Formalizing Complex Developmental Phenomena as Continuous-Time Systems: Learning Gains in Multiple Domains

Charles C. Driver and Martin J. Tomasik
Institute of Education, University of Zurich, Switzerland
Institute for Educational Evaluation, Associated Institute of the University of Zurich, Switzerland

This manuscript is not yet (September 25, 2022) peer reviewed. The development of academic competencies in mathematics and language appear strongly intertwined, but questions remain as to how and why. In this work we use data from the large-scale online learning system `Mindsteps`, to demonstrate how hierarchical continuous time dynamic systems approaches may be used to build up and formalize theories of development – specifically addressing mathematics and language competencies, and how such relations may change as students age. We use ability estimates from 95,649 observation occasions, across N = 2,786 3rd to 9th grade students and five ability domains. Substantial correlations between short-term changes in domains are found, making otherwise intriguing directional results somewhat difficult to interpret.

Keywords: Continuous time, dynamic model, cross-lagged panel, competence development, stochastic differential equation

Statistical modeling is a tricky business! When the goal is simply to generate predictions, some models may perform better and some worse, but at least there are reasonable feedback mechanisms available to the modeler, in that performance can always be checked against data that were not used for model development. When the goal is instead scientific inference, to learn something more fundamental about the phenomena studied, it can in some cases take many years or decades for further studies to debunk or confirm inferences made on the basis of models. This is easy to understand in the context we will address here: Mathematics and language performance of school students typically grows in a correlated manner, so it is 'relatively' straightforward to take some past performance scores and generate some reasonable predictions further ahead in time. If a researcher instead wants to use students scores to learn *why* domains like maths and language are correlated, and potentially to develop targeted interventions leveraging this understanding, there are many possible stories, such as genetics, parenting styles, overlap of content domains, etc. This makes it easy to weave together a plausible sounding story with a few fancy looking equations that is, nevertheless, completely wrong. Maybe a confounding variable was ignored, error prone measurements were not accounted for, or the relevance of parameter estimates for an inference was inflated – even the best intentioned make such mistakes, and for the worst there are many avenues to accrue interesting looking results. There are many paths to improved inferences, including aspects like linking to past findings, avoidance of treating exploratory work as confirmatory (maybe via pre-registration), and careful model fit checking, but there is unfortunately no simple recipe to ensure valid inferences. This work aims to contribute some small part by familiarizing developmental researchers with the hierarchical continuous-time dynamic systems modeling approach, which is in some cases more suited to specifying and testing complex theories of development than the standards of mixed-effects regression or structural equation modeling. We hope that by slowly building up in complexity, some of the possible confusion posed by parameter dependencies in multivariate developmental models becomes clear, helping readers become more critical users (and reviewers!) of such approaches. As such, although this paper addresses an empirical domain, it*is* relatively tutorial oriented, and in some cases expands in directions not explicitly relevant to the empirical enquiry put forward, and cuts short on some theoretical angles.

We start this paper by setting the stage with a rather high-level description of how human development can be conceptualized. Reverting to theoretical propositions of life-span developmental psychology (e.g., Baltes, 1987), we will argue that development under the constraints of limited resources (e.g., time, energy) comprises both gains and losses. This is because different domains, goals, actions etc. are not independent from each other but rather have supportive or competitive relations with each other (). We will then introduce selective optimization with compensation (Baltes et al., 1998) as universal elements of how successful development (defined as the maximization of gains and the minimization of losses) can be achieved given the fundamental challenge of limited resources.

In a subsequent step, we translate this high level concept into one specific area of functioning, namely academic competence development in multiple domains, focusing on the domains of language and mathematics and trying to answer how competence develops *within and between* these domains. We will approach this translation from two complementary directions. First, we will review notions and models conceptualizing the interrelations between different competence domains, argue that these interrelations, too, can be supportive or competitive, and introduce some formalizations that capture these interrelations. Second, we will review empirical literature that has investigated how competencies in the domains of language and mathematics develop over time. In the main part of the paper we look at what continuous-time dynamic models are and how they can help to formalize such investigations, then slowly build up a model from a very simple growth curve to a more elaborate form with distinct time scales for trends and dynamics, interacting dynamics, and age moderated parameters. While we are building the model up in continuous-time (i.e., differential equation) form, much of the procedure and commentary also applies to similar models formulated using structural equation modeling software or the like. We finish with the usual list of caveats and broad implications.

General Framework of Developmental Processes

It is beyond the scope of this paper to review all the propositions put forward by life-span developmental psychology in general (for overviews, see) or specifically with respect to cognitive development (for overviews, see). We want to focus on aspects that are most relevant in the context of the present paper, namely the notion of multidimensionality and multidirectionality and the notion of development as a dynamic between gains and losses.

First of all, human development is not a uniform process but needs to be conceptualized as comprising multiple dimensions developing in multiple directions. This is almost a truism when considering distinct areas of functioning such as physical, cognitive, or socio-emotional development. However, even within one area of functioning, multidimensionality and multidirectionality exists. A textbook example is cognitive development in adulthood in which different trajectories of crystallized and fluid intelligence are observed throughout the life span (e.g., Hartshorne and Germine, 2015). Not taking into account multiple areas or domains of functioning will necessarily result in an incomplete picture of human development.

Closely related to the notion of multidimensionality and, in particular, multidirectionality is the theme of gain-loss relations in development. Because humans are not equipped with unlimited resources, development always comprises both positive gains and negative losses. These dynamics can be observed throughout the entire life-span, including childhood (Uttal and Perlmutter, 1989). For instance, the transition to schooling seems to be causally associated with the increase of self-control capacities (Burrage et al., 2008) but at the same time a decline in spatial skills. We will elaborate on the causal mechanisms possibly underlying the gain-loss dynamics below. For the time being, we would like to emphasize that at least some aspects of these dynamics might be explained by inherently supportive or competitive relations between areas or domains of functioning. Given the challenge of limited resources, heavily investing in one area (e.g., academic competence) will necessarily come at the expense of expertise in another one (e.g., craftsmanship). The notion of gains and losses on multiple dimensions is quite a fundamental one and can also be found in biological (e.g., Waddington, 1957) and sociological (e.g., Featherman & Sorensen, 1983) conceptions of differentiation and specialization/canalization. There, too, is the idea that commitment to one developmental track and consequent gains therein implies the loss of alternative options, be it in areas such diverse as cell differentiation (Ho et al., 2011) or career progression (Heckhausen and Buchmann, 2019).

Individuals as active agents of their own development (Heckhausen and Buchmann, 2019) orchestrate processes of life management in order to negotiate the aforementioned dynamics of gains and losses. Generally, these processes comprise selection, optimization and compensation (SOC; for details, see Freund, Alexandra M., 2008). *Selection* (of developmental pathways, contexts, goals, action alternatives etc.) aims at reducing the number of potential alternatives and hence effectively investing resources in those that are viable. *Optimization* aims at improving the efficiency of resource investments within the selected alternatives and *compensation* becomes relevant when individuals reach the limits of their capacities. Selection, optimization, and compensation are very universal processes that can be meaningfully applied at different levels of analysis (e.g., cell, individual, society) and in different domains of functioning (e.g., cognition, emotion, motivation). The application of these processes is responsible for an age-related increase in specialization of motivational and cognitive resources and competencies (Baltes, 1987). Selection, optimization and compensation also become relevant when it comes to negotiating the supportive or competitive relations between areas or domains of functioning (Riediger and Freund, 2004). Conflicting goals due to limited resources, for instance, can be resolved by prioritizing them (Freund and Tomasik, 2021) to arrive at an optimal goal structure (Tomasik et al., 2017). This will become relevant when we will discuss competence development in multiple domains.

Competence Development as a Case in Point

Interestingly, the notion of fundamentally limited resources and the assumption of supportive and competitive relations between domains are also the starting point of some influential theories of general cognitive (e.g., van der Maas et al., 2006) and academic competence development (e.g., van Geert, 1991). In the following section, we will attempt to translate the high-level propositions of lifespan developmental psychology to cognitive development in the academic context and then summarize some of the recent reviews and meta-analyses that have investigated competence development in multiple domains.

The starting point of our elaborations is the observation of a *positive manifold* in both general cognitive abilities (e.g.,) and academic achievement (e.g.,). Scores in a battery of cognitive tests typically correlate to a substantial degree, as do scores on achievement tests in different subject domains. This very robust finding has been replicated in countless studies conducted in very different contexts. One popular explanation of the positive manifold is the existence of a higher order general factor *g* that accounts for the common variance between the single measures of intelligence (e.g., Carroll, 1993) or academic achievement (e.g. Roth et al., 2015). Although there is some controversy around the psychological meaning of *g* (e.g., Sternberg and Grigorenko, 2002) as well as methodological questions regarding the establishment of it by cross-sectional data (Borsboom et al., 2004), its psychometric interpretation as a summary measure of the positive manifold is straightforward (e.g., Bartholomew, 2004).

