top row, a-c) and estimates of 2nd-grade children at pretest following their individual classification according to the three different versions of the model (bottom row, d-f)

Procedure

The experimenter introduced the task: “What I’m going to ask you to do is show me where on the number line some numbers are. This number line goes from 0 at this end to 1,000 at this end. When I ask you where a number goes, I want you to make a line through the number line where you think the number goes.” Before each trial, the experimenter asked, “If this is 0 and this is 1,000, where does *N* go?” Children completed the pretest, then two trial blocks, then the post-test.

Feedback was given on the first three trials in each trial block for children in the 150-feedback and 725-feedback conditions. To introduce these trials, the experimenter explained: “After you mark where you think the number goes, I’ll show you where the number *really* goes, so you can see how close your guess was.” After the child marked the line, the experimenter occluded the paper, marked the correct location for the target number (*N*) using a hidden ruler, and recorded the number associated with the child’s response (*X*). The experimenter then showed the child the corrected line and said, “You said that *N* goes here, but *N* actually goes here. That line you marked is where *X* goes.” If the child’s response fell within 50 points of the target number, the experimenter said: “You can see these two lines are really quite close. How did you

know *N* went there?” If the child’s response was more than 50 points from the target number, the experimenter said: “Your guess was a bit too high/low. You can see these two lines are quite far from each other. Why do you think this is too high/low for *N*?”

Results and discussion

Children were excluded from further analysis if estimates at either pretest or post-test were uncorrelated with presented numbers (Spearman rank correlation, $r_s, p > .05; n=24$) or the child used only 10 % of the line when marking locations on over 90 % of trials ($n=2$). The remaining 91 children were included ($M = 7$ years and 5 months, range 6 years and 10 months to 7 years and 10 months).

Comparison with Opfer and Siegler (2007)

We first asked whether our data were consistent with Opfer and Siegler (2007; O&S), who investigated the effect of feedback on initially “logarithmic” estimators. We identified children better fit at pretest by a logarithmic model than a linear model ($n = 41$). As in O&S, analyses were performed after testing, so children couldn’t be distributed among feedback conditions

based on pretest data. However, initially “logarithmic” children were present in all conditions (no-feedback: $n = 17$, $M = 6$ years and 9 months; 150-feedback: $n = 11$, $M = 6$ years and 9 months; 725-feedback: $n = 13$, $M = 6$ years and 9 months). Median estimates for these “logarithmic” estimators were similar across conditions at pretest (results of model comparisons²: no-feedback, $\log R^2 = 0.96$, $\text{lin } R^2 = 0.74$, $\Delta \text{AICc} = 40.36$; 150-feedback, $\log R^2 = 0.91$, $\text{lin } R^2 = 0.73$, $\Delta \text{AICc} = 25.44$; 725-feedback, $\log R^2 = 0.96$, $\text{lin } R^2 = 0.69$, $\Delta \text{AICc} = 47.37$).

Post-test estimates for children in the 150-feedback condition were more linear than logarithmic ($\log R^2 = 0.77$, $\text{lin } R^2 = 0.96$, $\Delta \text{AICc} = 39.18$). For the O&S 725-feedback condition, a linear fit was slightly but not significantly better than a logarithmic fit. Our data were consistent with this: median estimates for children in our 725-feedback condition were also reasonably well described by logarithmic or linear functions ($\log R^2 = 0.86$, $\text{lin } R^2 = 0.85$, $\Delta \text{AICc} = 1.67$). No-feedback children remained more logarithmic than linear at post-test ($\log R^2 = 0.94$, $\text{lin } R^2 = 0.84$, $\Delta \text{AICc} = 24.16$). Overall, current data were comparable to O&S and successfully replicated the basic findings.

Characterizing estimates according to the proportion judgment framework

The proportion judgment framework proposes that the number-line task involves proportion estimation and that children’s knowledge and experience may lead to different strategies. Accordingly, children should be categorized not as logarithmic versus linear responders, but by the reference points they use when making estimates (see Fig. 1). Children making bounded judgments (Figs. 1b and e) know enough about the numbers involved and the structure of the task to appropriately make estimates relative to values at both endpoints, leading to a “one-cycle” pattern of over-/underestimation. Children using endpoints and a midpoint (Figs. 1c and f) produce a “two-cycle” pattern of over-/underestimation, typically with more accurate estimates overall. Children making unbounded judgments (Figs. 1a and d) apparently do not make effective use of the upper endpoint.

