Formalizing developmental phenomena as continuous-time systems: relations between mathematics and language development

Driver, Charles C ; Tomasik, Martin J

DOI: https://doi.org/10.1111/cdev.13990

Originally published at:
Driver, Charles C; Tomasik, Martin J (2023). Formalizing developmental phenomena as continuous-time systems: relations between mathematics and language development. Child Development, 94(6):1454-1471.
DOI: https://doi.org/10.1111/cdev.13990

DOI: 10.1111/cdev.13990

SPECIAL SECTION

CHILD DEVELOPMENT

Formalizing developmental phenomena as continuous-time systems: Relations between mathematics and language development

Charles C. Driver | Martin J. Tomasik

Institute of Education, University of Zurich, Zurich, Switzerland

Correspondence
Charles C. Driver, Institute of Education, University of Zurich, Zurich, Switzerland.
Email: charles.driver2@uzh.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung

Abstract

We demonstrate how developmental theories may be instantiated as statistical models, using hierarchical continuous-time dynamic systems. This approach offers a flexible specification and an often more direct link between theory and model parameters than common modeling frameworks. We address developmental theories of the relation between the academic competencies of mathematics and language, using data from the online learning system `Mindsteps`. We use ability estimates from 160,164 observation occasions, across N=4623 3rd to 9th grade students and five ability domains. Model development is step-by-step from simple to complex, with ramifications for theory and modeling discussed at each step.

Theories help organize concepts and define constructs and are particularly valuable because they *represent* systems or phenomena of interest (Swoyer, 1991). This representation allows us to reason using the theory to make predictions about the phenomena, and if the theory adequately represents the phenomena (and our reasoning is sound), our predictions will be accurate. Theories do not need to be highly complex and detailed to be useful—much usefulness is because theories are simpler to work with than the phenomena themselves. Robinaugh et al. (2021) describe learning to navigate Paris by map, which contains enough detail to locate oneself, but not so much that it is difficult to find points of interest. Rather than such maps, many theories in developmental science are more like faded signposts and approximate directions—clearly relevant and provide some guidance, but lacking the detail to make clear predictions. The imprecision of natural language is posed as one cause of this vagueness, as many theories are expressed only in words (Robinaugh et al., 2021; Smaldino, 2017). In contrast, through their use of a more precise language, formal mathematical theories facilitate more precise predictions. Moreover, formal theories help organize and integrate empirical observations made in different contexts and help generate novel predictions of unobserved phenomena.

However, on their own, formal theories are not necessarily easy to work with, and theoretical claims are usually not falsifiable at the broader framework or theory levels. For practical purposes, theories need to be expanded with auxiliary assumptions and instantiated in software. Computational modeling, wherein vague, verbal, descriptions of ideas are developed into formal mathematics and software (Guest & Martin, 2021), offers a clear and powerful approach to linking theory, data, and inferences. Being forced to explicitly state how data arise from a phenomenon of interest can help scientists understand the repercussions of their ideas.

Borsboom et al. (2021) outline a process of theory construction centered around modeling: (1) phenomena of interest are identified, (2) a "prototheory" is developed using broad theoretical principles that explain the phenomena, (3) the prototheory is used to construct a mathematical model encoding these principles, (4) the adequacy of the model is assessed by checking whether it reproduces the phenomena of interest, and (5) the adequacy of the theory is assessed by evaluating whether the phenomena are reproduced faithfully, and whether the explanatory principles are sufficiently parsimonious and plausible.

In terms of exactly how formal theories may look and how they can be instantiated, Vallacher and Nowak (1997) and van Geert (1994) offer extended discussions on the value of nonlinear dynamic systems and links to theorizing in psychology. Such approaches offer the flexibility needed to instantiate theories of complex processes and interactions developing over time. Many works address the topic (e.g., Boker, 2012; Haslbeck et al., 2022), yet the actual process of instantiating theory in dynamic system equations, and using the equations as a vehicle to assess and improve theory, remains challenging. This work aims to familiarize researchers with the hierarchical continuous-time dynamic systems approach, which sometimes offers a closer link between theories and model parameters than the standards of mixed-effects regression or structural equation modeling. This work is primarily intended as a tutorial on instantiating verbal theories of development as continuous-time dynamic systems models of varying complexity. Considering the theory development stages posed by Borsboom et al. (2021), this work focuses on the third stage, in that we describe how multiple prototheories may be turned into a statistical model. However, we also anticipate somewhat the fourth and fifth stages of model and theory checking, by including a range of known and expected phenomena in the model that were not explicitly described by the theories and also by building and fitting the model step-by-step and considering relations to theory at each point. By taking seriously sources of confusion when interpreting parameters and inherent limitations in the formalization and data, we hope such an approach to "initial formalization" can be fruitfully employed in theory development procedures such as posed by Borsboom et al. (2021) and Haslbeck et al. (2022). At the same time, we leverage a novel data source and approach for considering competence development in mathematics and language, providing a basis for more developed empirical works.

DEVELOPMENTAL RELATIONS BETWEEN MATHEMATICS AND LANGUAGE

Two hypotheses, that of the "medium function" and that of the "thinking function," are currently considered to explain the consistent finding of positive correlation between mathematics and language development (Bailey et al., 2020). The *medium function hypothesis* (Bruner, 1966; Cohen & Dahaene, 1995; Fetzer & Tiedemann, 2018) posits language as a tool for communicating mathematical concepts and building and retrieving mathematical knowledge from long-term memory. Accordingly, students with good language skills will not only profit more from instruction and hence grasp mathematical concepts more easily but also retrieve mathematical knowledge more easily from memory, and communicate this more competently.

In contrast, the *thinking function hypothesis* (Daneman & Merikle, 1996; Lombrozo, 2006; Peng et al., 2018) posits higher-level cognitive functions underlying both language and mathematics performance, similar to *g*-factor in research on intelligence. Indeed, research shows a strong overlap between cognitive processes underlying both (Cirino et al., 2018; Korpipää et al., 2017).

Drawing from the broader developmental literature, the idea that competence development in language and mathematics could show a competing relation arises from general theories of development such as selection, optimization, and compensation (Baltes, 1987; Freund, 2008). The selection of developmental pathways, contexts, and goals aims to effectively invest resources in the most viable directions, driving age-related increase in specialization of competencies and interests (Hofer, 2010; Köller et al., 2001). We include this notion as a third theory, which we will refer to as the *specialization hypothesis*. Specialization may explain some changes in the relation between competencies with age, given higher demands on students as they progress through schooling and the necessity of focusing on specific areas (van Geert, 1991).

Four recent reviews or meta-analyses investigated competence development in language and mathematics. Chow and Jacobs (2016) consider four works on the influence of language on fraction outcomes in school-age children, all suggest that language contributes to fraction performance. de Araujo et al. (2018) synthesize 75 qualitative studies focusing on multilinguistic educational contexts and foreign language students. Here language is assumed as a prerequisite for mathematical competence development, and the authors conclude that it may be easier for foreign language students to learn mathematics in their native language. Koponen et al. (2017) investigated the association between rapid automatized naming (a component of phonological processing) and mathematics performance, finding a significant correlation ($r = .37$). Almost half of the 33 studies reviewed were longitudinal, but the authors did not investigate the direction of effects. Finally, a meta-analysis by Peng et al. (2020) of 344 studies examined whether cognitive function could explain relations between language and mathematics. They found an overall relation between language and mathematics of $r = .42$, comparable in magnitude with the meta-analysis by Koponen et al. (2017). Working memory alone accounted for 8%–16%, and general intelligence alone accounted for 21%–23% of variance in the relation between language and mathematics, providing some evidence for the thinking function hypothesis. The longitudinal studies included were used to address the direction of effects between language and mathematics performance. There was an average path from language to mathematics of $r = .20$, whereas the reversed path from mathematics to language was $r = .22$—quite comparable in magnitude. Age was also investigated (among other variables) as a moderator of this relationship, but generally did not moderate between-construct relations. We have discussed three theories describing how mathematics and language performance may relate developmentally. The thinking-function hypothesis poses a common cause behind the two, the medium-function hypothesis poses supportive relations from language to mathematics, and specialization would suggest that learning in either domain may detract from the other, and that this may increase with age. Meta-analytic results appear mixed and inconclusive with respect to distinguishing between the theories, and indeed, given their vague specification, they may all be plausible to some degree.

EMPIRICAL RESEARCH AGENDA

Because of their vague specification and inconclusive empirical results, we do not propose any strong test of the three theories. For similar reasons, we will not and cannot build models based purely on the theories in question. Instead, we will build and fit models of competence development using a general approach that can encapsulate both the theoretical aspects under consideration and the necessary auxiliary model elements (e.g., age trends, individual differences) and can be easily expanded or adjusted as needed. Our aim is to provide indications of theoretical support or debunking, the potential to integrate the theories in a formal model, a basis from which additional models and/or data may be included, and an example of specifying and interpreting broad theories in a continuous-time dynamic systems framework.

The data come from the `Mindsteps` online learning platform (see https://www.mindsteps.ch/). Mindsteps offers practice and tests with questions drawn from a bank of many thousands across a range of subjects and is used by teachers and students in a variety of Swiss regions. The system covers topics from the third grade in elementary school until the third grade in secondary school, spanning 7 years of compulsory schooling. Currently, the item bank comprises up to 15,000 items per school subject.

In `Mindsteps`, there are two types of item banks. A practice item bank is available to all students and teachers for training and teaching. Students can use this item bank to create and answer an item set from a topic domain on which to practice. A testing item bank is used to evaluate students' abilities. Teachers can select items according to desired competency domains or topics of the curriculum, and create assessments for students.

