Modeling person-specific development of math skills in continuous time: New evidence for mutualism

Lu Ou
ACTNext by ACT, Inc.
500 ACT Drive
Iowa City, IA 52243
lu.ou@act.org

Abe D. Hofman
University of Amsterdam
Nieuwe Achtergracht 129-B
1018 WS Amsterdam
The Netherlands
a.d.hofman@uva.nl

Vanessa R. Simmering
ACTNext by ACT, Inc.
500 ACT Drive
Iowa City, IA 52243
vanessa.simmering@act.org

Timo Bechger
ACTNext by ACT, Inc.
500 ACT Drive
Iowa City, IA 52243
timo.bechger@act.org

Gunter Maris
ACTNext by ACT, Inc.
500 ACT Drive
Iowa City, IA 52243
gunter.maris@act.org

Han L. J. van der Maas
University of Amsterdam
Nieuwe Achtergracht 129-B
1018 WS Amsterdam
The Netherlands
h.l.j.vandermaas@uva.nl

ABSTRACT

In this study, we fitted a mixed-effects nonlinear continuous-time mutualism model of skill development proposed by van der Maas et al. (2006) to naturally collected irregularly spaced time series data from an online adaptive practice system for mathematics called Math Garden. Results showed that the mutualism model provided a better fit to the data than a g-factor model. The paper illustrates continuous-time modeling of irregularly-spaced multivariate time series data that are increasingly prevalent in modern learning systems.

Keywords

mutualism model, continuous-time model, mixed-effects model

1. INTRODUCTION

For the past century, generations of researchers have continued to pursue explanations for the consistent positive correlations between diverse sets of cognitive ability tests, known as *the positive manifold* [25, 29]. Heated debates went on about whether there is a potential biologically based g-factor that causes the development of general intelligence as well as the positive manifold [26, 7, 27, 11, 9]. Although researchers have not reached consensus, there is a shift from conceptualizing cognitive development as merely reflective, as in factor analysis, to thinking of it as formative [2, 14, 27]. In a formative model, the positive manifold is an emergent property that results from within-person changes and connections over time. This ontological stance implies that research needs to focus on understanding the causal relationships that underlie cognitive development to guide effective efforts to predict and intervene in students' learning.

Lu Ou, Abe Hofman, Vanessa Simmering, Timo Bechger, Gunter Maris and Han van der Maas "Modeling person-specific development of math skills in continuous time: New evidence for mutualism" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 372 - 377

Various mathematical representations encompassing continuous and discrete variables have been proposed to describe the mechanistic changes and sources of individual differences in cognition [28, 9, 32, 21]. From a developmental perspective, cognitive abilities develop as a dynamic system with reciprocal interactions between the elements of the system causing the developmental pathways of each of the elements [29]. In the *Mutualism* model of intelligence, elements of a system interact with each other in a collaborative way to achieve mutual benefits. This provides an alternative explanation for the positive manifold, other than the g-factor approach, and only requires sparse, weak, and even some negative interactions to produce positive correlations [28].

In the current study, we take advantage of massive time series data collected with an online learning environment for mathematics [12] and propose a method to fit the mutualism model to this dataset. We aim to examine potential reciprocal interactions of mathematical skills in two domains — counting and addition — in children's learning and practicing mathematics online. We build a model that takes into account individual differences in the learning processes by allowing individuals to start in different positions and by including random effects in key parameters of an otherwise group-based mutualism model. Note that this is the first application of the nonlinear mutualism differential equation model to empirical data, providing an evaluation of how well the theoretical account proposed by van der Maas et al. [29] can capture changes in children's mathematical skill development over time. In addition, we pioneer the use of continuous-time models to analyze irregularly-spaced data that arise when students use educational technology in realistic settings, and show that the estimation framework implemented in the dynr R package [19, 20] can handle nonlinear equations and mixed effects that explain both between-person and within-person differences.

In summary, the contribution of the work is three-fold: 1) providing new evidence of reciprocal interactions in mathematics skill development as a pioneer in fitting the nonlinear mutualism model to empirical data; 2) presenting a way to analyze the irregularly spaced multivariate time series data commonly seen in learning systems; and 3) demonstrating the use of state-space approaches in estimating parameters of mixed-effects dynamic models.