There are, however, at least two alternative explanations for the positive manifold in general cognitive abilities and academic achievement. The first explanation attributes at least part of the common variance between the single skills or abilities to (correlated) measurement error (e.g., Eysenck, 1987). Because it is virtually impossible to obtain purely independent measures of single skills or abilities, some correlation is inevitable resulting in the impression of a common factor. Most mathematical tasks, for instance, comprise some verbal instruction and hence require verbal skills to process them (e.g., Ajello et al., 2018). The second explanation is a developmental one and hence compatible with the propositions of life-span developmental psychology introduced above. It assumes (primarily supportive) mutual relations between single skills or abilities that over time result in parallel growth (van der Maas et al., 2006). Using cross-sectional correlative data and employing factor analytical methods it then *appears* as if there was a higher-order factor that would explain the positive manifold – which in actuality could be a result of mutually reinforcing processes over (longer periods of) time, possibly without *any* underlying common cause.

The "mutualism hypothesis" builds on a special case of a dynamic system model in which predominantly supportive relations exist. In SOC thinking this suggests that little selection is necessary, optimization works efficiently, and compensation probably is not necessary as resources are not exhausted by conflicting relations of growth in different ability domains. One might think of different, maybe high-demand situations where this is not the case and the dynamic system is also characterized by competitive relations, for instance due to resource conflicts. There then, selection pressure and lacking capacities for compensation could translate into growth in one skill or ability at the expense of another one. Also, one needs to consider that the characteristics of the system also depend on the temporal resolution at which it is observed. Zooming into a system down to moment-to-moment relations, for instance, could be associated with stronger competitive relations, simply because at some point we can do only one thing at a time and resource conflicts set in.

There is some evidence that such a dynamic systems perspective on skills and abilities could provide important insights into both the structure and the development of cognitive and academic competencies. With regard to general intelligence, van der Maas and colleagues (2006) have demonstrated that the positive manifold could perfectly be explained by the reciprocal reinforcement between single cognitive abilities over time. A general factor *g* no longer played a role and other phenomena known from research on intelligence such as differentiation effects or the observed increase in heritability of intelligence with age could be perfectly explained with a dynamic systems model.

A more sophisticated approach including more parameters in the dynamic systems model was introduced by van Geert (1991) to explain cognitive and language growth. Although building on a completely different theoretical tradition than life-span developmental psychology and using quite different terms, metaphors and examples, it shares some of its assumptions (such as the notion of limited resources) and comprises similar characteristics (such as the notion supportive and competitive relations between processes). Cognitive and language growth is modeled as an exponential process that asymptotically approaches some capacity limit and is influenced by self-reinforcing loops as well as growth in other domains in terms of supportive or competitive relations. Supportive relations exist when growth in one process (e.g., vocabulary) accelerates growth in another one (e.g., grammar) whereas competitive relations indicate growth in one process at the expense of another one. The latter occurs when processes compete for limited spatiotemporal, informational, energetic, and material resources. It is beyond the scope of this paper to elaborate the different variants of these models much more extensively (for details, however, see van Geert). Suffice to say that this dynamic systems approach allows modeling and predicting different learning phenomena such as vocabulary growth over time or the development of learning strategies. It does so without having to revert to a general factor and actually no such assumption is needed.

Mutual Relations Between Mathematics and Language

We have devoted some space to describe a general framework of developmental processes and introduced some modeling approaches because both are relevant for approaching the substantive question of how mathematics and language competencies develop over time. In the current literature, two competing hypotheses are currently discussed (see Bailey et al., 2020) and we want to introduce a potential third one before reviewing the empirical evidence for all of them. The *medium function hypothesis* () posits that language is a tool for communicating mathematical concepts with others and for building and retrieving mathematical knowledge from long-term memory. This would imply a causal relation from language performance to mathematics performance. According to this hypothesis, students who have a high level of language skills not only will profit more from instruction in general and hence grasp mathematical concepts more easily. They will also be better able to retrieve their mathematical knowledge from memory and to communicate these more competently. In modeling terms, the medium function hypothesis would be supported by finding that the cross-domain supportive effects from language to mathematics are positive and significantly stronger than those from mathematics to language.

Opposed to this notion of skill transfer is the *thinking function hypothesis* () that posits higher-level cognitive functions underlying both language and mathematics performance. The basic idea behind this hypothesis is the same as that behind a *g*-factor in research on intelligence. Indeed, research shows a strong overlap between cognitive processes underlying both (). Most prominent examples comprise nonverbal reasoning, working memory, processing speed, and even vocabulary. Furthermore, neuroimaging studies suggest a structural overlap of areas active when solving mathematical and reading tasks (). A common cause would also seem likely, if one observed that mathematics and language competencies develop in parallel over a longer period of time.

Not really reflected in the child development empirical literature but a somewhat more prominent theme in research on adolescence and young adulthood is the notion that competence development in language and mathematics could at some point develop a competing relation (). This competing relation may in part be explained by the sometimes high demands that are put on students in secondary and tertiary education, so that competence development in different domains has to compete for resources such as time and effort. This possibility has been taken into account by van Geert (1991) at least conceptually. Another explanation could be forced academic choices that adolescents and young adults (but not children) have to make (e.g., when enrolling into a language-oriented or a science-oriented school track). In any way, the result of a competitive relation between language and mathematics would be a specialization in one subject domain at the expense of the other.

To our knowledge, there are four recent reviews or meta-analyses that have investigated the mutual relations between competence development in language and mathematics (). Chow and Jacobs (2016) argue that the nature of the contribution of language to mathematics development has not been well understood, hence assuming a unidirectional relation between the two domains. The authors then focus on one specific subdomain in mathematics competence, namely fraction knowledge and understanding, and synthesize the relevant literature examining the role of language on fraction outcomes in school-age children. Three observational studies and one intervention study have been included in their analysis, all of them suggesting that language significantly contributes to fraction performance. No further details are given concerning the methodological approach to modeling the interrelation between language and mathematics. De Araujo and colleagues (2018) provide a synthesis of 75 qualitative studies with a strong focus on multilinguistic educational contexts and foreign language students. These authors, too, are implicitly assuming language as a prerequisite for mathematical competence development and based on their review suggest that, for instance, it might be easier for foreign language students to engage in mathematics learning in their first (or heritage) language rather than in their second (or school) language. Because the studies reviewed are exclusively qualitative in their methodology used, the interrelation between language and mathematics has not been explicitly modelled. Koponen and colleagues (2017) focused on investigating the association between rapid automatized naming (RAN) as a component of phonological processing and mathematics performance. As a result, RAN was significantly correlated with performance in mathematics (r = .37), in particular when calculation tasks were involved. Almost half of the 33 studies reviewed were longitudinal but the authors did not further investigate the direction of effects and were conceptually assuming RAN as a predictor of mathematics performance without even taking into account other causal directionality. Finally, there is the meta-analysis by Peng and colleagues (2020) that is remarkable for two reasons. First, with 344 studies and 393 independent samples comprising more than 360,000 participants, it was much more extensive than the other three reviews. Second, the authors made use of the available longitudinal studies to shed light on the direction of effects between language and mathematics performance. Two different conceptual models outlined above (i.e., medium function hypothesis vs. thinking function hypothesis) served as guidelines for the research questions that they addressed. For testing the thinking function hypothesis, they collected evidence as to whether working memory capacity and/or general intelligence could explain the relations between language and mathematics. In addition to this, the authors also investigated age (among other variables of less relevance here) as a moderator in in this relation. The authors found an overall relation between language and mathematics of r = .42 that was comparable in magnitude with the meta-analysis by Koponen and colleagues (2017). Working memory alone accounted for 8 to 16 percent and general intelligence alone accounted for 21 to 23 percent of the variance in the relation between language and mathematics. Working memory and intelligence together accounted for 41 to 54 percent of the common variance. This could be seen as some evidence for the thinking function hypothesis. Concerning the longitudinal relations in the studies sampled, there was an average path from language to mathematics (partialing out initial mathematics performance) of r = .20 whereas the reversed path from mathematics to language (partialing out initial language performance) was r = .22 and hence comparable in the order of magnitude. Age, in general, mostly did not moderate these relations, especially after controlling for general intelligence. Taken together, the authors found that language and mathematics performance share a substantial amount of variance with working memory and general cognitive abilities, which supports the thinking function hypothesis. Little evidence is found for a directed effect from language performance to mathematics performance as the medium function hypothesis would suggest. Rather, reciprocally supportive relations between language and mathematics seems to prevail. This somehow qualifies the results of the former three reviews and meta-analyses that all explicitly or implicitly assumed a stronger unidirectional relation.

