Empirical Research

What Applying Growth Mixture Modeling Can Tell Us About Predictors of Number Line Estimation

Jeffrey M. DeVries*[a], Jörg-Tobias Kuhn[b], Markus Gebhardt[a]

[a] Research in Inclusive Education, Faculty of Rehabilitation Science, Technical University Dortmund, Dortmund, Germany.
[b] Educational Research Methods, Faculty of Rehabilitation Science, Technical University Dortmund, Dortmund, Germany.

Abstract

Number line estimation tasks have been considered a good indicator of mathematical competency for many years and are traditionally analyzed by fitting individual regression curves to individual responders. We innovate on this technique by applying growth mixture modeling and compare it to traditional regression using a sample of 2nd graders (n = 325) who completed both 0–20 and 0–100 number line tasks. We explore the effects of gender, special education needs, and migration background. Using growth mixture modeling, more children were identified as logarithmic responders than were identified using regressions. Growth mixture modeling was able to identify the significant effects of gender on class membership for both tasks, and of special education needs for the 0–20 task. Overall, growth mixture modeling provided a more complete picture of individual response patterns than traditional regression techniques. We discuss the implications of these findings and provide recommendations for future researchers to use growth mixture modeling with future number line task analyses.

Keywords: number line estimation, growth mixture modeling, mixture modeling, latent growth modeling, special education needs, migration background

Number line estimation is a well-developed task with clear correlates to the development of numerical cognition (Fazio, Bailey, Thompson, & Siegler, 2014; Friso-van den Bos et al., 2015; Schneider, Thompson, & Rittle-Johnson, 2018; Siegler, Thompson, & Schneider, 2011). Since its early research (e.g., Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003), a great deal of literature has contributed to its development with few advances in analysis. Common analyses fit linear, logarithmic, power, segmented linear, or other regression curves to number line responses, creating a graphical depiction of either individual or aggregate response patterns. Opfer and Siegler (2007) used this technique to describe the progression of children from logarithmic to linear response patterns. Rouder and Geary (2014) further developed this technique to examine multiple possible curvilinear regression patterns. Then, curves were compared based upon absolute error without accounting for the probability that the underlying representation of a child is actually linear or logarithmic. This paper innovates on traditional number line analyses techniques by using growth mixture modeling (Ram

& Grimm, 2009) which can estimate the probability that a child's responses follow a linear, logarithmic, or other underlying representation.

The Logarithmic to Linear Shift

Early work showed that younger children provide logarithmic estimates (i.e., greater discrimination for smaller quantities) on number lines, while older children provide linear estimates (Siegler & Booth, 2004; Siegler & Opfer, 2003). This shift was described as evidence that a child developed a more consistent understanding of numbers over the course of development (Opfer & Siegler, 2012; Praet & Desoete, 2014; Siegler, Thompson, & Opfer, 2009). However, later work by Opfer and Siegler (2007) showed that this representational change can come from of a single item of feedback. Similar findings demonstrated other instances of very rapid representational shifts (Laski & Siegler, 2014; Siegler, 2009; Siegler & Ramani, 2009).

Other work has suggested that children commonly solve number line estimation tasks using an anchoring strategy (Friso-van den Bos et al., 2015; Peeters, Degrande, Ebersbach, Verschaffel, & Luwel, 2016; Peeters, Verschaffel, & Luwel, 2017). Rouder and Geary (2014) examined the way benchmarks predicted responses with novel regression functions, such as an S-shaped two-anchor model, where responses were predicted based on distance to the both end-points of the number line, and not just the origin. They calculated Bayesian regression models for each child and selected the model with the lowest error for each child and measurement point.

Bayesian regression was also used in other recent articles (Friso-van den Bos et al., 2015; Kim & Opfer, 2017). It represented an important innovation in number line task analysis. It can account for many assumption violations of regression analyses; however, the interpretation of response patterns still relied on nonprobabilistic diagnoses, where the model with the lowest error was selected.

Covariates of Number Line Estimation

Response patterns on the number line task vary based on the magnitude of the target number (Kim & Opfer, 2017), the child's age (Praet & Desoete, 2014; Siegler & Opfer, 2003), socioeconomic status (SES; Fuchs et al., 2013; Ramani & Siegler, 2011; Siegler & Ramani, 2009), special education needs (SEN; Opfer & Martens, 2012; Tian & Siegler, 2017) and other factors of culture and the task (Laski & Yu, 2014; Leibovich, Al-Rubaiey Kadhim, & Ansari, 2017). Similarly, there is substantial evidence linking multiple variables to the development of number cognition. Boys tend to outperform girls on math tasks as early as the start of school (Cimpian, Lubienski, Timmer, Makowski, & Miller, 2016), while children from a migration background tend to underperform (Stahl, Schober, & Spiess, 2018). Similarly children with SEN lag in math development (Gebhardt, Zehner, & Hessels, 2014; Hansen et al., 2015). Yet, relatively few studies have examined the specific effects of SEN and migration background on number line tasks.