For our analyses, we use data from both item banks. Each assessment consisted of at least 10 items. In an initial modeling step, data from the years 2018 to 2022 was fit to multiple large unidimensional item response theory-based models, and individual ability scores for each assessment session were extracted. More details on the `Mindsteps` software, as well as a discussion of an earlier iteration of the modeling approach, are provided by Tomasik et al. (2018) and Berger et al. (2019). For pragmatic (i.e., computational) reasons, we subset available data to students with at least 10 assessments (in any of five domains), who have used the system for at least 1 year, and for whom at least four out of five domains have been assessed at least once. This left us with $N=4623$ students and 160,164 observation occasions in total. We use variables from the two German language domains, reading and grammar, and three mathematics domains, "forms and space" (*MATHS1*), "measures, functions, and probability" (*MATHS2*), and "numbers and variables" (*MATHS3*).

CHALLENGES IN BRIDGING THEORY AND EMPIRICAL RESEARCH

Many of the studies that have investigated the topic cannot clearly distinguish between different theories of competence development in language and mathematics, partly due to limitations in the models employed. Such a disconnect between theory and formal mathematical specification is not uncommon and does serve some function—uncritical eagerness to specify complex models can lead to an unwitting dependency on auxiliary assumptions of the models, generating overconfidence in either uncertain or wholly wrong inferences. In contrast, without explicit formal specification, language can be vague and interpreted in many ways. This means that certain theories may sound perfectly sensible and garner apparent support from a range of studies, yet offer little to no predictive value for new circumstances.

However, even with a formal mathematical specification of theory, there is yet another gap to be bridged. As Fried (2020) points out, common statistical modeling approaches often do not allow for a direct connection to theories. This latter gap often receives even less consideration than the former, and modeling approaches sometimes appear to guide and limit the variety of formal theories considered. A related but distinct issue concerning longitudinal modeling is that statistical models of change are usually not formulated in terms of direct effects, but in terms of expected change over some length of time—making it difficult to link model parameters to theoretical concerns. The difficulties of theoretical vagueness, model inflexibility, and linking model parameters to theory, can all be seen in the theories of mathematics and language skill development we are considering.

Regarding theoretical vagueness: If there is some truth to the medium function, then improvements in language will result in improvements in mathematics—but *when*? How long should it take for gains in language to result in gains in maths? Considering the proposed mechanism of action behind the theory we can develop some expectations; for example, if it is really the case that language facilitates *retrieval* of mathematical knowledge, this suggests to us that resulting gains should be near instantaneous. On the other hand, to the extent that language facilitates *learning* mathematics, there needs to be time for improved learning to take place, before any gains in mathematics performance are realized—though how long this should take is not clear.

Considering the lack of model flexibility critique while staying with the medium function hypothesis: it is not explicitly stated whether the theory should apply equally to students at different levels of performance. It seems almost self-evident that *some* fundamental abilities to communicate are a prerequisite for learning mathematics in a classroom, but should we expect comparable gains for university students studying high-level literature? Most modeling frameworks can accommodate some forms of heterogeneity, but certain features tend to be much harder to handle. One such difficult-to-model feature is continuous (i.e., not group-based) differences in complex parameters such as those in a multivariate correlation (see, e.g., Driver & Voelkle, 2018), but these can be useful for representations of development, as theoretical components can become more or less correlated with age.

The final reason we focus on the disconnect between theory and statistical models is that of direct effects and linking parameters to theory. Three seemingly innocuous statements are key to this: First, theories tend to be formulated in terms of direct causal mechanisms between constructs—the medium function hypothesis, for instance, suggests that language plays a *causal* role in mathematics performance. Second, constructs tend to be thought of as continuously in existence—we assume that a person has an ability in maths, whether or not we measure it. Third, constructs are continuously subject to all manner of influences—a person's maths ability is subject to all known and unknown influences (e.g., cognitive maturation, schooling), and these take effect independent of how often we observe the individual. These points might all seem obvious, yet modeling approaches regularly used to represent developmental theories involve assumptions that run counter to these statements. A core assumption built into common multivariate modeling frameworks is that of *discrete time*, wherein constructs (typically) change and interact only at moments when an observation occurs. Such discrete-time approaches can be perfectly viable in terms of the predictions they offer, yet problematic when the parameters are used for inference about underlying causal mechanisms, as is typical of interest when psychological theories are being considered. While there are works that go into much greater detail on this (Aalen, 1987; Aalen et al., 2016; Deboeck & Preacher, 2016; Driver, 2022; Kuiper & Ryan, 2018; Ryan & Hamaker, 2021), as an example, consider the following: Assuming that maths and language ability are continuous-time processes (i.e., continuously existing and interacting); there is a causal effect of language on maths (i.e., medium function); the two processes are subject to *independent* causes; we only measure abilities once per year. When common representations, such as a cross-lagged panel or vector-autoregressive models, are applied to such a case they are likely to find correlated fluctuations in the two processes that are *not* explained by directed effects (Driver, 2022). This could lead some to the conclusion that there are common-causes operating between the two, when in fact there are none. Such erroneous conclusions easily arise when thinking in terms of direct causal influence, but using models that parameterize change over some time interval.

The three challenges discussed—theory vagueness, model inflexibility, and linking model parameters to theory—draw us in somewhat different directions. Given vague theories and worries about inflexible models, one could be tempted to limit formal or theory-based modeling and pursue data-driven approaches to simply describe observed patterns. However, such a direction all but ensures that model parameters themselves have little obvious meaning, and the link to theoretical concerns is more difficult or impossible to deduce. In contrast, going the other direction to a single statistical model that most closely represents the verbal theory in question will likely result in invalid or overconfident inferences, as there is so much uncertainty regarding statistical model choice. In this work, we approach the problem from somewhere in the middle and describe the use of a class of statistical models that is capable of representing psychological theories in a direct causal sense, while flexible enough to accommodate the fact that we know very little about the process dynamics and our models may need further updating or complexity.

LINKING THEORY AND STATISTICS WITH CONTINUOUS-TIME DYNAMIC SYSTEMS

The most common approaches to questions of change in developmental psychology are probably multilevel regression and structural equation modeling. While one can usually find ways to get any theory and data combination into a particular modeling approach, when the modeling approach is not a good fit for the theory and data, theories may be altered to suit the modeling, unnecessary auxiliary assumptions may be required, or even, as we described with the direct effects example above, the theory that was thought to be examined is unknowingly not! A hierarchical continuous-time dynamic systems perspective allows for a relatively direct link between parameters and typical developmental theories, reducing issues related to misfits between theory and model. A continuous-time approach helps to address the issue of linking theory and parameters because the temporal effects are direct rather than a lagged aggregate—more on this below. The broad concern of model flexibility, and the capacity of the model to represent varied hypotheses, is also easily addressed in a continuous-time framework—software is available that can specify a broad array of linear or nonlinear systems, measurement models, and heterogeneity across time, individuals, or contexts. The concern regarding theory vagueness can, of course, not be rectified by any modeling approach. Nevertheless, a modeling framework that is otherwise more appropriate may be more likely to lead the modeler to the right questions for reducing vagueness and clarifying theory.

The core of a continuous-time system is a stochastic differential equation. Differential equations are a mathematical language for continuously changing processes, and the "stochastic" addition simply allows for uncertainty in the direction of change. Such models have been discussed in the social sciences since Coleman (1964), with further development closer to psychology by Singer (1993), Oud and Jansen (2000), Boker (2012), Chow (2019) and Voelkle et al. (2012), among others. Some software packages for stochastic differential equation modeling in the social sciences include `ctsem` (Driver et al., 2017), `dynr` (Ou et al., 2019), OpenMx (Neale et al., 2016), and `BHOUM` (Oravecz et al., 2016). We will describe models in a general form, but in some cases include additional labeling relevant to the `ctsem` software. Code for model fitting with `ctsem` plus additional results are available in the supplementary material.

Discrete-time processes

To understand the continuous-time systems framework, it can be helpful to start from the discrete-time perspective. Most models that address coupled and fluctuating processes over time, such as vector autoregressive, cross-lagged panel, and multivariate change-score models, contain, or can be re-written to contain (e.g., Voelkle & Oud, 2015), equation components for modeling the processes that look like this:

$$\eta_u = \mathscr{A}\eta_{u-1} + \mathscr{B} + \mathscr{G}\mathbf{z}_u \quad \mathbf{z} \sim N(0,1), \tag{1}$$

where $\boldsymbol{\eta}$ is a d length vector of process values (which might be observed data or hypothetical latent states), u indexes measurement occasion, $\boldsymbol{\mathcal{A}}$ is a matrix of temporal regression coefficients, $\mathscr{B}$ is an intercept, and $\boldsymbol{\mathcal{G}}$ is the effect matrix of the d length system noise vector $\mathbf{z}$, where z contains independent and identically distributed deviations with zero mean. Typically, $\boldsymbol{\mathcal{A}}$ is where we would expect to see parameters reflecting directed influences between processes (i.e., psychological constructs) that are included *in* the model, and $\boldsymbol{\mathcal{G}}$ is where we would find the influence of fluctuations in unknown elements that are *not* included in the model. The intercept $\mathscr{B}$ sets the trend (which is in some cases flat) and can reflect the influence of unchanging (or very slowly changing with respect to our observations) external elements. In the case of observing mathematics and language performance, their influence on each other would likely be found in $\boldsymbol{\mathcal{A}}$, while the effect of fluctuations in unmodeled constructs such as motivation would appear in $\boldsymbol{\mathcal{G}}$. For a recent overview contrasting different approaches to cross-lagged models and software, see Ruissen et al. (2022).