Challenges in Bridging Theory and Empirical Research

Considering all the studies that have investigated the topic, it becomes obvious that the vast majority of them are not really able to answer the question of the mutual relations between competence development in language and mathematics. Especially when it comes to deciding between different hypothesized models (such as medium function hypothesis vs. thinking function hypothesis), many studies lack an appropriate study design, appropriate data, or an appropriate statistical method. It is hence not surprising that in most of the published studies hypotheses the relations between the two domains are not explicitly formalized.

This 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, and generate over-confidence in either uncertain or wholly wrong inferences. However, the flip-side of this coin is that without explicit formal specification, language is slippery and can be interpreted in many ways, ensuring that certain theories may sound rather fancy, garner apparent support from a range of studies, yet offer little to no predictive value for new circumstances.

A variety of recent works have argued for the usefulness of formal / computational models in theory development (), and certainly, the process of examining verbal / linguistic theories and explicitly specifying what the words mean in mathematical terms can be highly productive, both in terms of highlighting gaps and vagueness in the theory, and for determining how the theory may be compared to empirical data. Yet, as Fried (2020) points out, statistical modeling approaches in common use often do not allow for a direct connection to theories. So, not only is there a gap to be bridged between verbal theory and formal theory, but there is also a gap between formal theory and statistical model. This latter gap often receives even less consideration than the former, and in fact it often appears that statistical modeling approaches in common use tend to guide and limit the variety of formal theories considered. The following section draws some material from Driver (2022), and describes one such gap between typical longitudinal statistical models, and the underlying theory.

Misspecified Temporal Structure

Many commonly used approaches to multivariate developmental models are some form of 'discrete-time' model, which implicitly assumes processes evolve in discrete jumps in time. In reality researchers usually tend to think of the underlying processes as continually in flux. While the importance of the distinction when it comes to modeling has been discussed, (), this is regularly ignored and discrete-time cross-regressive models are still the dominant paradigm in such settings.

Driver (2022) shows that when a cross-lagged model is used to represent a hypothesis where it is believed that a) the processes actually evolve in discrete jumps, and b) the measurement occasions u actually capture *each and every one of* these discrete jumps, then there is no problem. To consider such a system, we could imagine 3 people exchanging information by post, and the post is delivered daily. Person x writes to person y, and person y in turn passes this information on to person z. If we know that person x has received new information, we can expect y to become aware of this the following day, while it will take an additional day before z knows it. In this case, we can precisely represent the causal pathways using a temporal regression matrix (or path diagram) that represents change over a time interval of one day or one postal delivery, as per Figure 1. y is entirely dependent on x for information, and z is dependent on y.

Using such a representation, the model will perform optimally in terms of both prediction and inference – we can use some knowledge of what information each individual has on a particular day to predict future days, and we can use the structure to understand that if for instance the link between y and z is removed, z will not receive any information from x. If we instead shift to an observation time interval of two days, our representation no longer matches the data gener-

$$\begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

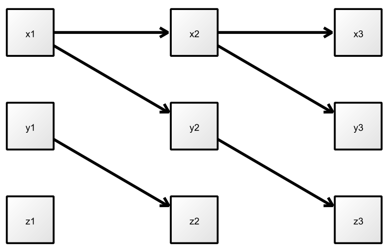

Figure 1

Daily packets of information flow between persons x, y, and z, over 3 days.

ating process, or causality, inherent to the system. For an interval of two days, the temporal regression matrix would look as follows, and the path diagram as per Figure 2.

$$\begin{array}{c} \\ x \\ y \\ z \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

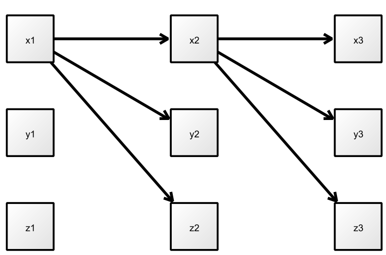

Figure 2

Packets of information flow between persons x, y, and z, across 6 days, when the time interval is two days.

In this case, while we have accurately represented how information flows through the system at time intervals of two days, we have *not* represented the actual causality in the system! One might infer from this representation that person y is unnecessary for information flow between x and z, but that is incorrect.

This discrepancy between the generating model, and the information flow over an arbitrary length of time, captures a problem inherent to many longitudinal modeling endeavors, and is why discrete-time cross regression parameters often should not be used as the basis of scientific inference. *Even when the true data generating function is a discrete time function*, whenever more than one step of change has occurred between observations the inferences become problematic. Since most psychological processes tend to be thought of as continuously existing and interacting with each other, rather than interacting only when we decide to measure them, this poses something of a problem! Regular longitudinal structural equation modeling and network modeling approaches just can't represent the data generating processes that researchers typically hypothesize, but the availability of modeling tools and mathematical understanding leads to their usage anyway.

Bringing Theory and Statistics Closer with Continuous-Time Dynamic Systems Modeling

Continuous-time dynamic systems are essentially a stochastic differential equation, paired with a measurement model. Many fields in the 'natural' sciences are well acquainted with differential equations, they are in a sense the natural mathematical language for continuously changing processes, and the 'stochastic' addition simply allows for uncertainty in the direction of change. Social scientists tend to be raised on a diet of regression, leading to models of change described in terms of regressions over time. Such an approach is certainly not 'wrong', but it is important to understand the limitations, and there may be circumstances where it is advantageous to switch to a differential equation based approach. While it certainly does not resolve all difficulties in linking theory to statistics, a significant virtue of any such differential equation approach is that it *does* resolve (or at least improve on) the issue discussed above, of misspecified temporal structure, when processes change and interact continuously.

Continuous-time / differential equation models have been discussed in the social sciences as early as Coleman (1964), with substantial further development closer to psychology by Singer (1993), Oud and Jansen (2000), S.-M. Chow et al. (2005) and Voelkle et al. (2012), amongst others. While software from the natural sciences / mathematics can of course be used for social science purposes, typical features of the research questions and data can make packages developed more explicitly with a social science orientation more functional and accessible. Such software packages include `ctsem` (Driver et al., 2017), `dynr` (Ou et al., 2019), and `BHOUM` (Oravecz et al., 2016). This article will describe models at a level of abstraction applicable to any software that can fit multivariate stochastic differential equations with a measurement model, but will at times also include additional labelling of matrices relevant to the `ctsem` software, and code for model fitting with `ctsem` plus additional results are available in the supplementary material.

Some additional benefits of longitudinal modeling in the `ctsem` framework, as distinct from classical structural equation modeling, are the possibilities of specifying a wider range of temporal relations, including features such as time-varying parameters, moderated correlations, random and covariate based individual differences in any parameters, and an easy shift between frequentist and Bayesian approaches. In the present article we do not explore the limits of the software, but instead endeavour to connect readers more familiar with regression and structural equation approaches to the possibilities for both thinking about and specifying theories of change as continuous-time dynamic systems.

Empirical Research Agenda

We have discussed three theories that describe alternative ways that mathematics and language performance may relate. The thinking-function hypothesis poses a common cause behind the two, the medium-function hypothesis poses supportive relations from language to mathematics, and a competitive hypothesis suggests that focusing learning in either may detract from the other. We do not propose any strong test of the theories here, the different theories may in fact all be plausible together and it may be a question of time span and granularity, however we plan to build and fit models of competence development that can encapsulate some aspects under consideration. This will hopefully both provide some indications of theoretical support or debunking, as well as provide a basis from which additional models and or data may be included and compared. Given the nature of the data available to us, we will focus attention on change at time-scales ranging in the approximately weekly to yearly range.

Data

The data we use come from the `Mindsteps` online learning platform (see `https://www.mindsteps.ch/`), which offers teachers and students in a variety of Swiss regions many thousands of questions in a range of subjects to practice and test from.

A distinctive feature of the system is that it covers topics and competencies from the third grade in elementary school until the third grade in secondary school, thereby spanning seven years of compulsory schooling. The item bank is based on a competency-based approach to learning (see Sampson and Fytros, 2008) that emphasizes learning progress and learning outcomes during the learning process. All items used are embedded in a curriculum, as suggested by Shepard () and Shavelson (2008). Currently, the item bank comprises up to 15,000 items per school subject.

In `Mindsteps`, there are two thematically identical types of item banks. The *p*ractice item bank is openly available to all students and teachers for training and teaching purposes. Students can autonomously use this item bank to create and answer an item set from a topic domain on which they choose to or are instructed to practice. This can be done virtually anywhere that has Internet access. The *t*esting item bank, in turn, can be used to evaluate students' ability levels and learning progress, as well as identify their strengths and weaknesses in a given content domain. Teachers can select items according to desired competency domains, single competencies, or topics of the official curriculum and create assessments that can be taken by students on computers at school. There are three 'use cases' for this item bank that result in three different types of feedback. First, teachers may want to assess their students' ability levels or learning progress regarding a general competency domain, such as reading comprehension or algebra. Second, teachers might be interested in a single competency among their students, such as comprehension of simple discontinuous texts or summation in the number range of a million. Finally, teachers can administer tests on topic-specific knowledge to assess students' level of mastery. Such topics usually are very narrowly defined and often refer to the content of single instructional units.