Some tasks have looked at related variables. Opfer and Martens (2012) found that adults with William's Syndrome did not change from logarithmic to linear response patterns, but their accuracy did improve. However, this is not a direct investigation of the overall impact of SEN, which more broadly refers to other needs including behavioral, linguistic, and attention difficulties. Meanwhile, Ramani and Siegler (2011) demonstrated that playing games that simulated a linear spatial representation facilitated linear performance on later number estimation tasks, particularly for learners with lower SES. Finally, Laski and Yu (2014) found that Chinese

children were more likely to provide linear responses than Chinese-American children. However, this study investigated cross-cultural difference and not the effect of a migration background. Direct investigation of these qualities will expand the overall picture of the development of number cognition, number line estimation, and mathematical competence. In particular, it can be uncovered whether the relationship between number line estimation and mathematical competence is a direct one, or if the relationship is differentially affected by related variables.

Growth Mixture Modeling for Number Line Tasks

Commonly, studies analyze number line tasks based on children transitioning from a better logarithmic fit to a better linear fit (Opfer & Siegler, 2007; Rouder & Geary, 2014) or the percentage of children which had a better linear or logarithmic fit (Opfer & DeVries, 2008). A linear or a logarithmic responder was identified based on which (frequentist or Bayesian) regression model had lower error, without regard to the difference between linear or logarithmic fits or the location of error. In this way, the data was reduced to nominal categories. Meanwhile, growth mixture modeling can provide probabilities of linear or logarithmic fits, where a child is given a value between 0 and 1 reflecting the probability of being best fit by the model. This allows a gradation of certainty in the categorization responders.

Traditionally, latent growth modeling techniques are used to assess change over time (Duncan & Duncan, 2004). They can be easily adopted to many different growth curves (e.g., linear, logarithmic, linear segments, etc.), allow for various missing data solutions, and can model data with many different measurement time points. Latent growth models can model number line tasks in the same way that regressions have. Instead of calculating growth curve over time, one calculates the position of estimates over the space of the line. Similar to a random-slope, random-intercept model, this produces slopes and intercepts for the entire sample as well as estimates for each individual. Thus, separate regression models for each child are unnecessary. Furthermore, the models can compensate for issues of lower error rates around end- or anchor-points and different error distributions. Because latent growth models can be estimated within a structural equation modeling (SEM) framework, the models can be extended into other modeling techniques, such as mixture modeling.

Mixture modeling allows for the identification of latent classes (for a review, see Collins & Lanza, 2010). This is a good match for number line estimation studies because responses are believed to be generated from two distinct representations (i.e., linear or logarithmic). Graduated probabilities of class membership are generated based on both the degree and location of deviations from the expected curve. Mixture modeling applied to latent growth models is called growth mixture modeling (Ram & Grimm, 2009). This is a major improvement over all-or-nothing categorizing based on absolute error. Additionally, mixture models can test covariates based on class membership probabilities. Many past studies have examined covariates (e.g., gender, cognitive ability, age, and others; Laski & Yu, 2014; Opfer & Martens, 2012; Siegler & Ramani, 2009) based solely and better linear or better logarithmic regression fits. These results can be improved upon within a growth mixture modeling framework.

Furthermore, growth modeling can accurately capture a spatial relationship, while mixture modeling can potentially identify a particular strategy used by children within a block of number line trials. A child may use an anchoring strategy, which may be identified based on lower error rates at the anchors, or alternatively, the child may use a random response strategy, which may be identified by a flat line. Because growth mixture modeling

allows for a curve statistics for each individual child, such strategies can be identified and sorted into a class of responder.

In Figure 1, an example of a growth mixture model applied to a number line task can be seen. Since we use an SEM framework, the model can be read very similar to other structural equation models. In this simplified model, the task is a 0–4 number line with three trials (Numbers 1, 2, and 3). The row at the bottom represents these trials. These compare directly to observations made at different time points in a traditional latent growth model. Above are I and S for the intercept and slope from the latent growth model. The intercept is fixed at 1 in all cases, but we can see that the slope is fixed at f(x), where x is the target number for the estimation task. The scaling of the slope depends on the intended curve shape. In the linear class,

$$f(x) = x$$

and in the logarithmic class,

$$f(x) = \ln(x)$$

where x is the target of the individual estimation task (i.e., 1, 2 or 3 in this example). Above the intercept and slope is C, which represents class membership. As described above, the classes differ only in slope loadings. Class membership then depends on the outcomes of the latent growth model with the differing slope loadings. Finally, onto class membership is regressed the background variables of interest, gender, SEN, and migration background.
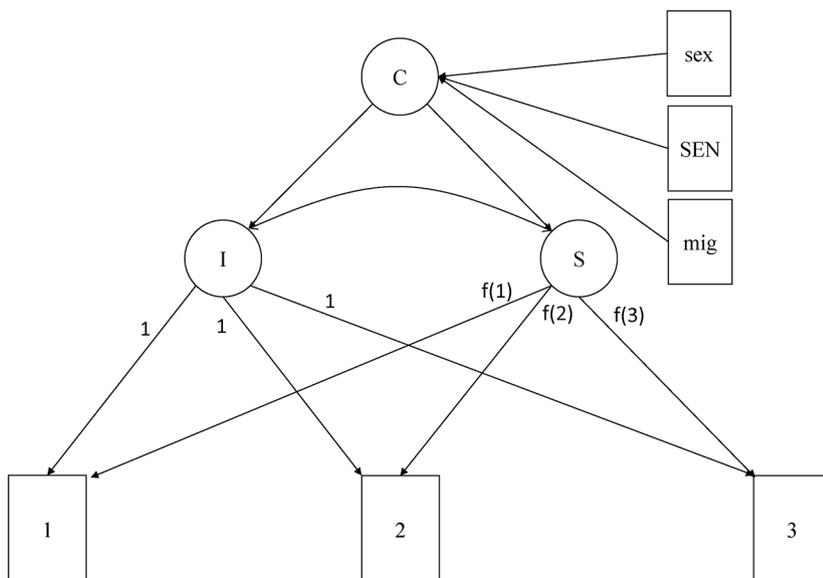
Figure 1. An example of a simplified growth mixture model for a number line task.