Continuous-time processes

In discrete-time systems, the temporal effects represent regression strengths between two points in time. When the processes we are interested in fluctuate and interact continuously (or near continuously), discrete-time regressions represent an aggregated effect over the time interval from all sources, rather than a direct effect. Consider a system where motivation to exercise causes exercise, and exercise causes fitness. If we use a discrete-time approach to examine this system with intervals of 1 year between measurements, then we will also obtain a positive cross-effect of earlier motivation on later fitness—even though achieving increases in fitness requires actual exercise, rather than simply motivation! Nothing is inherently "wrong" here, except that it is relatively common for researchers to assume that such cross-effects represent mechanistic causal relations between the variables. Driver (2022) discusses this example in more detail, and other works also address the topic (Aalen et al., 2016; Deboeck & Preacher, 2016; Kuiper & Ryan, 2018; Ryan & Hamaker, 2021). Continuous-time approaches allow for parameters that represent direct relations between changing constructs over time, thereby allowing more direct mapping between theories (which often involve direct relations) and parameters. In the exercise example, the continuous-time temporal relations matrix would always contain a zero for the effect of motivation on fitness, regardless of the time interval at which observations were made, allowing for a genuine test of the hypothesis that exercise mediates the effect of motivation on fitness.

One can intuitively think of continuous-time approaches as similar to discrete-time, but simply compressing the time interval to a "very small" value. A continuous-time form of the vector autoregression discussed is

$$d\boldsymbol{\eta}(t) = (\mathbf{A}\boldsymbol{\eta}(t) + \mathbf{b})dt + \mathbf{G}d\mathbf{W}(t). \tag{2}$$

This looks similar to the discrete-time form, but instead of telling us the new value of latent processes $\boldsymbol{\eta}$ given one step forward in time, it tells us how $\boldsymbol{\eta}$ is changing *at the moment*. Some complications due to stochastic differential equations are present. First, the dt on the right-hand side can be thought of as a very small step in time. Second, the $d\mathbf{W}(t)$ represents Gaussian white noise in continuous time and enters the system via the $\mathbf{G}$ matrix. We will at times refer to the

$\mathbf{G}d\mathbf{W}(t)$ in combination as the *system noise*. This represents all the changes in our latent processes $\boldsymbol{\eta}$ that are not accounted for by other parts of the system model. Ideally, this would capture short-term, unpredictable fluctuations such as changes in study habits due to exams leading to increased competence. However, allowing for such system noise also allows the model to dynamically compensate somewhat for the misspecification of other components, such as neglecting to model a trend when we know competencies increase with time. In such a case, reasonable forward predictions may still be possible, but parameter interpretation is confounded.

Because **A** now represents the influence of the current state of the system on the *direction of change* in the system, its interpretation differs somewhat from the discrete-time form. In continuous-time, both the cross (off-diagonal) and auto (diagonal) effects in the temporal effects matrix offer the same interpretation—when the effect is positive, higher values of the causal (column) variable lead to rises in the caused (row) variable, and vice versa. To understand the temporal effects interpretation, and how parameters in the matrix may relate to our theoretical interests, consider the example:

$$\mathbf{A} = \begin{array}{c} \\ \text{language} \\ \text{maths} \end{array} \begin{array}{c} \begin{array}{cc} \text{language} & \text{maths} \end{array} \\ \left[\begin{array}{cc} -0.5 & 0 \\ 0.2 & -0.8 \end{array} \right] \end{array}.$$

Here, we see that the diagonals, containing the auto-effects, are both negative. This means that whenever language or maths performance rises *above* its baseline trend (determined by other model elements), a *downwards* force pushes performance back toward the baseline—the system stabilizes itself. That maths has a more negative auto-effect indicates that fluctuations away from the baseline dissipate *faster*, as there is more force pushing deviations back to the baseline. We can crudely approximate how fast this dissipation is by calculating the rate of change at one moment and assuming that rate is constant for some length of time: If maths is 1.0 at time 0, then the rate of change in maths at that point (based only on the auto-effect for now) is $-0.8 \times 1 = -0.8$. Assuming a constant rate of change over 1 unit of time, we have a change of $1 \times -0.8 = -0.8$, leaving us with 0.2, or 20% of the initial change—also understandable as an auto-regression strength of 0.2 after a time interval of 1. More precise results can be obtained by taking multiple smaller steps in time or using an "exact" solving approach. The basic component of the exact approach is the matrix exponential, with temporal regression coefficients for particular time intervals given by $e^{\mathbf{A}\Delta t}$ (the full solution is described in Driver & Voelkle, 2018).

The **A** matrix is often of particular interest because it is generally where causal relations, or at least within-person temporal relations, are to be found—and indeed, it is relevant in the case of our three theories. The example **A** matrix shown above depicts a case where, as is highly typical, each process influences itself, these are the auto-effects on the diagonal already described. Then there are the cross-effects: The 0 in the top right corner reflects that maths ability (the column variable) does not influence language (the row variable), whereas the 0.2 in the lower left reflects a positive influence of language on maths—this would be consistent with the medium-function hypothesis, wherein language supports mathematics learning. The thinking function hypothesis, in which maths and language arise from higher-level cognitive functions, would also lead to patterns in the **A** matrix, but *only* if we had an appropriate cognitive function process in the system. In that case, the **A** matrix might look something like this, where we see that there are two cross-effect coefficients of 0.4 from cognition to each of maths and language:

$$\mathbf{A} = \begin{array}{c} \\ \text{language} \\ \text{maths} \\ \text{cognition} \end{array} \begin{array}{c} \begin{array}{ccc} \text{language} & \text{maths} & \text{cognition} \end{array} \\ \left[\begin{array}{ccc} -0.5 & 0 & 0.4 \\ 0 & -0.8 & 0.4 \\ 0 & 0 & -0.1 \end{array} \right] \end{array}.$$

Because we, unfortunately, do not have a cognition variable in the current data set, we would expect to see the influence of this and any other unmeasured common causes show up in other model elements, or somewhat confusingly, in different elements of the **A** matrix. A fast-changing (relative to our observation frequency) common cause could be found in the correlations or covariances given by the system noise effect matrix **G**, where the covariance of the system noise is $\mathbf{G}\mathbf{G}^{\top}$. For our purposes, the thinking-function hypothesis could lead us to expect a positive correlation in the system noise, to the extent that cognitive performance fluctuates unpredictably over our time span of assessment. It may also be that changes in cognitive performance are so slow and smooth that the common-cause effect instead shows up only or primarily in stable individual difference effects, discussed below. The trickiest circumstance is somewhere in the middle, where a common-cause changes too slowly to be well represented by white noise, but too quickly to be well captured as a stable individual difference. In such a case *both* off-diagonals of the **A** matrix could show (more) positive effects. Attempts to mitigate this with model sophistication are possible, but awareness of the possibility when making inferences is crucial.

The system noise correlation may also be relevant when considering the specialization hypothesis in which time spent improving one domain predicts reduced growth in the other domain. In this case, there is good reason to expect that there may be short-term fluctuations, for instance, before an exam on a particular subject more study in that subject may occur. Again, however, we cannot be sure of the time scale of such fluctuations, should they exist. It may be that specialization appears as a lowering of the system noise correlation, but if study patterns have some temporal consistency, it may also appear as more negative values in the cross-effects of the **A** matrix—time spent studying for maths leads to reduced growth or decline in language and vice versa.

Measurement model

To represent and estimate theoretical constructs in a statistical model of observed data, it is important to account for how the constructs are measured. Some constructs may be perfectly represented by available observations, others may be measured but imperfectly, while others may only be defined based on relations between latent constructs—with no direct observations at all. In our case, we have observed variables that we believe roughly represent math and language competence, but we do not want to assume that these observations are perfect, and we wish to combine multiple observed variables (e.g., from three maths subdomains). Additionally, in order to understand and model, our theoretical constructs of math and language competence better, we will (eventually) break each construct down into multiple latent processes, some reflecting long-term change with others capturing short-term change. This is all accomplished through the use of a measurement model. Regarding imperfect observations, observed variables are usually contaminated by factors that are extraneous to the construct of interest. If these other factors are not accounted for, all manner of problems can occur in the estimated relations at the process level. Commonly, it is assumed that these additional factors are numerous and varied enough such that their influence increases variance but does not bias results. Ignoring such *measurement error* in longitudinal models can lead to downward-biased autoregression coefficients and potentially spurious cross-regression relationships (Schuurman & Hamaker, 2019), as well as inducing a dependency wherein the apparent time scale of the processes (inferred from estimates) can depend strongly on the frequency of observation (Driver, 2022). It should also be noted that observations always occur in discrete time, in that there are gaps in time between observations—the continuous-time aspect is that we use such observations for inference about an underlying continuous-time process.

A state-space modeling approach, as used in this work, couples a latent system process model with a measurement model. This feature is shared with structural equation modeling, and indeed the two share many similarities (Chow et al., 2010; Oud & Jansen, 2000). The state-space representation shows a general state of the system and how this changes, rather than laying out every time point, and can be simpler to work with when there are many time points. These approaches allow for the variance of observations to be composed into two (or more) portions. A measurement error portion contains variation that is essentially discarded from one observation to the next, while the system noise component contains variation that, although it could not be predicted by the deterministic portion of the model, is nevertheless informative for predicting future states.

The classic linear factor model, widely used in state-space and structural equation modeling, offers a convenient and tractable approach to account for measurement error. Moreover, it allows for the possibility of combining multiple noisy measurements to improve the precision of estimated constructs, and vice versa, for one observation to arise due to multiple latent constructs.

The linear factor model for the observed variables in a system is

$$\mathbf{y}(t) = \boldsymbol{\Lambda}\boldsymbol{\eta}(t) + \boldsymbol{\tau} + \boldsymbol{\epsilon}(t) \quad \text{where}\, \boldsymbol{\epsilon}(t) \sim \mathrm{N}\left(\mathbf{0}_c, \Theta\right), \tag{3}$$

where $\mathbf{y}$ is the c length vector of observed variables, $\boldsymbol{\Lambda}$ contains the factor loadings, and $\boldsymbol{\tau}$ observation level intercepts. The manifest residual vector $\boldsymbol{\epsilon}$ has covariance matrix $\boldsymbol{\Theta}$, which captures the variation in the observed variables that do *not* aid in predicting future states of the system, which is often thought of as a measurement error.