In the following sections, we first explain the mathematical model we use to characterize the learning processes, and the estimation procedure. We then present how the empirical data in the current study were collected, and the sample characteristics. The paper ends with a discussion of the results and their implications for education.

2. THE MUTUALISM MODEL

In biology, the term *Mutualism* is used for a relation between species populations where different species organically interact with each other to maintain sustainable growth [1]. Biologists routinely use the Lotka-Volterra model [16, 30] to study the dynamics of such relations, which inspired van der Maas and colleagues [29] to propose the same model, referred to as the mutualism model, to study the dynamics of cognitive development, where elements of a cognitive system interact with each other to achieve mutual benefit.

2.1 The Lotka-Volterra Model

Mathematically, the mutualism model can be expressed using generalized N-subject Lotka-Volterra equations as

$$d\boldsymbol{x}(t) = \boldsymbol{F}(x_1(t), x_2(t), \cdots, x_N(t))dt \tag{1}$$

$$= \left[\rho_i x_i(t)\left(1 - \frac{x_i(t) + \sum\limits_{i\neq j} a_{ij}x_j(t)}{K_i}\right)\right] dt, \tag{2}$$

where $i, j = 1, 2, \cdots, N$ indicates different elements of a dynamic system, and t is continuous time. Here, the elements are the counting and addition skills. The differential of vector $\boldsymbol{x}(t)$ with respect to t denotes the change in $\boldsymbol{x}(t)$ within an infinitely small time interval.

The model assumes logistic growth. The ρ_i are growth parameters that determine the steepness of the logistic growth function associated with each $x_i(t)$, and the K_i are the carrying capacity parameters that represent the limited resources in the system, such as limited attention and working memory one can allocate in learning. The a_{ij} are interaction parameters that specify the relations between each pair of x_i and x_j in development. With all $a_{ij} = 0$, the change of the latent variable $x_i(t)$ follows a simple logistic curve that converges to an equilibrium state of K_i, regardless of its starting position. The system is collaborative if the Jacobian matrix $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}$ is positive definite, and is competitive otherwise. If, for all i, $x_i(t)$ and ρ_i only take positive values, then it is possible to show that as long as the combined consumption of resources $x_i(t) + \sum\limits_{i\neq j} a_{ij}$ does not exceed the carrying capacity K_i, $x_i(t)$ will continue to increase to its equilibrium. Further, when the interaction parameters a_{ij} are negative (or $-a_{ij}$ are positive) for all $j \neq i$, $x_i(t)$ can develop even beyond the original carrying capacity K_i, as a benefit of the collaboration with the other processes. On the other hand, when the parameters $a_{ij}, j \neq i$ are positive, $x_i(t)$ can never reach the full potential K_i, as a loss due to competition. van der Maas and colleagues [29] showed that when $-a_{i,j}$ is positive and less than 1, the mutualism model can result in the positive manifold.

2.2 State-space Representation

If we take into account individual differences in the mutualism model, as well as process noise and measurement errors[1] that may occur alongside the manifestation of the mutualism process, we obtain a state-space representation of the mutualism model:

$$d\boldsymbol{x}_s(t) = \boldsymbol{F}_s(\boldsymbol{x}_s(t))dt + d\boldsymbol{w}_s(t) \tag{3}$$

$$\boldsymbol{F}_s(\boldsymbol{x}_s(t)) = \begin{bmatrix} \rho_1 x_{1,s}(t)\left(1 - \frac{x_{1,s}(t)+a_{12}x_{2,s}(t)}{K_1+b_{1,s}}\right) \\ \rho_2 x_{2,s}(t)\left(1 - \frac{x_{2,s}(t)+a_{21}x_{1,s}(t)}{K_2+b_{2,s}}\right) \end{bmatrix} \tag{4}$$

$$\boldsymbol{y}_s(t_{s,k}) = \boldsymbol{x}_s(t_{s,k}) + \boldsymbol{\epsilon}_s(t_{s,k}), \tag{5}$$