Note. SEN = special education needs; Mig = migration background; I = intercept; S = slope; C = class membership. Residual variables are not shown.

The Present Study

The present study describes the application of a pre-existing analysis to a new area, number line tasks. We apply growth mixture modeling to assess whether participant responses more closely match a linear,

logarithmic, or two-anchor model. We also assess the effect of SEN, migration background, and gender on model fit probability. Along these lines, we have developed a series of hypotheses.

First, we assess performance on both 0–20 and 0–100 number line tasks. We expect that the vast majority of children will provide responses corresponding to a linear model on the 0–20 task, while fewer will do so on the 0–100 task. This corresponds to Siegler and Opfer's (2003) findings of linear responses in the second grade on similar scales, as well as more recent findings (Fuchs et al., 2013; Kim & Opfer, 2017; Laski & Yu, 2014; Opfer & DeVries, 2008; Siegler, 2009).

Second, we examine performance based on gender, SEN, and migration background. We expect that boys will respond more linearly than girls, based on other work estimating the effects of gender on mathematical competency (Cimpian et al., 2016). Similarly, we expect children with SEN will more show logarithmic response patterns (Gebhardt et al., 2014; Hansen et al., 2015; Opfer & Martens, 2012). Furthermore, we expect that those with a migration background will also demonstrate more logarithmic responses, based on past work indicating a possible lag in mathematical development (Fuchs et al., 2013; Siegler & Ramani, 2009).

Third, we examine the fits of a two-anchors model, where responses vary based on distance to both endpoints, and not just the origin (see Rouder & Geary, 2014). We expect few children to fit within a two-anchors model because past results indicate a logarithmic model is still more likely early in development.

Finally, we repeat our analyses following the fit-based regression techniques used by previous work. We then compare the results of both techniques.

Method

Participants

Participants were 325 second grade students attending regular primary schools in the Northwest of Germany. Participants were recruited through their school administrators and teachers following established protocols for education research within Germany. Slightly under half the participants were boys (n = 144, 44.3%). Teachers were asked to report SEN and migration background of their participants, based upon whether the child or the child's parents were born abroad. The proportion of learners with a migration background was relatively high, although commiserate with the region (n = 143, 44.0%). A smaller number had SEN (n = 57, 17.5%), including language problems (n = 26), learning (n = 11), cognitive development (n = 6) and other (n = 14). An overview of the gender proportion and age of our participants can be found in Table 1.

Table 1

Participant Statistics

Group	Percent Female	Age in Years *M* (*SD*)
Overall	55.7%	7.83 (0.42)
SEN		
With Any	35.1%	8.01 (0.49)
None	60.1%	7.79 (0.40)
Migration Background		
With	60.1%	7.85 (0.43)
Without	52.2%	7.81 (0.42)

Note. SEN = special education needs.

Procedure

All data were collected by the same trained research assistant. We updated previous similar procedures (e.g., Opfer & Siegler, 2007; Rouder & Geary, 2014; Siegler & Opfer, 2003) for use with a tablet computer. Tasks were implemented via the web-platform Levumi (www.levumi.de). Students were told they would play a game on a tablet computer. They were shown a number line task on a 10-inch tablet and the touch-interface was explained to them. Next, they received their own tablet with an example problem on it. When they clicked on the number line, a blue line indicating their choice appeared. They could then click a new position or click on the continue button. Once they completed the sample item, they had 3 minutes to complete as many of the number line problems as they could. The 3-minute limit ensured participants were engaged with the task and allowed for data to be collected with minimum disruption to the regular class. In the 0–20 condition, children received all possible numbers between 0 and 20. In the 0–100 children, children received 20 numbers, 2 from each decade of the range (e.g., 2 between 10 and 19, 2 between 20 and 29, etc.). This procedure was replicated for both the 0–20 and 0–100 number lines.

Due to the random ordering of trials, items that were not reached were treated as missing at random and deleted pairwise in all analyses. The average number of missing responses was 3.2 (*SD* = 4.0) in the 0–20 number line and 4.1 (*SD* = 5.9) in the 0–100 number line. Most responders had fewer than 10% missing (64.6% in the 0–20 task and 63.1% in the 0–100 task). A multiple linear regression indicated that gender, migration background and SEN did not have a significant effect on the number of missing responses (all *p*s > .10). In six cases, no responses were given at all. These cases were removed from all analyses, leaving 319 participants.