Individual differences

Stable individual differences, between-person effects, or sometimes simply "heterogeneity"—since Molenaar (2004) the field has gradually become more aware of the need to distinguish within-person change from stable characteristics that differ among individuals. Often though, this distinction is thought of in a black-and-white manner, as though there is some true individual difference, and then what is left is evidently within-person change. However, this split is highly dependent on both the data that are available and the model that is used. An example of data dependency is when what initially looks like a stable difference between individuals fluctuates substantially when examined at a longer timescale. Model dependency can arise via inadequate models, and an example of such can be seen in the typical pattern of heterogeneity in linear growth-curve models of learning, where a higher initial value correlates with a lower slope. Such slope differences are very often not because of some innate or stable individual difference, but because the true growth function is nonlinear, such that improvement gets more difficult and slower as performance rises—so those who start the assessment period at a higher level tend to show reduced growth (https://cdriver.netlify.app/post/heterogeneity/ elaborates on this example). Considered in light of such phenomena, individual differences in statistical models may best be understood as a tool to represent *differences in the ideal model parameters between subjects, conditional on the specified model and data*.

To estimate such individual differences, there are a variety of approaches: Given sufficient data, the process and measurement model may simply be fit separately for each individual, resulting in unique parameter estimates for every subject. Large and informative enough data sets for doing this with complex models are still quite rare, so approaches that assume some sort of similarity across subjects, thereby increasing the precision of parameter estimates, are helpful. Such approaches can assume certain parameters are (a) fixed across all subjects, (b) vary as a function of some observed covariate(s), and/or (c) vary according to some underlying distribution. For (b), the term "fixed effects" is often used, and for (c), "random effects," though the use of such terms can be heterogeneous and confusing across fields.

In linear models such as those we have described for the theoretical processes and measurements, individual differences in intercept type terms (i.e., initial values, continuous process intercepts, observation intercepts) are generally the most straightforward parameter types to model as random effects. Hamaker et al. (2015) detail the importance of this for a discrete-time structural equation model, whereas Oud (2002) demonstrates this for a continuous-time state-space form. Such an approach, wherein most parameters (such as temporal dynamics, factor loadings, measurement error variance, etc.) are assumed to take the same value across individuals, can offer a reasonable and tractable "first-pass" type model. The assumption that non-intercept type parameters do not differ over individuals is generally very strong, however, and it is often worth at least including a check for substantial between-person differences, whether or not there are theoretical concerns related to such differences.

Approaches to estimate individual differences in parameters such as a standard deviation or correlation can be complicated because only certain values for such parameters result in a valid model. In `ctsem` this is handled by having a linear "unconstrained" form of the parameter, which is subject to any covariate and random effects, *inside* a mathematical transformation that ensures the resulting "constrained" parameter is appropriate (e.g., has the correct sign or does not violate positive-definiteness of the matrix, etc.).

Although there is nothing in the formulation of our three theories that specifically focuses on individual differences, individuals' cognitive capacities, living situations, and all manner of other causes of performance differences can be stable over long timespans and will need to be accounted for in our model. This "could" be as simple as allowing for individual differences in initial performance level, though given the broad age range (developmentally speaking) in question, to us it seems more reasonable to assume that all aspects of the system may differ. One plausible possibility would be that fundamental levels of language competence, as attained in lower school grades, are critical for learning mathematics, but at higher levels, this is no longer the case—so the medium function becomes less relevant with age. In contrast, higher grades could see the emergence of the specialization phenomena, wherein time constraints and the importance of focusing on study areas become more substantial concerns.

MODEL BUILDING

If we had a complete, formalized theory of competence development, a reasonable starting point for modeling would simply be to instantiate that theory and work from there. Instead, we are faced with what may be the more typical scenario, wherein we have some vaguely specified theories and want to see to what extent each may have something to offer. Because the three theories of interest are only a small piece of the competence development puzzle, we also need many more pieces—these may be seen as auxiliary assumptions when it comes to considering the theories in question. Some auxiliary pieces are likely to be more important, in terms of explained variance in the data than those elements which are directly relevant from a theory comparison perspective. A good example of such is the general growth trend across subjects—ignoring this aspect would force all of the growth to be accounted for in the system dynamics and measurement error terms, likely leading to large temporal cross-effects between maths and language even though growth in the two may be only due to a common cause. Because of such considerations, we will walk through a progression of models, starting with basic components and building up in complexity, including parameters of more theoretical interest only in the later stages. The fact that we *begin* with a linear growth curve should not be taken to indicate that the model is "primarily" a growth curve—eventually the model will have components that function like latent change score models (Ghisletta & McArdle, 2012), common/dynamic factor models (Molenaar, 1985), and vector autoregressive models, in a hierarchical continuous-time framework with parameters moderated by covariates. For inference, we use the maximum a posteriori estimation approach of `ctsem` (Driver & Voelkle, 2021), which can also be thought of as a form of penalized likelihood. For main effect type parameters in a data set of this size, this will have very little influence, but for moderated parameters in complex models such as we will come to, it can lessen capitalizing on chance results by keeping estimates more conservative than maximum likelihood. Results from all model fits, both with priors and without, are available in the supplementary material. When time is discussed, we use the metric of years, unless explicitly specified.

Linear growth

Although linear growth without individual differences is a highly unrealistic specification for most developmental processes, it provides a basis on which to build. At this stage, none of the three theories are of specific relevance, we only include the most fundamental of auxiliary assumptions—that there might be non-zero growth over time. In this case, the process (Equation 2) can be simplified by dropping matrices of zeros and reverting to ordinary differential equation form (with *dt* on the left), leaving:

$$\frac{d\boldsymbol{\eta}(t)}{dt} = \mathbf{b}. \tag{4}$$

This tells us that the rate of change in maths and language processes $\boldsymbol{\eta}$ at time (t) is determined by the continuous intercepts **b**. In our model, there are two latent processes, language and maths, so η and b are vectors of length two. The measurement model is as shown in Equation (3) and involves five indicator variables. Two variables measure language skill and three measure mathematics. The first indicator for each process is constrained to a factor loading of 1.0 and an intercept of 0.0, to identify the model.

The point estimates of the model fit are shown in Appendix A in expanded matrix form, with the zero matrices left in place to facilitate understanding. From the estimates, we obtain the most basic sanity check of our data and model combination, as there is an upward growth trend in both mathematics and language performance over time, based on the positive values for the continuous intercept **b** vector.

Individual differences in linear growth

Of course, single values for the expected initial values and growth across individuals and age is a highly unlikely proposition. An obvious next step then is to allow individual differences in these parameters. At this point, we will still not have an instantiation of the three theories, but can start to look at the relation between maths and language development, and also changes in this relation over age.

Because we do not have sufficient data (i.e., hundreds or thousands of time points per subject) to estimate our more complex models separately for each individual, we will ignore such a possibility. Then, individual differences in model parameters can be achieved either via including additional latent variables to estimate the mean and standard deviation of the distribution of parameters (i.e., a 'random effects' model), or by (also) regressing model parameters on one or more covariates.

Linear growth—Random effects

Starting with random effects, we can allow for varying initial intercepts (in `ctsem` "T0MEANS" for "time zero latent process means") by estimating the initial covariance matrix parameters ("T0VAR" for "time zero variance/covariance matrix" in `ctsem`) for the latent processes. For random variation in any parameters other than the initial intercept, these need to be included as additional processes in the system model, which exert appropriate influence depending on which type of parameter they are (For users of `ctsem`, this occurs in the background when individually varying parameters are requested). Continuous intercepts are one of the simplest parameter types to specify as random-effects; they can be included as additional processes exerting an influence of 1.00 on the original processes of maths and language. This influence occurs via the temporal effects, or drift, a matrix that we could previously ignore as it contained only zeroes. Now, the **b** vector from Equation (2) is no longer needed—instead, the continuous intercepts are included in the extended state vector $\boldsymbol{\eta}$. Equation (5) shows the point estimates for this expanded initial latent state distribution. Here we see the pattern both the medium and thinking function theories would predict. The initial latent states for maths and language positively covary (top-left quadrant), the continuous intercepts do also (lower-right quadrant), and the initial states are *negatively* related to the continuous intercepts—performance growth is expected to be lower for those who start with higher performance (lower left/top-right quadrants). The terminology of initial states and continuous intercepts may be confusing to those accustomed to intercepts and slopes, but clarity here is important: In this case, because we have a simple linear model, the initial latent state $\boldsymbol{\eta}(t_0)$ is equivalent to the intercept term in a growth model and sets the expected value of the process for the first observation. The continuous intercept is then equivalent to the linear slope. Once more complex models are used, this one-to-one relation breaks down—for instance, in systems that fluctuate around a baseline, as in many vector autoregressive type models, the continuous intercept (in combination with temporal effects) serves to determine the baseline.

$$\underbrace{\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}(t_0)}_{\eta(t_0)} \sim \mathrm{N}\left(\underbrace{\underbrace{\begin{bmatrix} -0.31 \\ -0.52 \\ 0.14 \\ 0.1 \end{bmatrix}}}_{\text{T0MEANS}}, \underbrace{\underbrace{\begin{bmatrix} 0.5 & 0.33 & -0.07 & -0.05 \\ 0.33 & 0.27 & -0.05 & -0.04 \\ -0.07 & -0.05 & 0.02 & 0.01 \\ -0.05 & -0.04 & 0.01 & 0.01 \end{bmatrix}}_{\mathbf{Q}^*_{t0}}}_{\text{T0VAR}} \right). \tag{5}$$