$$\boldsymbol{\epsilon}_s(t_{s,k}) \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon = \begin{bmatrix} \sigma^2_{\epsilon,1} & 0 \\ 0 & \sigma^2_{\epsilon,2} \end{bmatrix}\right), \tag{6}$$

where the subscript s indexes individuals, and $k = 1, 2, \cdots, T_s$ indexes the kth discrete person-specific measurement occasions $t_{s,k}$. The vector $\boldsymbol{x}_s(t)$ contains the latent counting and addition skills $x_{1,s}(t)$ and $x_{2,s}(t)$ for an individual s, manifested as $\boldsymbol{y}_s(t_{s,k})$ in a measurement model with serially independent Gaussian measurement errors $\boldsymbol{\epsilon}_s(t_{s,k})$. The differential of $\boldsymbol{x}_s(t)$ is determined by the systematic dynamic functions $\boldsymbol{F}_s(\cdot)$ and the differential of process noise $\boldsymbol{w}_s(t)$ that follows a Wiener process (i.e., a continuous-time version of random walk, [10]), with a diffusion matrix $\boldsymbol{Q} = \begin{bmatrix} \sigma^2_{w,1} & 0 \\ 0 & \sigma^2_{w,2} \end{bmatrix}$. Person-specific random effects $\begin{bmatrix} b_{1,s} \\ b_{2,s} \end{bmatrix}$ are added to the carrying capacity parameters, and are assumed to follow a normal distribution with mean $\boldsymbol{0}$ and a covariance matrix of $\boldsymbol{\Sigma}_b = \begin{bmatrix} \sigma^2_{b,11} & \sigma^2_{b,12} \\ \sigma^2_{b,12} & \sigma^2_{b,22} \end{bmatrix}$.

The initial condition, or the distribution of the variables at the first available time point, of the dynamic process $\boldsymbol{x}_s(t_{s,1})$ is assumed to follow a multivariate normal distribution with mean $\begin{bmatrix} \mu_{1,1} \\ \mu_{1,2} \end{bmatrix}$ and variance $\begin{bmatrix} \sigma^2_{1,11} & \sigma^2_{1,12} \\ \sigma^2_{1,12} & \sigma^2_{1,22} \end{bmatrix}$.

2.3 An Alternative G-factor Model

In order to explore the fit of the mutualism model to empirical data compared to the g-theory, a comparable state-space

[1] Process noise is distinct from measurement error in that the former is associated with random behavior in the underlying process, whereas the latter depends on the measurement process, device, and other environmental influences that may affect the accuracy of the measurements. In an educational context, a correct guess without knowing an item can be seen as a measurement error, while a child having a good or bad day could contribute to the process noise. Whereas measurement error does not influence growth at the next time point, the process noise does steer the dynamical system.

one-factor model without interactions can be developed as

$$dx_s(t) = \rho_1 x_s(t)\left(1 - \frac{x_s(t)}{K_1 + b_{1,s}}\right) + dw_s(t) \quad (7)$$

$$x_s(t_{s,1}) \sim N(\mu_{1,1}, \sigma^2_{1,11}), \boldsymbol{Q} = \left[\sigma^2_{w,1}\right], \boldsymbol{\Sigma}_b = \left[\sigma^2_{b,11}\right]$$

$$\boldsymbol{y}_s(t_{s,k}) = \begin{bmatrix}1\\ \lambda\end{bmatrix} x_s(t_{s,k}) + \boldsymbol{\epsilon}_s(t_{s,k}), \quad (8)$$

$$\boldsymbol{\epsilon}_s(t_{s,k}) \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon = \begin{bmatrix}\sigma^2_{\epsilon,1} & 0\\ 0 & \sigma^2_{\epsilon,2}\end{bmatrix}\right),$$

where the observed variables are linearly linked with the single latent variable through a loading matrix of $\begin{bmatrix}1 & \lambda\end{bmatrix}^\top$.