For the present analyses, we have been using all data available from both item banks and all use cases. Each assessment consisted of at least 10 items. In a prior modeling step, data from the years 2018 to 2021 was fit to multiple large uni-dimensional item response theory based models, and individual ability scores based on point estimates for each assessment session were extracted. More details on the `Mindsteps` software, as well as discussion of an earlier iteration of the modeling approach, can be found in Berger et al. (2019). Epistemological, methodological, and practical issues of the software are discussed by Tomasik and colleagues (2018). For pragmatic (i.e., computational) reasons, we subset the available data to those students with at least 10 assessments (in any of 5 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 $N = 2{,}786$ students, and 95,649 observation occasions in total. For the purpose of this investigation, we use variables from the two German language domains, reading (*dles*) and grammar (*dsif*), as well as the three mathematics domains, 'numbers and variables' (*mzuv*), 'forms and space' (*mfur*), and 'measures, functions, and probability' (*mgfd*).

Continuous-Time Framework

The most common approaches to addressing questions of change in developmental psychology are probably multilevel regression, and structural equation modeling. While one can usually find ways to shoehorn a theory and data combination into a particular modeling approach, when the modeling approach is not a good fit for the theory and data, it can be that either theories are altered to suit the modeling, unnecessary auxiliary assumptions are required, or even, as we described above, that the theory that was thought to be examined is unknowingly not! For such reasons, we will describe the use of a continuous-time, stochastic differential equation approach. Such an approach is at times uncomfortable for the average psychological researcher, usually raised on a diet of ANOVA and regression rather than the differential equations that are more typically the domain of 'harder' sciences. While they take some time to get used to, for formalizing systems that tend to function and interact continuously, differential equations can often offer a more natural formalization with less need for compromises and 'slipperiness' between the specific formalization and the actual theory of interest. This section contains some material based on Driver (2022) and Driver and Voelkle (2018a), for more elaborated discussion one should see the sources and other references. For the less mathematically inclined, briefly skimming this section before focusing on the subsequent model building section may be a fruitful approach.

Discrete-time processes

To understand the continuous-time systems framework, it can be helpful to start from the (hopefully) more familiar 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 something like this:

$$\eta_u = A y_{u-1} + b + G z_u \quad z \sim N(0, 1) \tag{1}$$

Where η is a d length vector of process values (either observed data or hypothetical latent states), u is an index of measurement occasion, A is a matrix of temporal regression coefficients, b is an intercept, and G is the effect matrix of the d length system noise vector z, where z contains independent and identically distributed deviations with zero mean. Many extensions and variations are in use, non-linear forms may be written more generically as functions, but the issues discussed herein, while broadly applicable, are simplest to understand using the basic form shown. For a recent overview contrasting different approaches to cross-lagged models and software, see Ruissen et al. (2021).

Continuous-Time Processes

In discrete-time systems, the matrix of temporal effects represents regression strengths between two points in time, as in the day or two day examples discussed in the earlier letter writing example. Continuous-time approaches can be intuitively thought of in much the same way, but simply compressing the time interval to a 'very small' value. This allows for usage of a conceptually similar temporal effects matrix, the only difference being that the matrix represents the influence of the current state of the system on the *change* in the system – which is actually analogous to the approach used in change-score structural equation models, where an autoregression of 1.00 is always applied in addition to any estimated self and cross feedback parameters. A continuous-time form of the vector autoregression discussed is:

$$\mathrm{d}\eta(t) = (A\eta(t) + b)\mathrm{d}t + G\mathrm{d}W(t) \tag{2}$$

This looks similar to the discrete-time form, but instead of telling us the new value of η given one step forward in time, it tells us how y is changing *at the moment*. Some mathematical complications due to the nature of stochastic differential equations are present: the $\mathrm{d}t$ on the right hand side, which can simply be thought of as a very small step in time, and the $\mathrm{d}W(t)$, which represents white noise in continuous time. To compute η at some time point given an earlier value, one needs to solve the system. This is numerically involving and described in detail in Voelkle et al. (2012), however simple approximations using linear extrapolation can be done by hand, and are quite helpful – both for numerical simulation and to follow the basic logic.

$$\eta_u = \eta_{u-1} + \Delta t(A\eta_{u-1} + b) + G z_u \quad z \sim N(0, \Delta t)) \tag{3}$$

Essentially then, one computes the deterministic rate of change using A and b and the earlier value of η, then multiplies this by the length of time step Δt desired. The variance of the white noise element z is then also Δt, resulting in a system noise (co)variance of $GG^\top \Delta t$. Shorter steps are (up to a point) more accurate, but require more computations. Many refinements are possible – this simply represents the basic idea, which is called the Euler-Maruyama method in the context of stochastic differential equations. The linear system shown also allows an 'exact' (to numerical tolerances, anyway) one-step solution. While this is described in more detail elsewhere (e.g. Driver & Voelkle, 2018a, contains the equations and code for plotting), the basic component involves the matrix exponential, with temporal regression coefficients for particular time intervals given by $e^{A\Delta t}$.

Measurement Model

The basic linear measurement model for the system can be represented as:

$$\mathbf{y}(t) = \boldsymbol{\Lambda}\boldsymbol{\eta}(t) + \boldsymbol{\tau} + \boldsymbol{\epsilon}(t) \quad \text{where } \boldsymbol{\epsilon}(t) \sim \mathrm{N}(\mathbf{0}_c, \boldsymbol{\Theta}) \tag{4}$$

y is the c length vector of manifest variables, $\boldsymbol{\Lambda}$ represents the factor loadings, and $\boldsymbol{\tau}$ the manifest intercepts. The factor loadings and manifest intercepts govern how the latent process looks when observed by one or more of the indicators – the loading sets the scaling factor, and the intercept the offset. In cases with only one measured variable per latent process, they are commonly set to 1 and 0 respectively to ensure the model is identified, and does not have multiple parameters that can be varied to give the same predictions. The manifest residual vector $\boldsymbol{\epsilon}$ has covariance matrix $\boldsymbol{\Theta}$, and this captures the proportion of variation in the observed variables that does *not* aid in predicting future states of the system. This can be thought of simply as measurement error due to randomness, as genuine fluctuations on such a short time scale that it is more effective to ignore them, or as some mixture of the two.

Individual Differences

Stable individual differences in any of the process or measurement model components can be included. Given sufficient data, the model may be fit separately for each individual, resulting in unique parameter estimates for every subject. Large and informative enough datasets 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 be to 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 a little heterogeneous and confusing across fields. Approaches to actually estimate individual differences in parameters such as a standard deviation or correlation can be somewhat complicated, we will keep discussion of such aspects to a minimum, but for further details see **empty citation**.

Different Forms of Uncertainty – System and Measurement Noise

One of the core benefits to the state-space modeling approach, in which a latent system model is coupled to some form of measurement model, is the ability to decompose the variance of the system into two (or more) portions. The measurement error portion contains variation that is essentially discarded from one observation to the next, while a 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 visual of Figure 3 may help with this understanding.

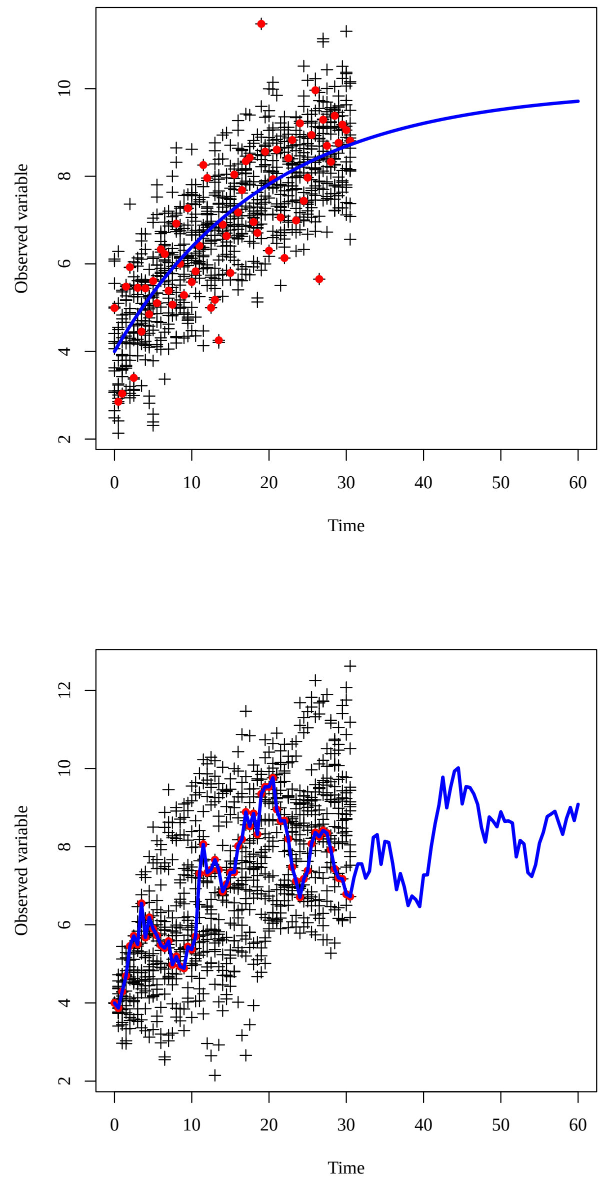

Figure 3

These two plots show the distinction between measurement error, on the top, and system noise, on the bottom. The plots show a set of data from multiple subjects as black points, the red points refer to a single individuals observations, and the blue line refers to that individuals latent process. In the top plot, our knowledge of the specific individuals points helps very little with predicting the trajectory of the latent process, because deviations away from the central tendency of the group are simply measurement error. In the lower plot, such deviations represent true change in the process, and are informative for future predictions.