With the incorporation of the continuous intercepts as latent states, the drift (temporal-effects) matrix is no longer all zero but contains some fixed values of 1.00 where the effect flows *from* the two continuous intercepts (columns 3 and 4) *to* our language and maths processes of interest (rows 1 and 2):

$$\underbrace{\begin{bmatrix} 0.62 & 0 & 0 & 0 & 0 \\ 0 & 0.62 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.49 & 0 \\ 0 & 0 & 0 & 0 & 0.49 \end{bmatrix}}_{\underbrace{\Theta}_{\text{MANIFESTVAR}}} \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}}_{\epsilon(t)} (t). \quad (6)$$

Our specified measurement model appears reasonable, with point estimates suggesting that all indicators load similarly on the latent factors (seen in the Λ matrix) and have similar measurement error standard deviations (the Θ matrix):

$$\underbrace{\begin{bmatrix} \text{READING} \\ \text{GRAMMAR} \\ \text{MATHS1} \\ \text{MATHS2} \\ \text{MATHS3} \end{bmatrix}}_{\mathbf{Y}(t)} (t) = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1.16 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1.25 & 0 & 0 \\ 0 & 1.25 & 0 & 0 \end{bmatrix}}_{\underbrace{\Lambda}_{\text{LAMBDA}}} \underbrace{\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}}_{\eta(t)} (t) + \underbrace{\begin{bmatrix} 0 \\ -0.17 \\ 0 \\ 0.24 \\ 0.36 \end{bmatrix}}_{\underbrace{\tau}_{\text{MANIFESTMEANS}}} + \quad (7)$$

$$\underbrace{d\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}(t)}_{d\eta(t)} = \left(\underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\underbrace{\mathbf{A}}_{\text{DRIFT}}} \underbrace{\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}(t)}_{\eta(t)} dt. \right) \quad (8)$$

At this point, we have the classic multivariate linear latent growth curve model with (co)varying slopes and intercepts, formulated as a state-space stochastic differential equation. With just this random-effects linear growth model, we cannot substantially distinguish between the three theories, except to say that specialization does not appear as a dominant feature—maths and language performance positively, rather than negatively, co-vary. The positive correlations in growth are consistent with both thinking and medium function.

Linear growth—Covariate moderated parameters

Although the random effects approach can account for individual differences in model parameters, and relations between such individual differences, there are often substantive and pragmatic reasons for including covariate effects to moderate parameters of the model. Substantive reasons are typically when system parameters are expected to vary along with the covariate, and one wishes to account for and understand this relation—as with our case and the specialization hypothesis, wherein changes in relations based on age are expected. Pragmatic reasons can be that individual differences are expected on more parameters, but the computational cost and difficulties are too high to include random effects everywhere. Including moderation effects can allow for at least some of the heterogeneity, and generally imposes substantially less computational burden than random effects. At present, `ctsem` allows only linear moderation effects (expandable to polynomials via the inclusion of transformed covariates), though in principle any functional form is possible. A similar approach but using step-functions (i.e., groups) can be seen in structural equation model trees (Brandmaier et al., 2013) and the proof-of-concept with `ctsem` (Brandmaier et al., 2018).

While we could restrict moderated parameters to only those parameters we included random effects for—our initial states and continuous intercepts—a key benefit of the sometimes awkward seeming approach to covariance matrices used in `ctsem` is that we can also allow for moderated standard deviations and correlations, along with any other parameter. We could of course treat all of these as random effects, but elect not to for ease of both computation and explanation. For our model development, we will include an age covariate, centered at 13 years. Not only does this allow us to account for changing relations between starting points and growth as children age, but this can also be used to detect and account for changes in measurement properties, such as factor loadings or measurement error variance. Although this is still not a direct theoretical concern of ours, we want to ensure that major changes across age are accounted for to avoid biasing the theoretical parameters of interest we will include in later steps. The one restriction we do impose on the covariate effects is that we do not allow a change in the measurement intercept with age, as this would be akin to allowing distinct growth curves for each indicator variable when we are really interested in characterizing the change in the latent process. Fitting this model results in a similar overall pattern to the unmoderated model, with all age moderation effects shown in Appendix Table B1. These moderation effects indicate that, as expected, initial language and maths performance rise with age. The growth rate for language appears largely unaffected by age, while the growth rate for maths also rises with age. Factor loadings are relatively unaffected, while measurement error standard deviation reduces with age for all indicators except German reading performance. In terms of correlations among initial states and continuous intercepts, there are two effects where the 95% confidence interval does not include zero. The correlation between initial maths performance and the rate of growth appears to *increase* with age, while the relation between mathematics growth and language growth appears to decrease. While such results are useful for inferences regarding the direction of effects, it is difficult to comprehend the magnitude of effects due to the nonlinear matrix transformations needed to convert the age-moderated "unconstrained correlation" parameters into correlations. For such purposes, plots of the expected correlations, conditional on age, are very helpful. Figure 1 shows all correlations within and between initial states and slopes. From this figure, it is clear that the general pattern is no change in the correlation between initial states, reduced correlation in slopes as age increases, and a lessening of the negative correlation between initial states and slopes with age. The lower correlation in slopes with age is what we would expect to see if specialization is occurring—children who devote energy to one domain may not be able to devote the same energy to the other, once the demands from schooling increase with age.

Including dynamics

So far, we have only developed a *static* model of change—the model predictions for an individual's state at any point in the future depend only on the system parameters for that

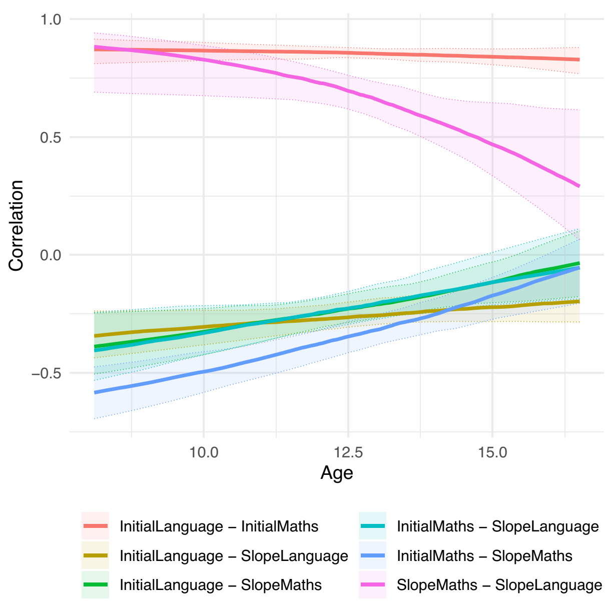

FIGURE 1 Correlations of initial states and slopes, conditional on age, for the moderated linear growth model. 95% confidence intervals are shaded.

individual, and time (Voelkle et al., 2019). If we shift to a truly *dynamic* model, in which we acknowledge that unpredictable changes occur in the constructs, the model is likely to perform (i.e., predict) better, and can be used to consider questions on the relation between these unpredictable fluctuations—including those from our three theories of competence development. To allow for such fluctuations, it is usually necessary to relax the implicit assumption that *all* information from earlier states is carried forward in time, represented by the zeroes in the temporal effects matrix diagonal. This is because if the new variation is now to be continually added, representing fluctuations, the total variance will continually increase, which is often unrealistic and undesirable. To relax this assumption, we (generally) need negative values instead of zeroes in the diagonals of the temporal effects matrix (similar to a discrete-time vector autoregressive model where autoregressions of less than 1.00 are needed). With a negative auto-effect (the diagonals of the temporal effects matrix), when a system fluctuates *above* its deterministic trend, the negative dependency on earlier states pushes the system back *down* toward the expected trend and vice versa for downward fluctuations. Positive dependencies may also be plausible over limited timespans but become "explosive" without other mitigating model elements. Allowing for temporal effects and system noise brings us back to the full Equation (2). At first, we will only free the diagonals of the temporal effects, such that any dependency between changes in the maths and language latent processes is solely due to the correlated system noise. Put differently, we assume that maths and language may have correlated changes that the deterministic trend component could not account for, but we do not allow for directionality between these changes at present. This means that we will at present *not* be able to distinguish between the common cause posed by the thinking function hypothesis and that of the directional effect from language to maths posed by the medium function hypothesis. However, to the extent that specialization is occurring, further evidence for these could appear in the form of lower correlations in system noise as age increases.

By freeing the auto-effects in the temporal dependency matrix, the deterministic trend becomes nonlinear and incorporates the idea that performance may rise more slowly, the higher performance is. This relation could already be seen when we fit the linear growth model with random effects, in that there was a negative correlation between the initial state and the growth rate—but in that case, it was only a relation *between-subjects*, and any specific subject would be (apparently wrongly) assumed to grow linearly. The updated formulation accommodates such a concept *within-subject*.

With free auto-effects and correlated system noise included, the point estimates for the system equation are shown in Equations (9) and (10). Expected trajectories for a typical student in the sample, *before* knowing any of their individual scores, are shown in Appendix Figure C1. After accounting for students' performance scores, the estimated forward predictions (i.e., conditioned on past scores) of the model are far less smooth, as Figure 2 demonstrates. In this plot, it is easy to see that for periods of time when there is more data on a student, the students' scores in this period influence predictions and can result in substantial fluctuations in the expected trajectory. This contrasts with periods of less data, where predictions are more dependent on the overall trend. In either case, the proportion of measurement error is estimated to be relatively large, meaning that predictions do not closely track each individual observation.