3. ESTIMATION

To estimate the random effects in the models, we augmented the latent variables $\boldsymbol{x}_s(t)$ with random effects $\boldsymbol{b}_s$ to yield a new latent variable vector, $\boldsymbol{x}^*_s(t) = \begin{bmatrix}\boldsymbol{x}_s(t) & \boldsymbol{b}_s\end{bmatrix}^\top$. We then modified the differential equations, the measurement model, and the initial condition to incorporate this change of $\boldsymbol{x}^*_s(t)$.

We used the dynr R package [19, 20] to estimate the parameters in the mixed-effects mutualism model, as well as the baseline g-factor model, by numerically optimizing an approximate log-likelihood function obtained as a by-product of the continuous-discrete extended Kalman filter [15]. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were constructed to compare models. Details of the estimation algorithms can be found in [4, 20].

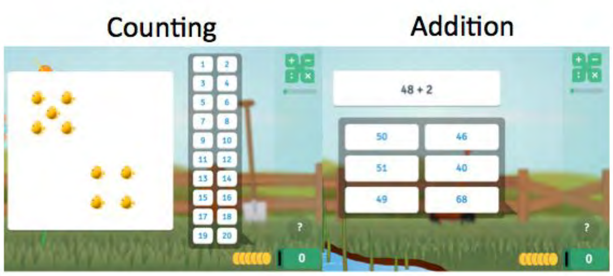

Figure 1: Screen shots of the counting and addition games in the Math Garden. Children give responses by clicking an option. The coins at the bottom disappear one per second, and reflect the scoring rule based on accuracy and response time.

4. EMPIRICAL STUDY

Here, we describe an application of the mutualism model.

4.1 Math Garden

We sampled data using a popular Dutch online adaptive practice and monitoring system called Math Garden [13]. The system consists of games that measure different mathematical skills, including counting and addition, as players practice their arithmetic skills through answering items. Figure 1 shows screen shots of two example items.

The system applies an explicit scoring rule for both speed and accuracy [17], visible to players as the number of coins they collect. For each item, a limit number of coins can be collected, and the number decreases by one at each additional second used to come up with the answer. In case of a correct answer, the score equals the remaining time. If the answer is incorrect, the score is the negative remaining time. The scoring rule takes speed-accuracy trade-off into account, penalizes quick but incorrect answers, and encourages thoughtful responses.

Skill rating and item difficulty are estimated on-the-fly using the Elo-algorithm [6] which was originally developed for chess competitions between two players, and now has been adapted for pairing a player with an item [13]. The skill and difficulty estimates for a player and an item are updated at each "match" they are involved in, depending on the weighted difference between observed and expected correctness, the latter of which is entailed by the measurement model [17]. Evidence has shown high validity and reliability of the skill and difficulty estimates [13].

In the current study, our observed data are the continuous end-of-day skill ratings in different domains, rather than binary correctness for each item. Comparisons of Math Garden's underlying measurement model [17] and the adapted Elo-algorithm [13] to other common models for binary responses in educational data mining — the Rasch model [23], additive factor models [3], performance factor analysis [22], and Bayesian knowledge tracing models [5] — are worth exploration, but beyond the scope of this paper.

4.2 Data Description

We selected a sample of children in grades 3–6, between the ages of 6 and 10 years old, who practiced counting and addition skills during at least 4 different months in the school year from September, 2016 to July, 2017, and had played at least 20 different days in each domain, with a minimum of 10 items per day. We excluded children whose parents indicated unwillingness to participate in Math Garden-related scientific research that was approved by the Ethics Committee of the psychology department of University of Amsterdam.

The resulting sample included a total of 2485 children, 51.07% male. The average age at which a child started to use Math Garden for practicing counting and addition during the school year was 7.23 years old (SD = 1.03). The original skill ratings could be negative, so we shifted them to the positive range by respectively adding 20 and 25 to the counting and addition scales. The over-time ebb-and-flow and variability of the skill ratings remain the same. From the second-by-second time stamps of the data points, we constructed continuous measures of time where each unit represents a week. Figure 2 shows the shifted skill ratings for three randomly selected individuals over time.