These categories do not map perfectly onto logarithmic and linear classifications, but there are predictable relations

² Formal model comparisons used Akaike’s Information Criterion corrected for small sample sizes (AICc; Burnham & Anderson, 2002; Burnham, Anderson, & Huyvaert, 2011). ΔAICc gives the difference in AICc scores between another model and the preferred model (which has the lowest AICc score). Burnham and Anderson (2002) suggest, “As a rough rule of thumb, models having a Δ within 1–2 of the [preferred] model have substantial support and should receive considerations in making inferences. Models having Δ within about 4–7 of the [preferred] model have considerably less support, while models with $\Delta > 10$ have either essentially no support and might be omitted from further consideration or at least fail to explain some substantial structural variation in the data” (p. 446).

between classification schemes. Children “unbounded” at pretest would be classified as “logarithmic” according to the representational shift scheme. Children who were “two-cycle” at pretest would generally fall into the “linear” category (although they are fairly accurate, their estimates are better explained by the two-cycle proportional model than a linear model because they show characteristic cyclical patterns of over- and underestimation). Children who were “one-cycle” at pretest included some who would be classified as “logarithmic” and some who would be “linear.”

We classified children’s pretest data in terms of these three categories, based on AICc differences. Fifty-one children produced “unbounded” pretest estimates (Fig. 1d), 33 children’s pretest estimates were “one-cycle” (Fig. 1e), and seven children’s pretest estimates were “two-cycle” (Fig. 1f). These findings show that this theoretical framework can describe individual differences in children’s estimation patterns.

Testing a new hypothesis about the role of feedback

To test our core hypothesis that feedback provides new reference points within the proportional structure of the number line, we looked in particular at one subset of children: the children with “one-cycle” pretest estimates (Fig. 1e), who appear to use two reference points (both endpoints). Eleven initially “one-cycle” children received 150-feedback and 13 received 725-feedback (nine were in the no-feedback condition).

Our main hypothesis is that corrective feedback provides new reference points in the vicinity of the feedback, providing these children with a third reference location in addition to the two endpoints they were already using at pretest. The patterns of over- and underestimation seen in the pretest data should thus repeat between each pair of reference points used at post-test (Hollands & Dyre, 2000). This leads to specific predictions about the post-test data of two groups: the children who initially showed a one-cycle pattern and then received feedback near 150, and the children who initially showed a one-cycle pattern and then received feedback around 725. If feedback provides additional reference points, initially one-cycle children who receive 150-feedback should produce post-test data better explained by a model with a third reference point around 150 (see Fig. 2a), but initially one-cycle children who receive 725-feedback should produce post-test data better explained by a model with a third reference point around 725 (see Fig. 2b).

Because informative feedback was given not only at 150 and 725 but at a range of locations surrounding those focal numbers and at the locations of children’s erroneous placements (as in O&S), true feedback locations were “noisier” than modeled feedback locations (see discussion below). However, data did conform to model predictions. Following feedback around 150, group median post-test data from