What can we learn from our model after including some dynamics? Two main aspects are the shape of the overall trend, which is now more flexible, and the shorter-term fluctuations. Regarding the overall trend, it seems that individual growth in performance slows as performance rises—this we can infer from the negative auto-effects or the expected trends shown in Appendix Figure C1. Regarding shorter-term fluctuations in performance, the system noise covariance in Equation (10) shows that unexpected changes in language and maths

$$d\underbrace{\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}(t)}_{\mathrm{d}\boldsymbol{\eta}(t)} = \left(\underbrace{\underbrace{\begin{bmatrix} -0.52 & 0 & 1 & 0 \\ 0 & -0.51 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\mathbf{A}}}_{\text{DRIFT}} \underbrace{\begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \end{bmatrix}}_{\boldsymbol{\eta}(t)}(t) + \underbrace{\underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{b}}}_{\text{CINT}} \right) dt + \tag{9}$$

$$\underbrace{\underbrace{\text{Cholesky}\left\{ \begin{bmatrix} 0.12 & 0.08 & 0 & 0 \\ 0.08 & 0.08 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\}}_{\mathbf{Q}}}_{\text{DIFFUSIONcov}} \underbrace{d \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}(t)}_{\mathrm{d}\mathbf{W}(t)}. \tag{10}$$

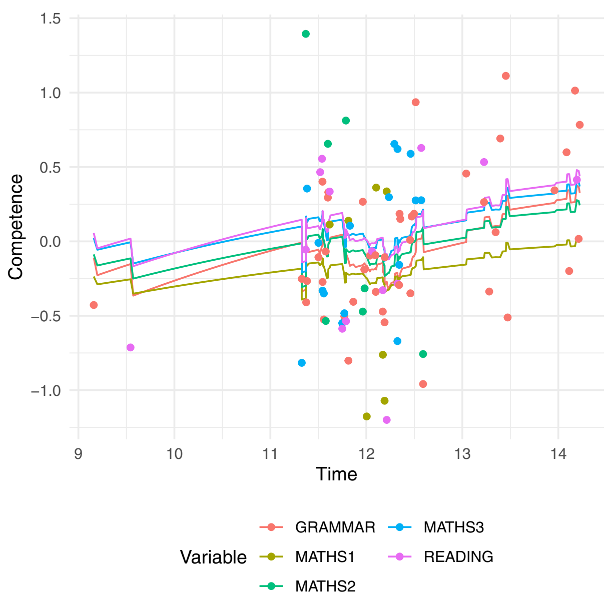

FIGURE 2 One step ahead forward predictions (shown by lines) for one student based on past data (shown by dots), generated based on the model fit with individually varying trends, auto-regression effects, and correlated system noise.

performance are highly correlated, with similar variances. Such results suggest that specialization is clearly not the primary feature driving short-term fluctuations in learning in the two domains. Had we seen negative correlations in the system noise, or at least substantially lower correlations than between the long-term trends, there would have been a good reason to consider specialization as a substantial phenomenon overall. Nevertheless, the age moderation effect on the system noise correlation *is* negative, as seen in Appendix Figure D1. This reduced correlation with age is supportive of the idea that some amount of specialization may be occurring, as this is thought to increase with age. Distinguishing between the thinking and medium function hypotheses is not yet possible, given the lack of directionality in the model at this point.

Multiple time scales and directional dynamics

So far, we have built up the basic components of the model, and see some level of support for specialization, but have not been able to distinguish between the medium and thinking function theories—to what extent does language support maths development and to what extent are they driven by common causes. To address this theoretical issue, our model will need directional dynamics incorporated, wherein changes in one process can lead to changes in the other. The results from such a directional dynamic model can contain lots of information, sometimes verging on overwhelming once individual differences in parameters are included. There is, however, a need for caution in interpretations.

One issue in the model specification, as it stands presently, is that the shape of the long-term trends is confounded with the time scale of fluctuations. That is, the auto-effects terms on the diagonal of the temporal effects matrix **A** are doing double duty, as they set the curvature of the overall deterministic trend, and also how long the short-term fluctuations persist in the system. While there, may of course, be no need to distinguish the two since the long-term trend might arise due solely to the temporal dynamics included in the model, this would be a strong assumption—we think it is more likely that there are persistent influences on maths and language development that are *not* maths and language performance themselves, such as cognitive function. Ideally, we would have well-measured data for all possible influences on the system and could simply model everything as interrelated dynamics. This is of course generally not possible, so it is important to model reasonable proxies for such influences and to remember when interpreting model parameters that they will depend on such assumptions.

This issue with the trend over time (i.e., nonstationarity) is why two-step procedures, where the data first have trends removed, are sometimes used; however, it can be more meaningful and statistically appropriate to include the trend and the dynamics in the same model, unless one component is known with fairly high certainty. To achieve this, we can expand the system in a similar way as when we included the continuous intercepts as a random effect. The full model specification is visible in Appendix E. Essentially, we create an extra process for each language and maths to contain the fluctuations or dynamics. The temporal effects and initial covariance matrices are fixed to zero in many elements such that these new dynamics processes do not interact or covary with the other processes in the system. The auto-effect of the dynamics processes is estimated, setting the speed of the fluctuations. The initial state and continuous intercept for each dynamics process is set to zero, to ensure that the processes capture only the fluctuations and not the general trends. We duplicate the first two columns of the factor loading matrix $\boldsymbol{\Lambda}$, as these additional dynamics processes will not interact with what is now the trend and continuous intercept processes but are essentially just added on top to generate the model predictions. In this way, we estimate an overall trend, and then *around* this trend we have dynamic fluctuations, with neither contaminated by parameter estimates for the other.

To address the theoretical question of common cause versus language-driven growth, we now also free the cross-effect parameters between language and maths dynamics. The system structure and fitted point estimates for the model are now seen in the temporal effects matrix of Equation (11) and the system noise matrix of Equation (12).

$$\underbrace{\underbrace{\begin{bmatrix} -0.22 & 0 & 1 & 0 & 0 & 0 \\ 0 & -0.14 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2.03 & -2.86 \\ 0 & 0 & 0 & 0 & 1.23 & -7.78 \end{bmatrix}}_{\mathbf{A}}}_{\text{DRIFT}} \begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \\ \text{dynLANG} \\ \text{dynMATHS} \end{bmatrix}. \tag{11}$$

$$\underbrace{\underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.37 & 0.32 \\ 0 & 0 & 0 & 0 & 0.32 & 0.38 \end{bmatrix}}_{\mathbf{Q}}}_{\text{DIFFUSIONcov}} \begin{bmatrix} \text{LANG} \\ \text{MATHS} \\ \text{cintLANG} \\ \text{cintMATHS} \\ \text{dynLANG} \\ \text{dynMATHS} \end{bmatrix}. \tag{12}$$

In the model with only one auto-effect for language and one for maths, the auto-effects were approximately −0.70, which implies autoregressions after 1 year of $\exp(-0.70 \times 1) = 0.50$. Now that we have allowed for different timescales by splitting trends and dynamics, auto-effects for the trend components are closer to zero, giving trends that are closer to the linearity of our earlier growth models. The timescale of the dynamics components is now faster (more negative), and there is a negative cross-effect from maths to language. These dynamics may be easier understood by referring to Figure 3. The top half of this diagram shows the typical interpretation, wherein an upward shift in maths performance tends to be followed in the coming months by a *drop* in language performance, while improvements in language do not (significantly) predict later changes in maths. The medium-function hypothesis suggests that language facilitates mathematics learning, so the most basic expectation would be that we would have seen a positive cross-effect from language to maths. Instead, the picture is more mixed, as one can still make some argument for support since improvements in language do appear to have a more positive influence going forward than improvements in maths—yet this "more positive influence" appears to be approximately zero.

The typical interpretation of the negative cross-effect from maths to language partly aligns with the specialization idea that time is scarce, and moments spent focusing on one domain inevitably mean other domains may suffer. However, such an interpretation is tricky to reconcile with the high correlation in the system noise term. Plotting the temporal dynamics *in combination with* the system noise, as suggested in Driver (2022), means the

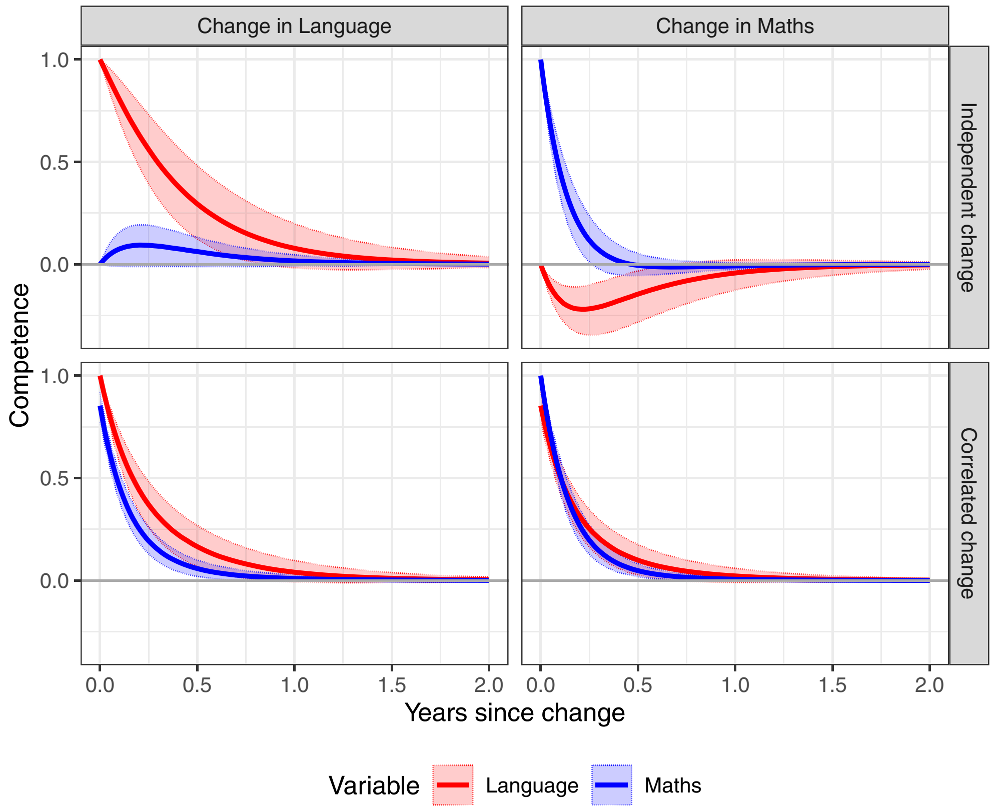

FIGURE 3 Temporal dynamics of fluctuations in maths and language performance, implied by the multiple timescale model with directed effects. 95% confidence intervals are shaded. At time zero, the process indicated in the column heading receives a perturbation of 1.00, leading to changes in both maths and language processes. The perturbation may be independent or correlated with a perturbation in the other process according to the estimated system noise correlation. Note the substantial difference in interpretation suggested by the independent compared with correlated perturbation—what looks like a drop in language due to rises in maths (top right) may just be a faster dissipation to baseline trend (bottom right).

lower half of Figure 3 tells quite a different story. In this case, we can see that whenever one process fluctuates upward, the other typically does the same but to a slightly lesser extent. Here then, the negative cross-effect from maths to language essentially means that when maths rises, so too does language, but that this perturbation in language *dissipates faster than is typical*, due to the negative cross-effect from maths. So, when language performance fluctuates *together* with maths performance, the change tends to dissipate quicker than when language performance fluctuates *alone* or in the *opposite direction* to maths performance.