Analysis

We applied growth mixture modeling (see Ram & Grimm, 2009 for a review) to estimate both linear and logarithmic response models. These were calculated using the robust maximum likelihood estimator (MLR; Yuan & Bentler, 2008). All growth mixture analyses were conducted using Mplus 7.4 (Muthén & Muthén, 1998–2017), and an example of our syntax and notes on model specification are provided in the Appendix. A two-class (linear and logarithmic) mixture model was applied to the 0–20 and 0–100 task. A three-class model (linear, logarithmic, and two-anchors; see Rouder & Geary, 2014) was also applied to the 0–100 number line task. A regression of gender, SEN, and migration background onto class membership probability was included in the mixture analysis via 1-step joint model estimation. The Appendix contains an example of our syntax and the scaling functions used to create a separate growth model for each latent class.

In order to compare analysis techniques, we also conducted a regression analysis. Separate linear and logarithmic regressions were calculated for each child. A child was categorized as linear or logarithmic based on which model had a better R^2.

Finally, we compare the results to linear and logistic regressions using gender, SEN, and migration background as predictors of linear R^2 or class membership, as determined by individual linear and logarithmic regressions of each child's responses.

Results

Overall Model Performance

The models converged successfully in all cases, successfully replicating the log-likelihood values in multiple random starts. Fits of all models are described in Table 2, and additional test statistics for the models are described in Table 3. As seen in Table 4, good latent class separation was also achieved in all models. Based on all three fit metrics, the Akaike information criterion (AIC), Bayesian information criterion (BIC), and adjusted BIC, the three class model performed better; however, very as seen in Table 3, very few responders (5%) fit the two-anchors class in the three-class model. We explore the two and three class models in more detail below.

Table 2

Information Criteria Fits of the Growth Mixture Models

Task	Model	AIC	BIC	Sample-Size Adjusted BIC
0–20	Two Class	22,585	22,699	22,604
0–100	Two Class	41,298	41,416	41,317
	Three Class	41,133	41,273	41,156

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 3

Likelihood Ratio Test Statistics for Model in Compared to Model With One Fewer Class

Task	Model	Lo-Mendell Rubin LRT *p*-value	Adjusted LRT *p*-value	Parametric Bootstrapped LRT Approximate *p*-value
0–20	Two Class	.071	.077	< .001
0–100	Two Class	.027	.029	< .001
	Three Class	.013	.015	< .001

Note. LRT = likelihood ratio test.

PsychOpen GOLD

Table 4

Entropy and Latent Class Membership Proportion and Class Separation

Model	Task	Entropy	Class	Class Proportion	Average Probability of Class Membership		
					Lin	Log	Two-Anchor
Two Class	0–20	1.00	Lin	.98	1.00	0.00	
			Log	.02	0.00	1.00	
	0–100	0.67	Lin	.57	0.92	0.08	
			Log	.43	0.12	0.88	
Three Class	0–100	0.84	Lin	.46	0.92	0.08	0.00
			Log	.49	0.07	0.92	0.01
			Two Anchor	.05	0.02	0.01	0.97

Note. Lin = linear class; Log = logarithmic class; Class proportion = relative proportion of children assigned to the given class.

Two Class Model

Graphical summaries of the two-class models for the 0–20 task can be seen in the left half of Figure 2. The linear class is well defined and shows a straightforward linear trajectory; however, the logarithmic class shows a very high intercept with a low slope. Responses for the logarithmic class are more erratic. It is therefore unlikely that the logarithmic class represents actual logarithmic responses on this scale, but instead indicates random responders. This is supported by the relative small proportion of membership in this class (less than 3%).
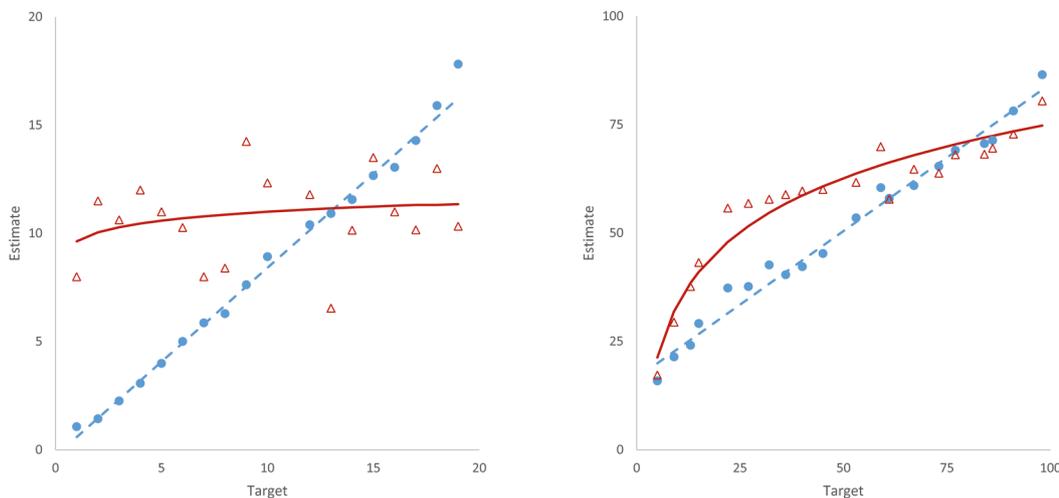

Figure 2. Two class model (linear and logarithmic) for 0–20 and 0–100 number lines.