Model Building

To illustrate how the different components of the process and measurement models can be used and combined to represent a wide range of developmental theories, we will walk through a progression of models. We will start with a simple linear growth model, then incorporate individual differences, nonlinear trends, common and or causal fluctuations, and multiple time-scales. While we describe such an approach here largely for tutorial purposes, it is generally good practice in model building to start simple, and build up from the large foundational components to the more sophisticated. This can also provide a form of sensitivity checking, in that if different models lead to wildly different inferences for reasons that are not obvious, caution in interpretation is definitely warranted! Note also that in terms of inference, we are using the maximum a posteriori estimation approach of `ctsem` (Driver & Voelkle, 2021), which can also be thought of as a form of penalised likelihood. For general parameters in a dataset of this size this will have very little influence, but when we discuss moderated parameters in complex models, it can reduce capitalising on chance results by keeping estimates more conservative and 'boring' as such. Results from all model fits, both with priors and without (i.e., maximum likelihood) are available in the supplementary material.

Linear Growth

While linear growth without individual differences is a highly unrealistic specification for most, if not all, developmental processes, it is a helpful starting point from which to develop understanding and build upon. In this formulation, the basic process (Equation 2), can be simplified by dropping matrices of zeroes, to just:

$$\mathrm{d}\eta(t) = b\mathrm{d}t \tag{5}$$

Which tells us quite simply that the slope of change (d) in our maths and language latent processes η at time (t) is determined by the continuous intercept b, multiplied by the change in time dt. In the model we are developing there are two latent processes, language and maths, so η and b are vectors of length two. The measurement model is as already shown in Equation 4, and in this case involves 5 indicator variables. Two variables are used to measure language skill, and three to measure mathematics. The first indicator for each latent process is constrained to a factor loading of 1.00 and an intercept of 0.00, to ensure identification of of the model. After fitting the model, we obtain estimates translate to growth expectations as shown in Figure 4. The point estimates of the model fit are shown here in expanded matrix form, with zero matrices left in place to facilitate understanding. Note that throughout this work, lower triangular matrices along with a matrix transformation function *UcorrSDtoCov* or *UcorrSDtoChol* are occasionally used in place of a covariance matrix. This transformation function takes a matrix of standard deviation and 'unconstrained correlation' parameters, ensuring that the free parameters can take on any value and still result in a positive-definite covariance matrix, or the Cholesky factor of a covariance matrix. See Driver and Voelkle (2018a) for more details:

Initial latent state:
$$\underbrace{\begin{bmatrix}\text{lang}\\ \text{maths}\end{bmatrix}(t_0)}_{\eta(t_0)} \sim \mathrm{N}\left(\underbrace{\underbrace{\begin{bmatrix}-0.03\\ -0.07\end{bmatrix}}}_{\text{T0MEANS}}, \underbrace{\underbrace{\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}}_{\mathbf{Q^*}_{t0}}}_{\text{T0VAR}}\right)$$

Deterministic change:
$$\underbrace{\mathrm{d}\begin{bmatrix}\text{lang}\\ \text{maths}\end{bmatrix}(t)}_{\mathrm{d}\eta(t)} = \left(\underbrace{\underbrace{\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}}_{\mathbf{A}}}_{\text{DRIFT}} \underbrace{\begin{bmatrix}\text{lang}\\ \text{maths}\end{bmatrix}(t)}_{\eta(t)} + \underbrace{\underbrace{\begin{bmatrix}0.17\\ 0.21\end{bmatrix}}_{\mathbf{b}}}_{\text{CINT}}\right)\mathrm{d}t \quad +$$

Random change:
$$\underbrace{\underbrace{UcorSDtoChol\left\{\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}\right\}}_{\mathbf{G}}}_{\text{DIFFUSION}} \underbrace{\mathrm{d}\begin{bmatrix}W_1\\ W_2\end{bmatrix}(t)}_{\mathrm{d}\mathbf{W}(t)}$$

Observations:
$$\underbrace{\begin{bmatrix}\text{dles}\\ \text{dsif}\\ \text{mfur}\\ \text{mgfd}\\ \text{mzuv}\end{bmatrix}(t)}_{\mathbf{Y}(t)} = \underbrace{\underbrace{\begin{bmatrix}1 & 0\\ 1.25 & 0\\ 0 & 1\\ 0 & 0.87\\ 0 & 1.24\end{bmatrix}}_{\mathbf{\Lambda}}}_{\text{LAMBDA}} \underbrace{\begin{bmatrix}\text{lang}\\ \text{maths}\end{bmatrix}(t)}_{\eta(t)} + \underbrace{\underbrace{\begin{bmatrix}0\\ -0.06\\ 0\\ 0.01\\ 0.08\end{bmatrix}}_{\tau}}_{\text{MANIFESTMEANS}} \quad +$$

Observation noise:
$$\underbrace{\underbrace{\begin{bmatrix}1.11 & 0 & 0 & 0 & 0\\ 0 & 1.06 & 0 & 0 & 0\\ 0 & 0 & 1.11 & 0 & 0\\ 0 & 0 & 0 & 1.13 & 0\\ 0 & 0 & 0 & 0 & 1.13\end{bmatrix}}_{\mathbf{\Theta}}}_{\text{MANIFESTVAR}} \underbrace{\begin{bmatrix}\epsilon_1\\ \epsilon_2\\ \epsilon_3\\ \epsilon_4\\ \epsilon_5\end{bmatrix}(t)}_{\epsilon(t)}$$

System noise distribution per time step: $\Delta[W_{j\in[1,2]}](t-u) \sim \mathrm{N}(0, t-u)$ Observation noise distribution: $\left[\epsilon_{j\in[1,2]}\right](t) \sim \mathrm{N}(0,1)$

Individual Differences in Linear Growth

Of course, single values for the expected initial values and growth across individuals and age is an obscenely unlikely proposition, a big reason we are interested in such models is because individuals differ in growth! So an obvious next step is to allow for individual differences in these parameters. Such individual differences 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` parlance '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' in the model. For random variation in any parameters other than the initial intercept,

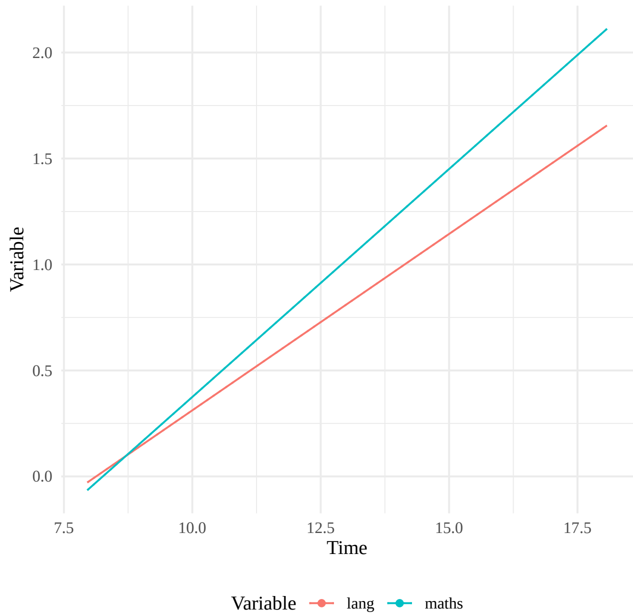

Figure 4

Expectations from linear growth model fit.