Interestingly, when we look at the age moderation effect on system parameters in Appendix Table F1, both cross-effect parameters become more negative with age. While we still urge caution because of the strong correlation in system noise, these more negative effects with age *do* neatly map onto the specialization idea—there is increased competition for resources between domains as students progress in their education (Hofer, 2010; Köller et al., 2001). Moreover, as seen in Appendix Figure G1, the cross-effect of language on maths changes substantially with age, such that there is some evidence for the medium function hypothesis at younger ages when language skills are more likely to be decisive for general skill learning. In this plot, we also see that fluctuations in competence at younger ages tend to persist for longer, though our theories have nothing to say on such a matter.

DISCUSSION

To examine different theories of competence development, we have built up models of development beginning with linear growth, through to combined long-term growth and dynamic fluctuations with individual differences. This development was designed to allow the incorporation and examination of the different theories, which in themselves were not sufficiently specific to lead to an obvious single formal instantiation. As such, any interpretation of the parameters and their relation to the theories is also contingent on the many auxiliary elements included in the models. While this is in some sense less than ideal for serious consideration of theory, it is also very much the norm across fields of psychological research, where theories tend to be only vaguely specified in words. Our step-by-step build of the model hopefully helps to familiarize readers with some of the many difficulties when it comes to instantiating developmental theories as formal statistical models and drawing inferences with respect to the theories.

In no sense do we pose this work as a perfect exemplar—it is intended as an informative walk-through of an initial theory formalization process, hopefully similar enough to what other researchers encounter in their empirical works to offer useful insights, rather than some Platonic ideal. The formalizations in use here should be viewed skeptically, as initial forays to be further developed based on iterative model and theory-checking processes. That said, there *are* consistent patterns found across the range of models considered, and the relatively unique data set offers new opportunities to consider the relation between short- and long-term changes in competence.

In general, we found that initial states and long-term growth trends all correlated positively, random fluctuations in maths and language were highly correlated, and a small negative temporal relation is apparent wherein upward fluctuations (i.e., above the expected growth trend) in maths predict downward fluctuations in language. In older children, initial states were (unsurprisingly) higher, measurement error tended to be lower, contemporaneous changes became less positively related, and directed temporal effects became more negative. The pattern of age relations that applies to short-term fluctuations can also be seen in the long-term trends, as growth in maths and language becomes less correlated with age.

The thinking function hypothesis poses both language and maths as driven by the common cause of higher-level cognitive functions, and that at least some of this occurs is perhaps so obvious as to not require verification. Nevertheless, modeling can help to distinguish the relative contributions, and in this case, we see high correlations in both short- and long-term change, yet these reduce as children get older. So, there is clearly some form of developmental differentiation (Breit et al., 2021; Garrett, 1946) occurring, with abilities becoming less correlated with age. The mechanism behind the differentiation is not clear though—is the common cause of cognitive function merely becoming less relevant to one or both skills or is the differentiation driven by the increased specialization and time demands suggested by the specialization idea? To the extent that we accept cognitive function as relatively stable across weeks and months, then specialization or some other phenomena must be at play. This is because the reduced correlations in change between maths and language are seen not just for the long-term change, but also the short. With the available data, it would be quite a leap to *comprehensively* claim that the reduced correlations are due to specialization—for this, we would ideally have data on both study behavior and general cognitive function—yet the results are *consistent* with such an idea, as long-term trends, short-term fluctuations, and the forward predictions from such fluctuations all reduce or become more negatively related with age.

The medium function hypothesis places language as a facilitator of mathematics learning. This too, seems at least somewhat obvious, as without *any* language capacity a child would surely struggle in a modern classroom. Nevertheless, it is interesting to know to what extent this idea may be relevant for the regular student population. Based on the directed temporal effects in our final model, there may be limited support for this idea, from two directions. First, while the cross-effect from maths to language is negative, from language to maths, it is positive but nonsignificant. This is qualified, however, by a substantial shift in this cross-effect across age, with a positive effect of language on maths at younger ages shifting to a negative effect at older ages, as seen in Appendix Figure G1. Considering these results together does provide some support for the idea of a medium function to language *if* one also accepts that this medium function is likely to be most relevant at lower levels of language ability, which generally occur at younger ages. Although plausible to us, we could find no such theorizing in the literature. While further work should verify such findings, this provides an example of how preexisting theory may be updated and constrained by computational modeling.

Looking at the models we have developed, the "theoretical" portion of the models seems relatively small in comparison with all the auxiliary elements. We do not think this is unusual, as there are many complexities related to both development and measurement that, while critical for modeling, may be less interesting or relevant when formulating initial vague theories. Perhaps counter-intuitively, relatively *more* auxiliary assumptions are likely necessary when the number of variables included is relatively low, or at least, when only some aspects of the core system are observed. In such cases, the model needs to account for structural forces from unobserved variables (such as cognition or brain development in this case) in hopes of being able to tease apart relations between the variables of interest, without (too much) confounding due to unaccounted for additional variables. In our case we accounted for persistent influences on growth by including a state-dependent trend structure, then considered how the shorter-term temporal dynamics we included lined up with the various theories. By including a second form of long-term development structure—that of age moderated parameters for each individual—we were also able to address how the relevance of the different theories may change throughout development. While we believe and have argued for some justification to the model building process, it is as noted still an initial formalization, and any elements of the models are open to critique, both from theoretical as well as data-driven perspectives. Do our measurements sufficiently reflect competence in maths and language, across the age range we have assessed? Is it reasonable to think that fluctuations in observed performance at least partly reflect fluctuations in actual competence?

Is a linear age effect sufficient to capture major patterns of developmental change in the parameters of interest? These are but a few examples of the sorts of questions researchers engaged in similar modeling exercises will have to consider.

Limitations

While the Mindsteps online-learning data are wonderfully rich in some dimensions, the unsurprisingly large amount of measurement uncertainty, as well as the relatively short observational period for most students (median observed time range is 1.73 years), does somewhat hinder the endeavor to distinguish long-term trends from genuine short-term variation. More years of the Mindsteps system running should offer further opportunities.

The lack of covariates such as working memory also means that while we could address some questions regarding relations between domains, questions as to the cause of commonalities, or change in such commonalities, were unable to be addressed. This means that there are substantial influences on the system for which we do not have data, making interpretations of system parameters very dependent on how we have modeled these unmeasured influences, via the trend and system noise components. Ideally, alternative plausible specifications would also be considered. A core example here is the modeling of the trend and dynamics as separated, which may be more conservative than necessary—a higher-order model combining both slower and faster dynamics would be an interesting contrast (Boker et al., 2016).

The two-step modeling approach, in which ability estimates for each occasion were first generated from binary test data, is inevitably not ideal, but necessary for computational reasons. Although including the age-moderated measurement model should have minimized the cost of such a two-step approach, we are investigating possibilities for combining both steps.

We did not discuss model fit throughout this paper, though note that as models increased in complexity, out-of-sample likelihoods (based on 10-fold cross-validation) continued to rise, suggesting improved model performance. Nevertheless, a more comprehensive model validation approach would be important to increase confidence in inferences. This might include comparing various quantities of interest between simulations from the fitted model and the raw data (Gelman et al., 1996), assessment of residuals (Lin et al., 2002), or comparisons to alternative more flexible models (Chow, 2019). A small example of residual checks is shown in Appendix Figures H1 and H2. A more comprehensive approach should also consider if inferences depend on data-driven modeling choices—leaving a portion of data untouched except for a verification model fit at the end can be helpful to break any such dependency.

CONCLUSION

Probably, the most striking result in this work is the high correlation between short-term fluctuations in performance of maths and language. Unfortunately, there are a variety of possible reasons for this high correlation that need further research to disentangle—is it an artifact of our low-stakes measurement procedure or measurement model, or genuine fluctuations in a common cause such as motivation? With respect to the three theories relating maths and language considered, the directional results we have provided at best very tenuous evidence for the medium function hypothesis (Bruner, 1966; Cohen & Dahaene, 1995; Fetzer & Tiedemann, 2018) in which language facilitates mathematics performance. The high correlation in system noise and correlated trends between maths and language would seem to support the thinking-function hypothesis (Daneman & Merikle, 1996; Lombrozo, 2006; Peng et al., 2018). Although this high correlation in system noise would seem to speak *against* the idea that learning mathematics or language draws potential resources away from learning the other, the fact that cross-effects in both directions became more negative with age, when students may struggle more with time pressure and demands, does fit with the ideas of specialization discussed by Hofer (2010) and Köller et al. (2001).