Note. The dashed line is the predicted aggregate linear model, and the solid line is the predicted aggregate logarithmic model. Circles represent linear class mean responses, and triangles represent logarithmic class mean responses.

The 0–100 task also appears in Figure 2. In contrast to the 0–20 task, it demonstrates a clear logarithmic curve. Coupled with the high proportion of membership, a logarithmic interpretation of this class is appropriate.

Effects of Subject Variables on Class Membership

The odds ratio effects on the probability of being a linear responder is described in Table 5. On the 0–20 number line, the only significant predictor of linear or logarithmic group membership was SEN, who were significantly less likely to respond linearly, $p < .05$. On the 0–100 number line, males were significantly more likely to produce linear responses than females, $p < .01$.

Table 5

Odds Ratios for Linear Responders in the Two Class Model

	Number Line Range	
Group	**0–20**	**0–100**
Male	6.53	2.40**
Migration Background	0.97	1.27
Special Education Needs	0.19*	1.00

Note. An odds ratio describes the likelihood change of the linear class matching given the child with the described quality (e.g., male, possessing a migration background, or having special education needs). Values that are significantly different from 1 are marked.
*$p < .05$. **$p < .01$.

Three Class Model

We examined the three-class model for only the 0–100 number line task because the responders in the 0–20 task were overwhelming linear, and interpretation of the 0–20 model's second class was already unclear due to the flat slope of the logarithmic class. As seen in Table 4, the class separation remained good in the three class model for the 0–100 task. The graphical summary of the three-class model can be seen in Figure 3.
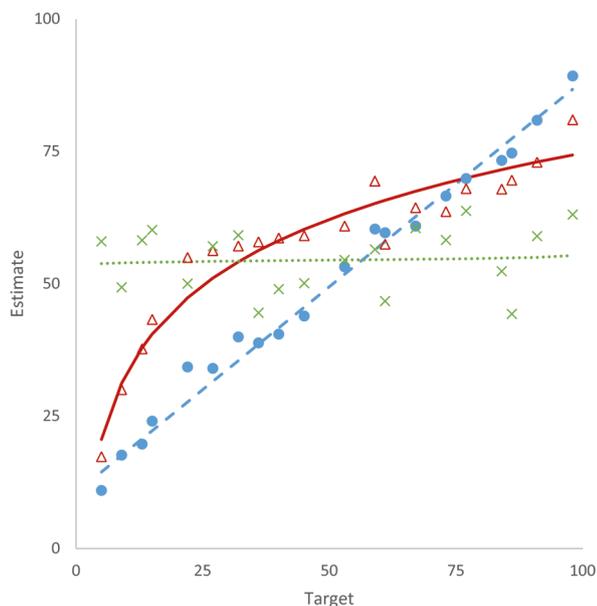

Figure 3. Three class model (linear, logarithmic, two anchors) on the 0–100 number line task.

Note. The dashed line is the predicted aggregate linear model, the solid line is the predicted aggregate logarithmic model, and the dotted line is the predicted two-anchors model. Circles represent linear class mean responses, triangles represent logarithmic class mean responses, and Xs represent two-anchors class mean responses.

Here the slope is quite flat, and there is relatively little sinusoidal curvature. This model appears to fit mostly a small group of random responders.

Table 6 shows that males were more likely to be in the linear class than the logarithmic class, $p < .01$. Migration background had no effect, $p > .05$. Children with SEN were more likely to the linear or logarithmic class than in the two-anchors class, ps $< .05$. Based on the flat curvature of the two-anchors solutions, this suggests that children with SEN were more likely to be random responders.

Table 6

Path Values Predicting Class Membership in the Three Class Model

Reference Class	Class	Predictor	Path Value	SE
Linear	Logarithmic	SEN	0.06	0.39
		Male	−0.88**	0.28
		Migration Background	−0.21	0.29
	Two-Anchors	SEN	1.65*	0.63
		Male	−1.25	0.70
		Migration Background	0.77	0.64
Logarithmic	Two-Anchors	SEN	1.59*	0.61
		Male	−0.37	0.70
		Migration Background	0.98	0.64

Note. SEN = special education needs. Positive numbers indicate that the stated class was more likely for that group than the reference class, negative values mean it was less likely, and zero means it was equally likely.
*$p < .05$. **$p < .01$.

Model Selection

While the information criteria fit values for the three-class model are better and the likelihood ratio test values indicate a significantly better fit, theoretical interpretation is also a critical factor in model selection. In this case, the interpretation of the three-class model does not follow the theoretical goals of the model. Instead of representing a logarithmic function of the distance from either end-point, it represented relatively flat, random responses of a very small (5%) proportion of responders. Thus, the three classes of the model were linear, logarithmic, and random responders. We therefore prefer the simpler 2-class model.