these need to be included as additional latent processes in the system model, which exert appropriate influence depending on which type of parameter they are (Note that for users of `ctsem`, this is automated and occurs in the background when individually varying parameters are requested). Continuous intercepts are one of the simplest parameter types to include, they come in as additional processes that exert an influence of 1.00 on the original processes of maths and language. This influence occurs via the temporal effects, or drift, matrix that we could previously ignore as it contained only zeroes. Once this is configured, the b vector from Equation 2 is no longer needed – the relevant components are now part of the extended state vector η. Equation 6 shows the point estimates for this expanded initial latent state distribution.

$$\underbrace{\begin{bmatrix} \text{language} \\ \text{maths} \\ \text{cint\_language} \\ \text{cint\_maths} \end{bmatrix}(t_0)}_{\eta(t_0)} \sim \mathrm{N}\left(\underbrace{\begin{bmatrix} -0.1 \\ -0.23 \\ 0.31 \\ 0.34 \end{bmatrix}}_{\text{T0MEANS}}, \underbrace{\begin{bmatrix} 0.56 & 0.48 & -0.08 & -0.06 \\ 0.48 & 0.56 & -0.07 & -0.09 \\ -0.08 & -0.07 & 0.09 & 0.07 \\ -0.06 & -0.09 & 0.07 & 0.1 \end{bmatrix}}_{\underbrace{\mathbf{Q}^*_{t0}}_{\text{T0VAR}}} \right) \quad (6)$$

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):

$$\underbrace{\mathrm{d}\begin{bmatrix} \text{language} \\ \text{maths} \\ \text{cint\_language} \\ \text{cint\_maths} \end{bmatrix}(t)}_{\mathrm{d}\eta(t)} = \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\underbrace{\mathbf{A}}_{\text{DRIFT}}} \underbrace{\begin{bmatrix} \text{language} \\ \text{maths} \\ \text{cint\_language} \\ \text{cint\_maths} \end{bmatrix}(t)}_{\eta(t)} \quad (7)$$

Our specified measurement model appears at least not obviously terrible, 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):

$$\underbrace{\begin{bmatrix} \text{dles} \\ \text{dsif} \\ \text{mfur} \\ \text{mgfd} \\ \text{mzuv} \end{bmatrix}(t)}_{\mathbf{Y}(t)} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.95 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1.03 & 0 & 0 \\ 0 & 0.94 & 0 & 0 \end{bmatrix}}_{\underbrace{\mathbf{\Lambda}}_{\text{LAMBDA}}} \underbrace{\begin{bmatrix} \text{lang} \\ \text{maths} \\ \text{cint\_lang} \\ \text{cint\_maths} \end{bmatrix}(t)}_{\eta(t)} + \underbrace{\begin{bmatrix} 0 \\ -0.09 \\ 0 \\ 0.02 \\ 0.15 \end{bmatrix}}_{\underbrace{\boldsymbol{\tau}}_{\text{MANIFESTMEANS}}} + \quad (8)$$

$$\underbrace{\begin{bmatrix} 0.82 & 0 & 0 & 0 & 0 \\ 0 & 0.82 & 0 & 0 & 0 \\ 0 & 0 & 0.85 & 0 & 0 \\ 0 & 0 & 0 & 0.85 & 0 \\ 0 & 0 & 0 & 0 & 0.89 \end{bmatrix}}_{\underbrace{\mathbf{\Theta}}_{\text{MANIFESTVAR}}} \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}(t)}_{\epsilon(t)} \quad (9)$$

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. The results in Equation 6 show the pattern one would expect for such constructs, wherein the initial latent states for maths and language positively covary, the continuous intercepts do also, 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. The terminology of initial states and continuous intercepts may seem 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 $\eta(t_0)$ is equivalent to the intercept term in a growth model, and the continuous intercept is equivalent to the linear slope. Once more complex models are used, this one to one relation breaks down to a degree – for instance in systems that fluctuate around a baseline, as in many vector autoregressive type models, the continuous intercept may determine the baseline.

Linear Growth – Covariate Moderated Parameters. While the random effects approach can do a good job accounting for and understanding individual differences in the model parameters, as well as relations between such individual differences, there are often both substantive and pragmatic reasons for including additional covariate effects to moderate parameters of the model. Substantive reasons are typically when one expects that system parameters will vary along with the covariate and one wishes to account for and understand this relation. Pragmatic reasons can be that one would really wish to allow random effects on more parameters, but the computational cost and difficulties are too high. Including covariate 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 standard technique of including additional 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), for which a proof of concept was also implemented with an earlier version of the `ctsem` software (Brandmaier et al., 2018).

While we could simply allow for moderation of the same 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 the factor loadings or measurement error variance. Fitting this model results in a similar overall pattern to the unmoderated model, with age moderation effects shown in Table 1. These moderation effects indicate that, as expected, initial language and maths performance (e.g. *T0m_lang*) rises with age. The growth rate for language (*T0m_cint_lang*) appears largely unaffected by age, while the growth rate for maths also rises with age. Factor loadings (e.g. *lambda_mgfd* for one of the maths indicators) are relatively unaffected, while measurement error standard deviation for maths indicators (e.g. *mvarmfur*) reduces with age. Some indicators *do* show a tendency to a higher measurement intercept parameter (e.g. *mm_dsif*) with increasing age. This tendency suggests that, as we would expect given the clear structure in the data and the amount of it, additional factors may improve the model from a statistical standpoint – there is some residual growth in the indicators that is not accounted for by the latent processes. In terms of the standard deviations and correlations amongst 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 (*T0var_cint_maths_maths*) appears to *increase* with age, while the relation between mathematics growth and language growth (*T0var_cint_maths_cint_lang*) appears to decrease. While such results are useful for inferences regarding direction of effects, it is difficult to comprehend the magnitude of effects due to the nonlinear matrix transformations used 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 5 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.

Table 1

'Age moderation effects on the linear latent growth with random effects model.

	Est.	SD	2.5%	97.5%
T0m_lang	0.05	0.01	0.04	0.07
T0m_maths	0.07	0.01	0.05	0.09
T0m_cint_lang	0.00	0.01	-0.01	0.01
T0m_cint_maths	0.03	0.01	0.01	0.04
lambda_dsif	0.00	0.01	-0.01	0.01
lambda_mgfd	0.00	0.01	-0.02	0.01
lambda_mzuv	-0.01	0.01	-0.03	0.00
mvardles	0.00	0.00	-0.01	0.00
mvardsif	-0.02	0.00	-0.03	-0.02
mvarmfur	-0.03	0.00	-0.03	-0.02
mvarmgfd	-0.03	0.00	-0.03	-0.02
mvarmzuv	-0.02	0.00	-0.03	-0.02
mm_dsif	0.03	0.00	0.02	0.04
mm_mgfd	0.01	0.01	0.00	0.02
mm_mzuv	0.04	0.01	0.03	0.05
T0var_lang	0.00	0.01	-0.02	0.02
T0var_maths_lang	-0.01	0.01	-0.02	0.01
T0var_maths	0.00	0.01	-0.02	0.01
T0var_cint_lang_lang	0.02	0.01	0.00	0.04
T0var_cint_lang_maths	0.02	0.01	-0.01	0.04
T0var_cint_lang	0.00	0.01	-0.01	0.01
T0var_cint_maths_lang	0.01	0.01	-0.01	0.04
T0var_cint_maths_maths	0.04	0.01	0.02	0.07
T0var_cint_maths_cint_lang	-0.07	0.03	-0.12	-0.02
T0var_cint_maths	0.01	0.01	0.00	0.02

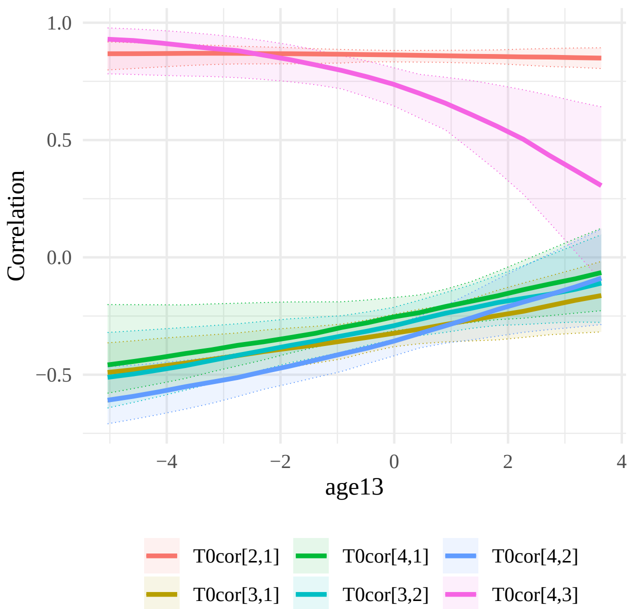

Figure 5

Expected correlations between initial states and slopes, conditional on the (centered at 13 years) age covariate, for the moderated linear growth model. Index 1 and 2 are initial language and maths performance, respectively, while 3 and 4 are language and maths slopes, respectively. 95% confidence intervals are shaded.