Distributions of the initial and ending skill ratings are plotted in Figure 3. At the first available time point for each individual, the counting skill ratings for all sampled children had a mean of 13.84 and a variance of 1.51, whereas the addition skill ratings had mean 12.94 and a larger variance of 10.56. At the last available time point for each individual, the ending counting estimates had a mean of 14.83 and variance of 1.49, while the ending addition estimates had a mean of 15.92 and variance of 9.05. Generally speaking, during the school year, more development is observed in the addition skill compared to the counting skill. There was more variability in children's initial and ending addition skill ratings

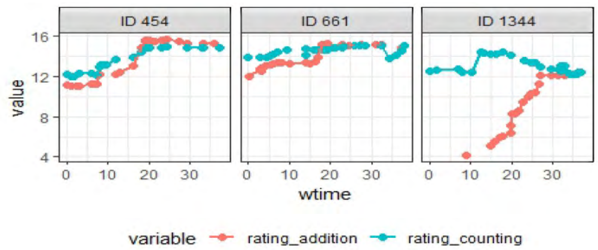

Figure 2: Individual time series data of counting and addition skill ratings for three children.

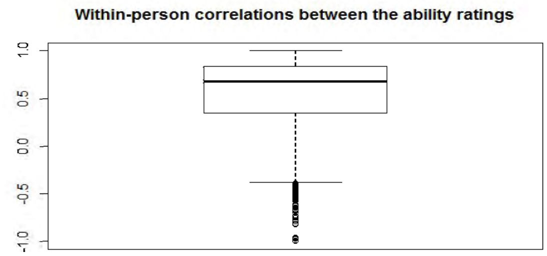

Figure 4: A boxplot of the within-person correlations of the addition and counting skill ratings.

than in the counting domain.

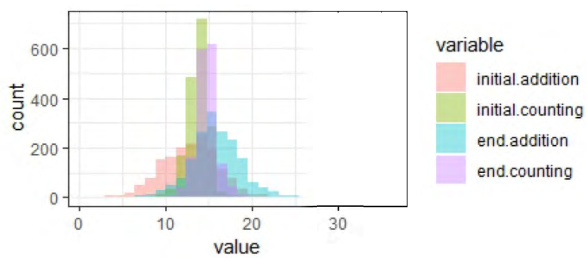

Figure 3: Histograms of the initial and ending skill ratings

The correlation between the initial addition and counting skill ratings was .73, and the correlation between the ending ratings was .79, confirming the positive manifold. Figure 4 shows the boxplot of within-person correlations between the skill ratings in the two domains. The mean of the within-person correlations was .55 (SD = 0.38). However, some negative values were observed at the significance level of 0.05 of the asymptotic p-values computed by the Hmisc R package [8]. For example, in Figure 2 the child with identification number 1344 had upward growth in the addition skill ratings and downward decline in the counting skill ratings. The downward decline may be due to an unexpected bump in the skill ratings that was higher than the child's equilibrium and hence resulted in a return to the equilibrium. Another possible explanation would be that, the child learned and practiced counting at school before addition, but forgetting took place as the child started to learn addition and practiced counting less. In such a case, there was a competition for attention and learning time between the skills in different domains, instead of a collaboration. Either case can be captured by the mutualism model.

The length of the individual time series of the counting skill ratings ranged from 20 to 177 days with a median of 29 days, and that of the addition skill ratings ranged from 20 to 192 days with a median of 42 days. The length of the interval between two time points represents the inactivity gap between two practice days of an individual for a mathematical skill, and ranged from 1 day to 18.29 weeks. The minimal gap length of a single time series had a median of 1 day across the sample in both domains, whereas the median maximum gap length was 4.86 weeks in the counting domain and 4 weeks in the addition domain. In Figure 5, the data of ten randomly selected individuals illustrate the irregularly-spaced measurement occasions, as well as the unbalanced practices in each domain on a single day and across time. The mutualism model assumes a continuous integrative process of change even though we do not have measurements of each skill at all times.