Comparison to Individual Regressions

We compared growth mixture modeling to traditional techniques which fit individual regression curves to each participant for the 0–100 task. We excluded the 0–20 task because of the relatively low variability in class membership at this range. These analyses produced very different results than the mixture modeling. Nearly 30% more responders were classified as linear (83%) than in growth mixture modeling. Additionally, we conducted three separate regressions, summarized in Table 7. The first regression was a logistic regression of SEN, gender, and migration background onto whether the individual linear regression fit better than the individual logarithmic regression. Here, both gender and migration background were significant, $p < .05$. We also conducted linear regression of SEN, gender, and migration background onto both linear R^2 and logarithmic R^2, no significant coefficients were found in either case, $p > .05$.

Table 7

Regression Results of Subject Variables Onto Individual Regression Model Results for the 0–100 Number Line Task

Model	Predictor	Unstandardized Beta	SE	p
Logistic	Male	.84	0.37	.023*
	SEN	.87	0.58	.135
	Migration Background	−.76	0.34	.027*
Linear R^2	Male	.66	0.34	.054
	SEN	−.03	0.05	.556
	Migration Background	−.06	0.04	.096
Logarithmic R^2	Male	.05	0.03	.147
	SEN	−.03	0.04	.495
	Migration Background	−.04	0.03	.257

Note. SEN = special education needs. The logistic model was coded as a better or equal linear R^2 = 1, and worse linear R^2 = 0. Thus, positive coefficients indicate an increase likelihood to be better fit by the linear function.
*$p < .05$.

Discussion

Our paper describes a new application of growth mixture modeling for the analysis of number line tasks. This technique identified a significant effect of gender on linear response probability in the 0–100 number line, and of SEN for the 0–20 number line. These results differed from traditional number line analyses, which involve fitting individual regressions for each child, where far fewer logarithmic responders were identified and significant effects were found for gender and migration background in the 0–100 task. Our analysis represents an important innovation in data analysis of number line tasks and allows for superior data analysis and better treatment of missing data. Furthermore, this growth mixture modeling can deal with violations of normalcy and homoscedacity. Importantly, it provides continuous estimates of probability of linear, or logarithmic (or other) response patterns. This technique can also be extended to analyze other functions including power, segmental, and other curves. The technique can be further extended to more measurement points and other response ranges via simple changes to the syntax and scaling functions. A further extension of the models to latent transition analysis could also model changes in response patterns over time or due to an intervention.

Our new analysis identified more children into the logarithmic class than the regression analyses did. This difference can be explained by error distribution and model overlap. Number line errors are not consistent throughout the entire line (see Rouder & Geary, 2014 for a detailed explanation of this issue). This is further complicated by some individuals using an anchoring strategy (e.g., Friso-van den Bos et al., 2015; Kim & Opfer, 2017; Rouder & Geary, 2014). Thus within expected error ranges, the linear and logarithmic representations overlap at many points. Responses at these points are less informative when identifying a linear or a logarithmic responder, but regression techniques typically use overall error of the regression model for this purpose. Mixture models give these responses less weight, which allows for a better, probabilistic assessment of responders.

This helps to explain the differences between growth mixture modeling and individual regression results. In the two class growth mixture model for the 0–100 number line task, boys were significantly more likely than

girls to fit the linear class. Similarly, they were more likely to have a better linear R^2 in the logistic regression, although their the R^2 did not differ significantly in the linear and logarithmic regressions. The logistic regression also found that children with a migrant background were significantly less likely to be better fit by the linear regression. However, the effect of migrant background was not significant in the growth mixture model. The effect of migration background was biased due to the small number of logarithmic responders. Growth mixture modeling avoided this by identifying a larger group of logarithmic responders.

While the data match the logarithmic curves well in the 0–100 condition, there are some additional issues about model performance. In the 0–20 condition, the vast majority (over 97%) of responders fit into the linear class, but the logarithmic class appears to be random responders and not true logarithmic responders. Therefore, for the 0–20 range, it would be better to interpret the logarithmic class as a class of random responders. This represents only a few outlier cases and may be an artifact of using a two-class model when almost all responders were linear. We see a similar artifact in the three-class model of the 0–100 task. It is therefore critical to carefully interpret each class when applying growth mixture modeling.

Higher class proportions of linear responders on the 0–20 task than the 0–100 task replicated the critical finding of a logarithmic to linear shift with increasing magnitudes (Friso-van den Bos et al., 2015; Laski & Yu, 2014; Opfer & Siegler, 2007). Our participants were overwhelming (over 97%) linear responders in the 0–20 condition, and only slightly above 50% linear responders in the 0–100 condition.

We also explored a three-class model, with a linear, logarithmic, and two-anchor class (based on Rouder & Geary's 2014 *M2*). In the two-anchor class, curves are connected from the end-point to the midpoint, suggesting participants' compressed representation starts at the end points, and is fitted through the midpoint of the number line. However, response patterns in the two-anchors class responded along a straight, flat line, and not along a sinusoidal curve. This suggests that they instead adopted a random response pattern. Further work fitting such models across more age groups may identify true two-anchor strategies.