Including Dynamics

So far, we have only developed a 'static' model of change – the model predictions for an individuals state at any point in the future depends only on the system parameters for that individual, and time (Voelkle et al., 2019). If we shift to a truly 'dynamic' model, in which we acknowledge that unpredictable changes occur in our constructs of interest, the model is likely to perform (i.e., predict) better, and can be used to consider a range of new hypothesis regarding relations between these unpredictable fluctuations. For instance, while long term maths and language performance trends tend to correlate, shorter term fluctuations could be driven by a 'more common cause' such as motivation, or may show reduced correlation due to competitive processes. 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, otherwise the variance goes to infinity as everything of the past is kept and new noise is continually added. To relax this assumption we 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* it's deterministic trend, the negative dependency on earlier states pushes the system back *down* towards the expected trend, and vice versa for downwards fluctuations. Positive dependencies are also plausible over limited time-spans, 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 wholly dependent on 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, and we do not allow for any directionality between these changes at present. Directionality, wherein changes in one process lead to changes in the other, will be incorporated in a later step.

By freeing the auto-effects in the temporal dependency matrix, the deterministic trend becomes non-linear. Instead of the trend being simply linear, the trend now 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 grow linearly. The updated formulation accommodates such a concept *within-subject*.

With free auto-effects and correlated stochastic change included, the point estimates for the system equation are shown in Equation 10, and the expected trajectories for a typical student in the sample, *before* knowing any of their individual scores, is shown in Figure 6. 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 7 demonstrates. In this plot it is easy to see how 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 tightly follow each individual observation.

What can we really learn from our updated model however? Two main aspects regard 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 Figure 6. Regarding shorter term fluctuations in performance, the system noise covariance in Equation 11 shows that the random changes in language and maths performance are highly correlated, with similar variances. Such results speak *against* the notion that

$$\underbrace{\mathrm{d}\begin{bmatrix}\text{lang}\\ \text{maths}\\ \text{cint\_lang}\\ \text{cint\_maths}\end{bmatrix}(t)}_{\mathrm{d}\boldsymbol{\eta}(t)} = \left(\underbrace{\begin{bmatrix}-0.64 & 0 & 1 & 0\\ 0 & -0.66 & 0 & 1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}}_{\underbrace{\mathbf{A}}_{\text{DRIFT}}}\underbrace{\begin{bmatrix}\text{lang}\\ \text{maths}\\ \text{cint\_lang}\\ \text{cint\_maths}\end{bmatrix}(t)}_{\boldsymbol{\eta}(t)} + \underbrace{\begin{bmatrix}0\\0\\0\\0\end{bmatrix}}_{\underbrace{\mathbf{b}}_{\text{CINT}}}\right)\mathrm{d}t+ \tag{10}$$

$$\underbrace{Cholesky\left\{\begin{bmatrix}0.38 & 0.34 & 0 & 0\\ 0.34 & 0.48 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\right\}}_{\underbrace{\mathbf{Q}}_{\text{DIFFUSIONcov}}}\underbrace{\mathrm{d}\begin{bmatrix}W_1\\ W_2\\ W_3\\ W_4\end{bmatrix}(t)}_{\mathrm{d}\mathbf{W}(t)} \tag{11}$$

there may be competitive relations in the short term between 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, the competitive hypothesis would have been more plausible. Distinguishing between the thinking and medium function hypotheses is not yet possible, given the lack of directionality in our model. When we consider the age moderation effect on the system noise correlation, there *is* some influence, with the system noise correlation reducing in older children, shown in Figure 8. The most obvious interpretation of this result is that in older (or perhaps higher performing) children, fluctuations in performance are due somewhat less to common causes, such as general motivation, and more to unique(er) factors such as time spent studying for the particular subject recently.

Multiple Time Scales and Directional Dynamics

The results from such a dynamic model can be highly informative, and once individual differences in parameters are included sometimes verging on overwhelming. There is, however, a need for caution in interpretations. One particular issue in the model specification as it stands presently is that the shape of the long term trends are 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 – they set the curvature of the overall deterministic trend, as well as how long the short term fluctuations persist in the system. This kind of issue 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 resolve this, we can expand the system in a similar way

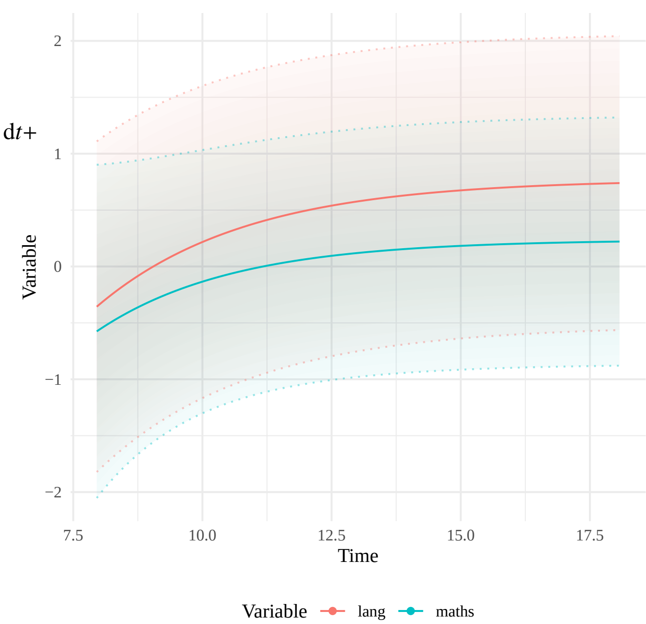

Figure 6

Expected trends before conditioning on data, once auto-effects and stochastic change are included in the model. 95% confidence intervals are shaded.

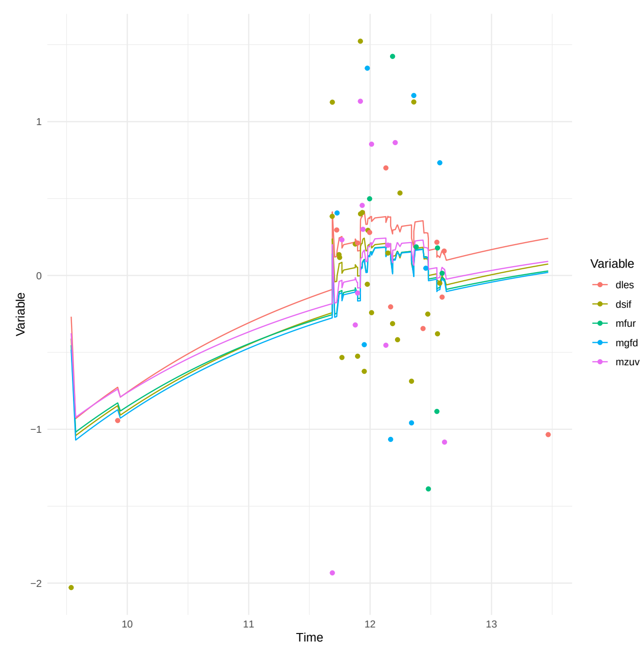

Figure 7

Predictions for one student based on past data, once auto-effects and correlated stochastic change are included in the model.

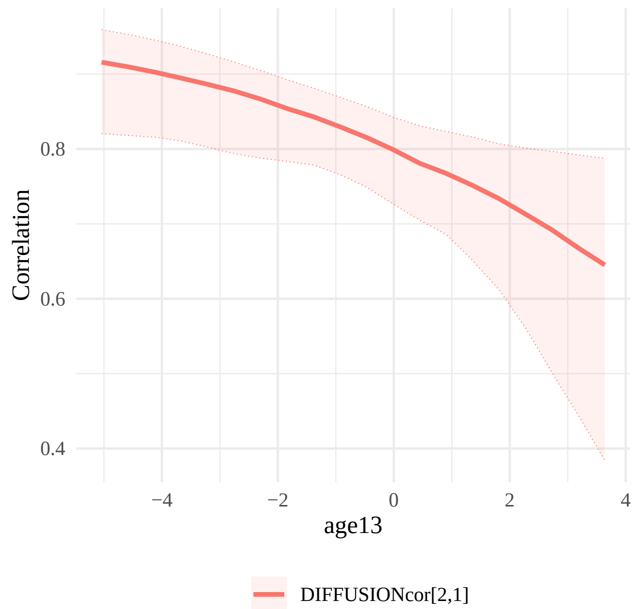

Figure 8

Correlation between random changes in language and maths performance, conditional on the (centered at 13 years) age covariate, for the model with freely estimated auto-effects and system noise. 95% confidence intervals are shaded.

$$\underbrace{\begin{bmatrix} -0.11 & 0 & 1 & 0 & 0 & 0 \\ 0 & -0.05 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6.89 & -0.27 \\ 0 & 0 & 0 & 0 & -2.02 & -3.28 \end{bmatrix}}_{\underbrace{\mathbf{A}}_{\text{DRIFT}}} \begin{bmatrix} \text{language} \\ \text{maths} \\ \text{cint\_language} \\ \text{cint\_maths} \\ \text{dynlanguage} \\ \text{dynmaths} \end{bmatrix} \tag{12}$$

$$\underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.56 & 1.33 \\ 0 & 0 & 0 & 0 & 1.33 & 1.43 \end{bmatrix}}_{\underbrace{\mathbf{Q}}_{\text{DIFFUSIONcov}}} \begin{bmatrix} \text{language} \\ \text{maths} \\ \text{cint\_language} \\ \text{cint\_maths} \\ \text{dynlanguage} \\ \text{dynmaths} \end{bmatrix} \tag{13}$$

as when we included the continuous intercepts as a random effect. In this case, we create an extra process containing the system noise for each of language and maths, fix various parameters so this process is centered around zero, and estimate separate auto-effects for each. We then duplicate the first two columns of the factor loading matrix Λ, as these additional dynamics processes will not interact with what are 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.