In terms of the modeling framework, while it is certainly not the only approach to instantiating theories and fitting them to data, the flexibility of combining stochastic differential equations with measurement models and individual differences does offer opportunities. The models discussed in this paper can easily be extended using ctsem or other software to include additional processes (cognitive performance such as working memory would be invaluable) as well as more sophisticated dynamics and measurement as needed. Some downsides of this framework are computational cost, the unfamiliar mathematical language (for social scientists), and sometimes unexpected dependencies in parameterization—some of which we have discussed in hopes of mitigating problems. However, given that the continuous-time parameterization employed here is, arguably, closer to the way we think about developmental processes than common modeling frameworks, these costs are likely worth bearing in at least some cases. By taking the step to explicitly formalize developmental theories in such a way, we take steps to interrogate and expand our understanding of the theories, hopefully leading to more thoughtful theory development and model instantiation.

FUNDING INFORMATION

This work was funded by the Swiss National Science Foundation.

ACKNOWLEDGMENTS

Open access funding provided by Universitat Zurich.

DATA AVAILABILITY STATEMENT

Analyses were not preregistered, scripts to reproduce are available in the online supplementary material. Mindsteps data are from private industry and cannot be made public.

ORCID

Charles C. Driver https://orcid.org/0000-0002-4174-2970
Martin J. Tomasik https://orcid.org/0000-0001-5235-3551

REFERENCES

Aalen, O. O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, *1987*(3–4), 177–190. https://doi.org/10.1080/03461238.1987.10413826

Aalen, O. O., Røysland, K., Gran, J. M., Kouyos, R., & Lange, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical Methods in Medical Research*, *25*(5), 2294–2314. https://doi.org/10.1177/0962280213520436

Bailey, D. H., Oh, Y., Farkas, G., Morgan, P., & Hillemeier, M. (2020). Reciprocal effects of reading and mathematics? Beyond the cross-lagged panel model. *Developmental Psychology*, *56*(5), 912–921. https://doi.org/10.1037/dev0000902

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*(5), 611–626.

Berger, S., Verschoor, A. J., Eggen, T. J. H. M., & Moser, U. (2019). Development and validation of a vertical scale for formative assessment in mathematics. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00103

Boker, S. M. (2012). Dynamical systems and differential equation models of change. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, vol 3: Data analysis and research publication*. American Psychological Association. https://doi.org/10.1037/13621-000

Boker, S. M., Staples, A. D., & Hu, Y. (2016). Dynamics of change and change in dynamics. *Journal for Person-Oriented Research*, *2*(1–2), 34–55. https://doi.org/10.17505/jpor.2016.05

Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. https://doi.org/10.1177/1745691620969647

Brandmaier, A. M., Driver, C. C., & Voelkle, M. C. (2018). Recursive partitioning in continuous time analysis. In K. Van Montfort, J. H. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences* (pp. 259–282). Springer.

Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods*, *18*(1), 71–86. https://doi.org/10.1037/a0030001

Breit, M., Brunner, M., & Preckel, F. (2021). Age and ability differentiation in children: A review and empirical investigation. *Developmental Psychology*, *57*(3), 325–346. https://doi.org/10.1037/dev0001147

Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.

Chow, J. C., & Jacobs, M. (2016). The role of language in fraction performance: A synthesis of literature. *Learning and Individual Differences*, *47*, 252–257. https://doi.org/10.1016/j.lindif.2015.12.017

Chow, S.-M. (2019). Practical tools and guidelines for exploring and fitting linear and nonlinear dynamical systems models. *Multivariate Behavioral Research*, *54*(5), 690–718. https://doi.org/10.1080/00273171.2019.1566050

Chow, S.-M., Ho, M.-H. R., Hamaker, E. L., & Dolan, C. V. (2010). Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 303–332. https://doi.org/10.1080/10705511003661553

Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002

Cohen, L., & Dahaene, S. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Coleman, J. S. (1964). *Introduction to mathematical sociology*. Free Press of Glencoe.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

de Araujo, Z., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, *88*(6), 879–919. https://doi.org/10.3102/0034654318798093

Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 61–75. https://doi.org/10.1080/10705511.2014.973960

Driver, C. C. (2022). *Inference with cross-lagged effects—Problems in time and new interpretations*. https://doi.org/10.31219/osf.io/xdf72

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with r package ctsem. *Journal of Statistical Software*, *77*(5). https://doi.org/10.18637/jss.v077.i05

Driver, C. C., & Voelkle, M. C. (2018). Hierarchical Bayesian continuous time dynamic modeling. *Psychological Methods*, *23*(4), 774–799. https://doi.org/10.1037/met0000168

Driver, C. C., & Voelkle, M. C. (2021). Hierarchical continuous time modeling. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp. 887–908). Academic Press. https://www.sciencedirect.com/science/article/pii/B9780128139950000340

Fetzer, M., & Tiedemann, K. (2018). The interplay of language and objects in the process of abstracting. In J. N. Moschkovich, D. Wagner, A. Bose, J. Rodrigues Mendes, & M. Schütte (Eds.), *Language and communication in mathematics education: International perspectives* (pp. 91–103). Springer International Publishing. https://doi.org/10.1007/978-3-319-75055-2_8

Freund, A. M. (2008). Successful aging as management of resources: The role of selection, optimization, and compensation. *Research in Human Development*, *5*(2), 94–106. https://doi.org/10.1080/15427600802034827

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, *1*, 372–378. https://doi.org/10.1037/h0056380

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760. https://www.jstor.org/stable/24306036

Ghisletta, P., & McArdle, J. J. (2012). Latent curve models and latent change score models estimated in r. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(4), 651–682. https://doi.org/10.1080/10705511.2012.713275

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., & Borsboom, D. (2022). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, *27*(6), 930–957. https://doi.org/10.1037/met0000303

Hofer, M. (2010). Adolescents' development of individual interests: A product of multiple goal regulation? *Educational Psychologist*, *45*(3), 149–166. https://doi.org/10.1080/00461520.2010.493469

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*(5), 448–470. https://doi.org/10.2307/749801

Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology*, *109*, 977–992. https://doi.org/10.1037/edu0000182

Korpipää, H., Koponen, T., Aro, M., Tolvanen, A., Aunola, K., Poikkeus, A.-M., Lerkkanen, M.-K., & Nurmi, J.-E. (2017). Covariation between reading and arithmetic skills from Grade 1 to Grade 7. *Contemporary Educational Psychology*, *51*, 131–140. https://doi.org/10.1016/j.cedpsych.2017.06.005

Kuiper, R. M., & Ryan, O. (2018). Drawing conclusions from cross-lagged relationships: Re-considering the role of the time-interval. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 809–823.

Lin, D. Y., Wei, L. J., & Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics*, *58*(1), 1–12. https://doi.org/10.1111/j.0006-341X.2002.00001.x

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*(2), 181–202. https://doi.org/10.1007/BF02294246

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2016). Bayesian data analysis with the bivariate hierarchical Ornstein-Uhlenbeck process model. *Multivariate Behavioral Research*, *51*(1), 106–119. https://doi.org/10.1080/00273171.2015.1110512

Ou, L., Hunter, M. D., & Chow, S.-M. (2019). What's for dynr: A package for linear and nonlinear dynamic modeling in R. *The R Journal*, *11*(1), 91–111.

Oud, J. H. L. (2002). Continuous time modeling of the cross-lagged panel design. *Kwantitatieve Methoden*, *69*(1), 1–26. http://members.chello.nl/j.oud7/Oud2002.pdf

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, *65*(2), 199–215. https://doi.org/10.1007/BF02294374

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, *146*(7), 595–634. https://doi.org/10.1037/bul0000231

Peng, P., Wang, C., & Namkung, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research*, *88*(3), 434–476. https://doi.org/10.3102/0034654317753350

Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743. https://doi.org/10.1177/1745691620974697

Ruissen, G. R., Zumbo, B. D., Rhodes, R. E., Puterman, E., & Beauchamp, M. R. (2022). Analysis of dynamic psychological processes to understand and promote physical activity behaviour using intensive longitudinal methods: A primer. *Health Psychology Review*, *16*(4), 492–525. https://doi.org/10.1080/17437199.2021.1987953

Ryan, O., & Hamaker, E. L. (2021). Time to intervene: A continuous-time approach to network analysis and centrality. *Psychometrika*, *87*, 214–252. https://doi.org/10.1007/s11336-021-09767-0

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods*, *24*, 70–91. https://doi.org/10.1037/met0000188

Singer, H. (1993). Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components. *Journal of Time Series Analysis*, *14*(5), 527–545. https://doi.org/10.1111/j.1467-9892.1993.tb00162.x

Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, A. Nowak, & S. J. Read (Eds.), *Computational social psychology*. Routledge. https://doi.org/10.4324/9781315173726

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*(3), 449–508. https://doi.org/10.1007/BF00499820

Tomasik, M. J., Berger, S., & Moser, U. (2018). On the development of a computer-based tool for formative student assessment: Epistemological, methodological, and practical issues. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.02245

Vallacher, R. R., & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry*, *8*(2), 73–99. https://doi.org/10.1207/s15327965pli0802_1

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review*, *98*(1), 3–53.

van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos*. Harvester Wheatsheaf.

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2019). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, *53*, 782–805. https://doi.org/10.1080/00273171.2018.1496813

Voelkle, M. C., & Oud, J. H. (2015). Relating latent change score and continuous time models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 366–381.

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192. https://doi.org/10.1037/a0027543

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Driver, C. C., & Tomasik, M. J. (2023). Formalizing developmental phenomena as continuous-time systems: Relations between mathematics and language development. *Child Development*, *94*, 1454–1471. https://doi.org/10.1111/cdev.13990