While few studies have precisely examined the effects of gender on number line estimation, our finding that boys responded more linearly than girls was well predicted by previous work relating to mathematical competency and gender. More work precisely measuring and accounting for this effect remains necessary. We also found that children with SEN were significantly less likely to belong to the linear class on the 0–20 task, although not on the 0–100 task. However, this result requires further examination as it represented a very small number of our participants. Additional work involving classrooms from multiple countries and cultures, and students from diverse backgrounds, with and without SEN is necessary. Further work modeling the change in class membership after feedback, over time, and throughout development is also necessary.

Conclusion

We apply growth mixture modeling for the first time to number line estimation tasks. The new method can effectively discern both linear and logarithmic representations. This approach identified effects of SEN and gender, while no effects of migration background were found. Our method can be readily adapted to many different number line estimation tasks in future analyses.

Funding

Competing Interests

The authors have declared that no competing interests exist.

Acknowledgments

The authors have no support to report.

Data Availability

For this article, a dataset is freely available (see the Supplementary Materials section).

Supplementary Materials

All data used in this article are freely available as supplementary material (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

DeVries, J. M., Kuhn, J.-T., & Gebhardt, M. (2019). *Supplementary materials to "What applying growth mixture modeling can tell us about predictors of number line estimation"* [Research data]. PsychOpen. https://doi.org/10.23668/psycharchives.2526

References

Cimpian, J. R., Lubienski, S. T., Timmer, J. D., Makowski, M. B., & Miller, E. K. (2016). Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts. *AERA Open, 2*(4), Article 233285841667361. https://doi.org/10.1177/2332858416673617

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* [Wiley Series in Probability and Statistics, Vol. 718]. Hoboken, NJ, USA: Wiley.

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy, 35*(2), 333-363. https://doi.org/10.1016/S0005-7894(04)80042-X

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53-72. https://doi.org/10.1016/j.jecp.2014.01.013

Friso-van den Bos, I., Kroesbergen, E. H., van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., van der Schoot, M., . . . Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12-29. https://doi.org/10.1016/j.jecp.2015.02.002

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., . . . Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*(3), 683-700. https://doi.org/10.1037/a0032446

Gebhardt, M., Zehner, F., & Hessels, M. (2014). Basic arithmetical skills of students with learning disabilities in the secondary special schools: An exploratory study covering fifth to ninth grade. *Frontline Learning Research, 2*(1), 50-63. https://doi.org/10.14786/flr.v2i1.73

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development, 35*, 34-49. https://doi.org/10.1016/j.cogdev.2015.02.001

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developmental Psychology, 53*(6), 1088-1097. https://doi.org/10.1037/dev0000305

Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology, 50*(3), 853-864. https://doi.org/10.1037/a0034321

Laski, E. V., & Yu, Q. (2014). Number line estimation and mental addition: Examining the potential roles of language and education. *Journal of Experimental Child Psychology, 117*, 29-44. https://doi.org/10.1016/j.jecp.2013.08.007

Leibovich, T., Al-Rubaiey Kadhim, S., & Ansari, D. (2017). Beyond comparison: The influence of physical size on number estimation is modulated by notation, range and spatial arrangement. *Acta Psychologica, 175*, 33-41. https://doi.org/10.1016/j.actpsy.2017.02.004

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide.* Los Angeles, CA, USA: Muthén & Muthén.

Opfer, J. E., & DeVries, J. M. (2008). Representational change and magnitude estimation: Why young children can make more accurate salary comparisons than adults. *Cognition, 108*(3), 843-849. https://doi.org/10.1016/j.cognition.2008.05.003

Opfer, J. E., & Martens, M. A. (2012). Learning without representational change: Development of numerical estimation in individuals with Williams syndrome. *Developmental Science, 15*(6), 863-875. https://doi.org/10.1111/j.1467-7687.2012.01187.x

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*(3), 169-195. https://doi.org/10.1016/j.cogpsych.2006.09.002

Opfer, J. E., & Siegler, R. S. (2012). Development of quantitative thinking. In K. Holyoak & R. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 585–605). New York, NY, USA: Oxford University Press.

Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., & Luwel, K. (2016). Children's use of number line estimation strategies. *European Journal of Psychology of Education, 31*(2), 117-134. https://doi.org/10.1007/s10212-015-0251-z

Peeters, D., Verschaffel, L., & Luwel, K. (2017). Benchmark-based strategies in whole number line estimation. *British Journal of Psychology 108*(4), 668-686. https://doi.org/10.1111/bjop.12233

Praet, M., & Desoete, A. (2014). Number line estimation from kindergarten to grade 2: A longitudinal study. *Learning and Instruction, 33*, 19-28. https://doi.org/10.1016/j.learninstruc.2014.02.003

Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565-576. https://doi.org/10.1177/0165025409343765

PsychOpen GOLD

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology, 32*(3), 146-159. https://doi.org/10.1016/j.appdev.2011.02.005

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*(4), 525-536. https://doi.org/10.1111/desc.12166

Schneider, M., Thompson, C. A., & Rittle-Johnson, B. (2018). Associations of magnitude comparison and number line estimation with mathematical competence: A comparative review. In R. S. Siegler & P. Lemaire (Eds.), *Cognitive development from a strategy perspective: A Festschrift for Robert Siegler* (pp. 100–119). Abingdon, United Kingdom: Routledge.