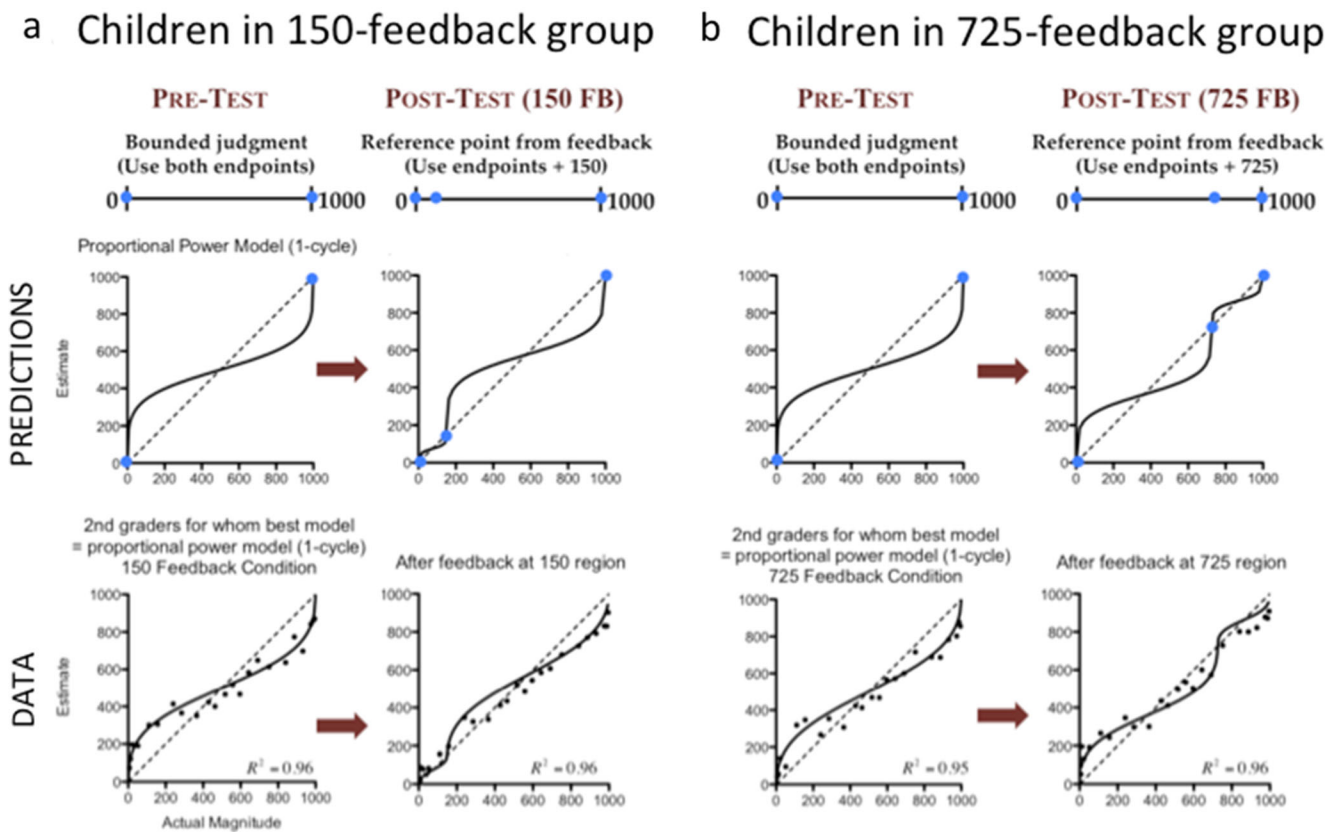


Fig. 2 Predictions (top row) and data (bottom row) regarding the effects of feedback for children who were classified at pre-test as making a bounded judgment using two reference points (the endpoints), resulting

in a one-cycle estimation pattern. Predictions and data for children receiving feedback around 150 are shown in (a); predictions and data for children receiving feedback around 725 are shown in (b)

children who were one-cycle at pretest ($n = 11$) were better explained at post-test by the model with a third reference point specifically located at 150 (vs. 725, $R^2 = 0.96$, $\Delta AICc = 10.35$; Fig. 2a). Following feedback around 725, group median data from children who were one-cycle at pretest ($n = 13$) were better explained at post-test by the model with a third reference point at 725 (vs. 150, $R^2 = 0.96$, $\Delta AICc = 19.41$; Fig. 2b).³

Effects of feedback on initially unbounded estimates

We also looked at the effects of feedback on the 51 initially “unbounded” children, for whom our framework does not make such specific quantitative predictions. For the initially “one-cycle” children discussed above, we could infer that they had the necessary knowledge to benefit from a new reference point (as their pretest estimates indicated that they already used both endpoints

appropriately), and we could predict specific patterns that should arise following feedback at 150 versus feedback at 725 (Fig. 2). The situation is different for children whose pretest estimates were “unbounded.” The unbounded fit could indicate a lack of knowledge of the magnitudes of the numbers near the upper end of the range, a lack of necessary ordinal knowledge, a failure to understand or respond appropriately to the bounded, proportional nature of the task, or some combination of these (see also Cohen & Sarnecka, 2014; Hurst et al., 2014). Because of this, our framework doesn’t make clear predictions about differential effects of feedback on reference point use for these children. For example, feedback clustered around 150 might lead these children to adopt a new reference point, but without sufficient knowledge to calibrate estimates relative to the upper endpoint of 1,000, they would not be expected to suddenly begin using the upper endpoint following feedback (i.e., feedback wouldn’t necessarily cause them to shift between categories in our classification scheme).