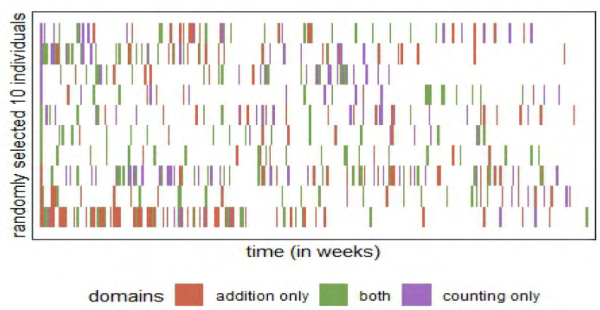

Figure 5: An illustration of the irregularly spaced time intervals of ten randomly selected individuals. Different colors represent the domains that an individual practiced during a specific day.

4.3 Empirical Results

The parameter estimates and model fit indices of both the mutualism model and the g-factor model were summarized in Table 1. All parameters were estimated to be significantly different from zero ($p < .05$). The estimates of the initial condition parameters ($\mu_{1,1}$, $\mu_{1,2}$, $\sigma^2_{1,11}$, $\sigma^2_{1,12}$, and $\sigma^2_{1,22}$) in both models were consistent with the sample mean and variance of the initial states. With lower AIC and BIC values, the mutualism model provided a better fit to the data compared to the g-factor model. Figure 6 shows the fit of the mutualism model to the observed data of four randomly selected individuals. The fitted trajectories were able to capture the changes of the observed paths for the individuals in both domains, suggesting a decent fit of the model to the data. In the mutualism model, the steepness parameters ρ_1 and ρ_2 were estimated to be close to zero, indicating that the overall development in skills was small and slow. The group-level equilibrium states K_1 and K_2, for when there was no interaction between the processes, were estimated to about 10, but individual differences captured by the random effects b_1 and b_2 contribute to an estimated co-variance of $\begin{bmatrix} 1.04 & 0.09 \\ 0.09 & 1.06 \end{bmatrix}$. Estimates of the interaction parameters a_{12} and a_{21} were found to be significantly negative, so the inter-

Table 1: Parameter estimates (standard errors) and model fit indices

	Mutualism Model	g-factor Model
ρ_1	0.08 (0.002)	0.02 (0.001)
ρ_2	0.09 (0.001)	
a_{12}	-0.48 (0.005)	
a_{21}	-0.58 (0.004)	
K_1	10.04 (0.005)	15.98 (0.124)
K_2	10.05 (0.004)	
$\sigma^2_{w,1}$	0.34 (0.002)	0.10 (0.001)
$\sigma^2_{w,2}$	0.42 (0.001)	
$\sigma^2_{\epsilon,1}$	0.29 (0.001)	4.92 (0.025)
$\sigma^2_{\epsilon,2}$	0.34 (0.002)	0.08 (0.001)
$\mu_{1,1}$	13.85 (0.006)	13.87 (0.013)
$\mu_{1,2}$	12.92 (0.001)	
$\sigma^2_{1,11}$	1.67 (0.014)	1.74 (0.068)
$\sigma^2_{1,12}$	1.34 (0.045)	
$\sigma^2_{1,22}$	10.55 (0.056)	
$\sigma^2_{b,11}$	1.04 (0.005)	1.71 (0.310)
$\sigma^2_{b,12}$	0.09 (0.004)	
$\sigma^2_{b,22}$	1.06 (0.005)	
λ		1.02 (0.001)
AIC	388981.81	472908.58
BIC	389155.46	472995.41

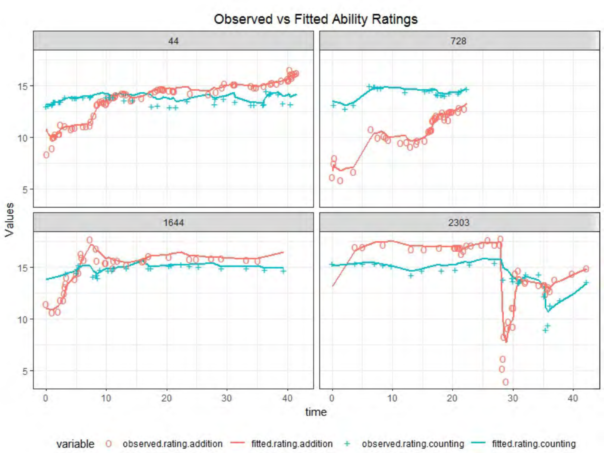

Figure 6: Observed and fitted skill ratings from the mutualism model.

actions between counting and addition ratings had a positive effect on their level changes. These results indicated that counting and addition skills collaborate, instead of competing, to form a positive manifold in the long run.