Siegler, R. S. (2009). Improving the numerical understanding of children from low-income families. *Child Development Perspectives, 3*(2), 118-124. https://doi.org/10.1111/j.1750-8606.2009.00090.x

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444. https://doi.org/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*(3), 237-243. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—But not circular ones—Improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545-560. https://doi.org/10.1037/a0014239

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*(3), 143-150. https://doi.org/10.1111/j.1751-228X.2009.01064.x

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273-296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Stahl, J. F., Schober, P. S., & Spiess, C. K. (2018). Parental socio-economic status and childcare quality: Early inequalities in educational opportunity? *Early Childhood Research Quarterly, 44*, 304-317. https://doi.org/10.1016/j.ecresq.2017.10.011

Tian, J., & Siegler, R. S. (2017). Fractions learning in children with mathematics difficulties. *Journal of Learning Disabilities, 50*(6), 614-620. https://doi.org/10.1177/0022219416662032

Yuan, K.-H., & Bentler, P. M. (2008). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 30*(1), 165-200. https://doi.org/10.1111/0081-1750.00078

Appendix

Syntax

!Mplus 7.4 syntax excerpt for 0–100 number line with 2-classes.

```
VARIABLE:
names are
    ID MP Sex SEN Migrat Avg
    hu22 hu59 hu98 hu53 hu45 hu9 hu86 hu84 hu15 hu13
    hu5 hu27 hu32 hu77 hu67 hu61 hu40 hu91 hu73 hu36;
Idvariable is ID;
Missing are all (-99);
usevariables are
    hu22 hu59 hu98 hu53 hu45 hu9 hu86 hu84 hu15
    hu13 hu5 hu27 hu32 hu77 hu67 hu61 hu40 hu91 hu73 hu36
    SEN Sex Migrat;
classes are c(2);

MODEL:
%Overall% !Linear class as the default model
i s | hu5@5 hu9@9 hu13@13 hu15@15 hu22@22 hu27@27 hu32@32
    hu36@36 hu40@40 hu45@45 hu53@53 hu59@59 hu61@61
    hu67@67 hu73@73 hu77@77 hu84@84 hu86@86 hu91@91 hu98@98;
[i s];
c on SEN sex migrat; !Class membership on covariates

%C#2% !Class 2 uses the logarithmic model
i s | hu5@1.609 hu9@2.197 hu13@2.565 hu15@2.708 hu22@3.091 hu27@3.296
    hu32@3.466 hu36@3.584 hu40@3.689 hu45@3.807 hu53@3.970 hu59@4.078
    hu61@4.111 hu67@4.205 hu73@4.290 hu77@4.344 hu84@4.431 hu86@4.454
    hu91@4.511 hu98@4.585;
[i s];

ANALYSIS:
starts are 50 (20);
type is mixture;
estimator is MLR;
```

Details on Model Specification

This approach can be extended to additional functions by adjusting the y-intercept and scaling parameters. By editing the syntax, many different linear or nonlinear growth curves can be fit. Below are the functions we used to calculate each class. In addition to the three classes we describe in the full paper, we attempted to fit a four-class model, with the extra class following Rouder and Geary's (2014) 3-anchor model (i.e., both end points and the midpoint as anchors). This model converged with comparable AIC, but the three-anchors class contained only one responder. So it was excluded from our main discussion. Different classes and models can be similarly fit by fixing the time scores to different curves. This includes segmented linear regressions, power functions, sinusoidal functions and more.

In our syntax, the time scores for each class were provided by the appropriate function to the target estimates. The results of the function provided the time score for each response on for the given class. The linear class was given by

$$f(x) = x$$

The logarithmic class was given by

$$f(x) = \ln(x)$$

The two-anchors class was given by

$$f(x) = \ln\left(\frac{x}{100-x}\right)$$

The three-anchors class was given by

$$f(x) = \begin{cases} \ln\left(\frac{x}{50-x}\right) & x < 50 \\ \ln\left(\frac{x-50}{100-x}\right) & x > 50 \end{cases}$$

while we chose to not to include the three-anchors class in our paper, it is worth noting a latent growth model avoids the problem faced by Rouder and Geary (2014) of individual estimates crossing over the midpoint of the line because all scaling is done with the target numbers and not participants' estimates.

A further extension via a latent transition analysis could be attempted. This would be an ideal method for comparing group membership over the course of multiple measurement points; however, it may prove computationally complex. Future refinements may improve modeling efficiency allowing for such a model.

Lastly, alternative methods for entering covariates may be used. For instance, the 3-step process may be implemented in Mplus via the "auxiliary" command in the "variable" section. We explored this method for both the 2- and 3-class models for the 0–100 number line, and there were no changes to significant results.

PsychOpen GOLD