Nineteen initially “unbounded” children received 150-feedback and 20 received 725-feedback (12 were in the no-feedback condition). To assess how feedback

³ While the small size of this critical subset (24 children across two feedback conditions) may be a potential limitation of the study, it is comparable to the previous study which included 61 initially “logarithmic” children distributed unevenly in unspecified numbers across four feedback conditions (Opfer & Siegler, 2007).

influenced children's estimates, we computed an average pretest/post-test change score for each child for each feedback condition (subtracting post-test estimate from pre-test estimate for each number⁴). A one-way ANOVA on Condition (150-feedback, 725-feedback, no-feedback) showed a significant main effect of Condition, $F(2, 48)=5.149$, $p=.009$. Tukey post hoc analyses showed significant differences between no-feedback and 150-feedback conditions ($p=.026$) and between 150-feedback and 725-feedback conditions ($p=.021$), but no difference between 725-feedback and no-feedback conditions ($p=.957$).

Thus, the location of feedback mattered, with feedback trials using target numbers clustering around 150 (vs. 725) leading to a greater influence on initially "unbounded" children's estimates, similar to the initially "logarithmic" estimators of O&S. These results do not require explanation in terms of representational discrepancy. Rather, the vicinity of 700 is the region in which group estimates are relatively unbiased for children who initially produce either "logarithmic" or "unbounded" estimates. Due to the design of the feedback procedure, feedback revealed information not only about the correct location of the target number, but also about the numbers corresponding to the locations of the child's erroneous placements, wherever they might be. Thus feedback around 150 was likely both broader and more informative than feedback around 725, in both O&S and our study. For example, feedback received by initially "unbounded" children in our 150-feedback condition had a range of 883 (from 0 to 883, $M = 281$, $SD = 202$) while in our 725-feedback condition the range was 521 (from 407 to 928, $M = 719$, $SD = 93$). This difference across conditions could explain differences in estimates following these two types of feedback – there is no reason to call upon discrepancies with mental representations of number for an explanation.

Conclusions

We tested the hypothesis that local feedback about the accuracy of number-line estimates simply provides children with new reference points in the vicinity of the feedback, rather than supporting a shift to a different mental representation of number. This hypothesis arises from a theoretical framework according to which number-line estimation tasks should be treated as proportion judgments. Second graders completed a 0–1000 number-line task with a pretest, two feedback trial blocks, and a post-test. We used pretest behavior to classify children according to this theoretical framework, showing that it can describe individual differences in second graders' estimation patterns, and to identify a subset of children for whom the

theory makes specific quantitative predictions. The effects of feedback do not support the representational shift hypothesis. Rather, they are consistent with the proportional-reasoning framework and the idea that local feedback provides children with new reference points on the number line. More broadly, this work provides no evidence for the idea that cognitive change can occur rapidly at the level of entire mental representations.

Author note This work was supported in part by NSF DRL-0950252 to H.B. and a Wesleyan University Psychology Department Postdoctoral Fellowship to E.S. We thank Martha Liskow and Rachel Santiago for assistance with data collection. We also thank the Middletown School District, in particular the participating school leaders, teachers, families, and students who made this work possible.

References

- Ashcraft, M., & Moore, A. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*, 246–267.
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125–135.
- Barth, H., Slusser, E., Cohen, D., & Paladino, A. M. (2011). A sense of proportion: Commentary on Opfer, Siegler, & Young. *Developmental Science*, *14*, 1205–1206.
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*, 189–201.
- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*, 1016–1031.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference. A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23–35.
- Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on 'Log or Linear? Distinct intuitions of the number scale in Western and Indigene cultures'. *Science*, *323*, 38b.
- Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, *18*, 331–338.
- Cohen, D. J., & Sarnecka, B. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, *50*, 1640–1652.
- Hollands, J. G., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*, 500–524.
- Hollands, J. G., Tanaka, T., & Dyre, B. P. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 563–574.
- Hurst, M., Monahan, K. L., Heller, E., & Cordes, S. (2014). 123s & ABCs: Developmental shifts in logarithmic to linear responding reflect fluency with sequence values. *Developmental Science*, *17*, 892–904.
- Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169–195.

⁴ A similar analysis using absolute values yielded the same results.

- Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science, 14*, 1194–1204.
- Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2015). Children's use of number line estimation strategies. *European Journal of Psychology of Education*. doi:10.1007/s10212-015-0251-z
- Rouder, J., & Geary, D. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*, 525–536.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*, 237–250.
- Siegler, R. S., & Ramani, G. B. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375–394.
- 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.
- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General, 142*, 193–208.
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 683–692.
- Sullivan, J., Juhasz, B., Slattery, T., & Barth, H. (2011). Adults' number-line estimation strategies: Evidence from eye movements. *Psychonomic Bulletin & Review, 18*, 557–563.
- Teghtsoonian, R. (2012). The standard model for perceived magnitude: A framework for (almost) everything known about it. *American Journal of Psychology, 125*, 165–174.