In summary, we have found beneficial interactions between children's addition and counting skill ratings as being better at one skill helps being better at the other. The mutualism model was a better fit to the data than the g-factor model. Individual differences are present in the data in both starting positions of the change trajectories and key model parameters that represent limited resources in the system, providing potential evidence for both the g-theory and the mutualism model of general intelligence, according to [29]. We concur with van de Maas and colleagues (2006) that individual differences cannot be ignored in educational applications.

5. CONCLUSIONS

In this paper, we presented a state-space expression of the continuous-time mutualism model proposed by [29] where individual differences, process noise, and measurement errors were taken into account. The mutualism model allowed us to tackle the underlying mechanism of the skill development from a micro perspective. We fitted the theoretical model to empirical data naturally collected online in authentic educational settings. Results showed that improvement in addition skill could positively influence the development in the counting domain, and vice versa. The better fit of the mutualism model to the data compared to the g-factor model suggested that the collaboration between the counting and addition skills in their co-development served as a better interpretation of the observed positive manifold.

The characteristics of the time series data in the current study are not uncommon in education as digital technology has transformed our way of collecting data about learning. The paper illustrates one way to fit dynamic models to the multivariate noisy irregularly spaced data that are rich in our real life. We appreciate the potential to apply the current method to different learning data to improve our understanding of cognitive and non-cognitive developments.

Nevertheless, this work has limitations that future work should aim to overcome. First, only two variables were considered in the current sample, while the mutualism model could be extended to multiple dimensions. The estimation algorithm is well suited for multivariate time series data, but the interpretation of the multivariate model can become complicated. Second, the estimation framework permits only a limited number of random effects in the current study [18]. In addition to the two carrying capacity parameters, one may be interested in adding random effects in the interaction parameters because of the potential competition between skills under time and attention constraints as we discussed above. The limitation of the estimation framework may be circumvented by utilizing sampling-based algorithms although they may be computationally heavy.

The fitting of the model to the data does not exclude other probable ways of interpreting cognitive development. Intervention studies with deliberate experimental designs are needed to establish causal relations in a dynamic system. These interventions may take the form of randomized assignment of skills to practice, for example, with groups of students assigned to practice only counting or only addition, but with progress measured on both skills after some period of practice. The cross-skill influence of practice can then be evaluated relative to practiced skill improvement.

Future work should also aim to evaluate how the mutualism account of skill development relates to other findings in education. For example, evidence suggests that interleaving practice on different problem types produces more robust learning and generalization than does blocking practice by problem type [31, 24]. It is possible that some of the benefit from interleaving relates to mutualism, with practice from different problem types influencing the development of the other skills.

6. ACKNOWLEDGMENTS

The work is supported by ACTNext by ACT, Inc.

7. REFERENCES

[1] M. Begon, C. R. Townsend, and J. L. Harper. *Ecology: from individuals to ecosystems*. Number Sirsi) i9781405111171. 2006.

[2] D. Borsboom, G. J. Mellenbergh, and J. V. Heerden. The Theoretical Status of Latent Variables. *Psychological Review*, 110(2):203–219, 2003.

[3] H. Cen, K. Koedinger, and B. Junker. Comparing two irt models for conjunctive skills. In B. P. Woolf, E. Aïmeur, R. Nkambou, and S. Lajoie, editors, *Intelligent Tutoring Systems*, volume 5091, pages 796–798, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[4] S.-M. Chow, L. Ou, A. Ciptadi, E. Prince, D. You, M. D. Hunter, J. M. Rehg, A. Rozga, and D. S. Messinger. Representing sudden shifts in intensive dyadic interaction data using differential equation models with regime switching. *Psychometrika*, 83(2):476–510, 2018.

[5] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[6] A. E. Elo. *The rating of chessplayers, past and present*. Arco Pub., 1978.

[7] G. E. Gignac. Residual group-level factor associations: Possibly negative implications for the mutualism theory of general intelligence. *Intelligence*, 55:69–78, mar 2016.

[8] F. E. Harrell Jr, with contributions from Charles Dupont, and many others. *Hmisc: Harrell Miscellaneous*, 2018. R package version 4.1-1.

[9] A. D. Hofman, R. A. Kievit, C. E. Stevenson, D. Molenaar, I. Visser, and H. L. van der Maas. The dynamics of the development of mathematics skills: A comparison of theories of developing intelligence. *OSF Preprints:https://doi.org/10.31219/osf.io/xa2ft*, 2018.

[10] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer New York, 1998.

[11] R. A. Kievit, A. Hofman, and K. Nation. Mutualistic coupling between vocabulary and reasoning in young children: A replication and extension of kievit et al.(2017). *Psychological Science*, 2018.

[12] S. Klinkenberg. High speed high stakes scoring rule: Assessing the performance of a new scoring rule for digital assessment. *Communications in Computer and Information Science*, 1(439):114–126, 2014.

[13] S. Klinkenberg, M. Straatemeier, and H. L. Van der Maas. Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2):1813–1824, 2011.

[14] K. Kovacs and A. R. A. Conway. Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, 27(3):151–177, 2016.

[15] G. Y. Kulikov and M. V. Kulikova. Accurate numerical implementation of the continuous-discrete extended kalman filter. *IEEE Transactions on Automatic Control*, 59(1), January 2014.

[16] A. J. Lotka. *Elements of Physical Biology*. Williams & Wilkins, Baltimore, MD, 1925.

[17] G. Maris and H. Van der Maas. Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, 77(4):615–633, 2012.

[18] L. Ou. *Estimation of nonlinear mixed-effects continuous-time models*. PhD thesis, The Pennsylvania State University, 2018.

[19] L. Ou, M. D. Hunter, and S.-M. Chow. *dynr: Dynamic Modeling in R*, 2017. R package version 0.1.11-5.

[20] L. Ou, M. D. Hunter, and S.-M. Chow. What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, 2019.

[21] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In P. De Bra, A. Kobsa, and D. Chin, editors, *User Modeling, Adaptation, and Personalization*, pages 255–266, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[22] P. I. Pavlik Jr., H. Cen, and K. R. Koedinger. Performance Factors Analysis — A New Alternative to Knowledge Tracing. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education, (AIED09)*, pages 531–538, Amsterdam, 2009. IOS Press.

[23] G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Studies in mathematical psychology. Danmarks Paedagogiske Institut, 1960.

[24] D. Rohrer, R. F. Dedrick, and S. Stershic. Interleaved Practice Improves Mathematics Learning. *Journal of Educational Psychology*, 106(4):1–9, 2014.

[25] C. Spearman. "General Intelligence," Objectively Determined and Measured Source : The American Journal of Psychology , Vol . 15 , No . 2 (Apr ., 1904), pp . 201-292 Published by : University of Illinois Press Stable URL : https://www.jst. *The American Journal of Psychology*, 15(2):201–292, 1904.

[26] L. L. Thurstone. Primary mental abilities. *Psychometric monographs*, pages ix–121, 1938.

[27] H. van der Maas, K.-J. Kan, and D. Borsboom. Intelligence is what the intelligence test measures. Seriously. *Journal of Intelligence*, 2(1):12–15, feb 2014.

[28] H. Van Der Maas, K.-J. Kan, M. Marsman, and C. E. Stevenson. Network Models for Cognitive Development and Intelligence. *Journal of Intelligence*, 5(2):16, apr 2017.

[29] H. L. van der Maas, C. V. Dolan, R. P. Grasman, J. M. Wicherts, H. M. Huizenga, and M. E. Raijmakers. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological review*, 113(4):842, 2006.

[30] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[31] Y. Weinstein, C. R. Madan, and M. A. Sumeracki. Teaching the science of learning. *Cognitive Research: Principles and Implications*, 3(2):1–17, 2018.

[32] M. V. Yudelson. Individualizing bayesian knowledge tracing. are skill parameters more important than student parameters? *EDM*, 2016.