In addition to separating trend from dynamics, we will also address the question of directionality between fluctuations now. We achieve this by freeing 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 12 and the system noise matrix of Equation 13

In the model with only one auto-effect for language and one for maths, the auto-effects were approximately -0.60, resulting in autoregressions after 1 year of $exp(-0.60 \times 1) = 0.55$. Now that we have allowed for different time-scales by splitting trends and dynamics, auto-effects for the trend components are close to zero, resulting in expected trends that look much closer to the linearity of our initial simpler models. The time-scale of the dynamics components are now much faster (more negative), and there are negative cross-effects between the two. These dynamics may be easier understood by referring to Figure 9, which shows that an upwards shift in language performance tends to be followed in the coming months by a *drop* in maths performance. The reverse effect, of maths on language, is also slightly negative but the confidence interval heavily overlaps zero. This pattern of results is very much contrary to the medium-function hypothesis, which suggests that improvements in language should lead to *improvements* in maths, and more so than maths improves language ().

The typical interpretation of such a negative cross-effect partly aligns with the idea that time is scarce, and moments spent focusing on one domain inevitably mean other domains may suffer. However, such an interpretation is hard to reconcile with the high correlation in the system noise term. Plotting the temporal dynamics in *combination with* with the system noise, as suggested and described in detail by Driver (2022), means Figure 10 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 lower magnitude. In such a scenario, the negative cross-effect from language to maths essentially just means that when language rises, so too does maths, but that this perturbation in maths *dissipates faster than is typical*. So, when maths performance fluctuates *together* with language performance, the change tends to dissipate quicker than when maths performance fluctuates *alone* or in the *opposite direction* to language performance.

Interestingly, when we look at the age moderation effect on system parameters in Table 2, both cross-effect parameters become more negative with age. While we would still

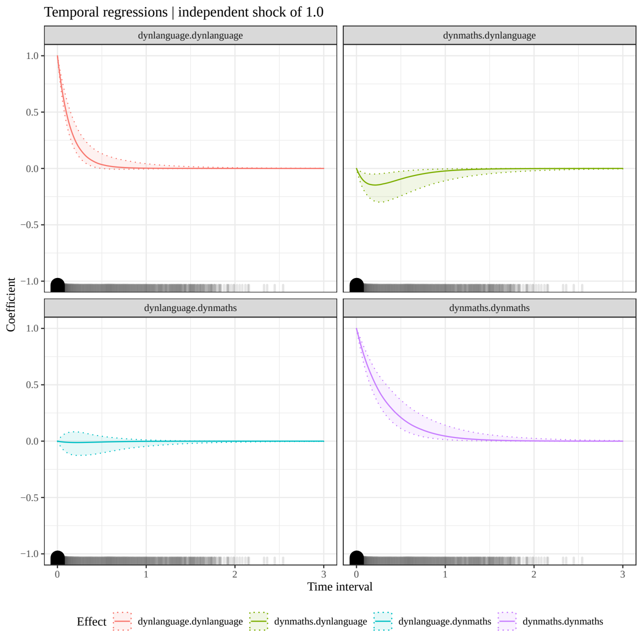

Figure 9

Temporal dynamics of fluctuations in maths and language performance, for the multiple time-scale model with directional effects, assuming independent initial changes. 95% confidence intervals are shaded. At time zero, the second process in the title is given an exogenous perturbation of 1.00, leading to changes in the first process. This order relates to standard row, column indexing for matrices and matches the temporal effects matrix.

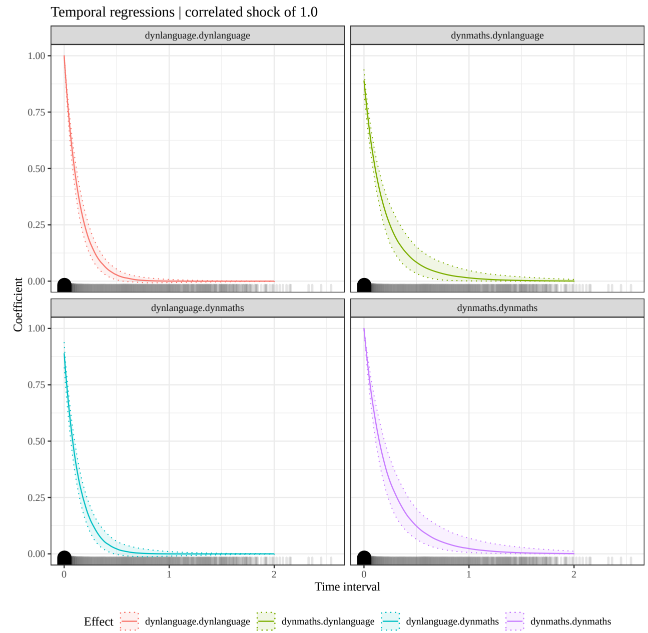

Figure 10

Temporal dynamics of fluctuations in maths and language performance, for the multiple time-scale model with directional effects, after accounting for correlated system noise. 95% confidence intervals are shaded. At time zero, the second process in the title receives a perturbation of 1.00, while the first process receives the amount of perturbation expected according to the system noise correlation. Plots then show how the perturbation propagates through the first process. This order relates to standard row, column indexing for matrices and matches the temporal effects matrix.

urge caution because of the strong positive correlation in system noise, this increasing negative relation with age *does* neatly map onto some ideas regarding increasing competition for resources between domains as students progress in their education ().

Discussion

Researchers in developmental and educational science are inclined to use the notion of "dynamics" (e.g.,), sometimes in an inflationary way and without having a theoretical or empirical foundation for doing so. Dynamics, sometimes it seems, is then used to refer to the complexity of a process that is simply not very well understood. Or to change in general that is inherent to any developmental or educational process.

On closer inspection, dynamics can have a variety of meanings. First, dynamics might refer to short-term fluctuations or deviations from a longer term developmental trajectory of one developmental process. This up and down can either be treated as unsystematic error as it is often done in latent growth modeling when researchers assume some functional (e.g., linear) growth process over time and discount deviations from the respective function into the error part of the model. A more nuanced approach to capture this type of dynamics is provided by introducing the distinction between states and traits (e.g., Steyer et al., 1999) and by using respective statistical models to distinguish both from error (e.g., Eid et al., 2017). It goes without saying that repeated measurements are necessary to empirically capture dynamics even in this most simple interpretation.

The notion of dynamics can also be used to paraphrase the acceleration or deceleration of a developmental process. Such changes of pace can either occur when a process is self-reinforcing or self-weakening or when there are endogenous (e.g., maturation) or exogenous (e.g., intervention program) factors that influence it. Technically speaking, such dynamics can be described in terms of either non-linear parameterizations of a latent growth model or by inspecting the derivatives of a growth function. Repeated measurements with a minimum number of measurement occasions (e.g., at the very least least three for a saturated quadratic growth model)

Table 2

'Age moderation effects for the multiple time-scale model with directional effects model.

	Est.	SD	2.5%	97.5%
T0m_language	0.05	0.01	0.04	0.07
T0m_maths	0.07	0.01	0.05	0.09
T0m_cint_language	0.01	0.00	0.00	0.02
T0m_cint_maths	0.04	0.01	0.03	0.04
lambda_dsif	0.00	0.01	-0.01	0.01
lambda_mgfd	0.00	0.01	-0.02	0.02
lambda_mzuv	-0.02	0.01	-0.03	0.00
drift_language	0.01	0.02	-0.03	0.04
drift_maths	-0.05	0.02	-0.08	-0.03
drift_dynlanguage	0.45	0.37	-0.29	1.15
drift_dynlanguage_dynmaths	-1.94	0.41	-2.75	-1.17
drift_dynmaths_dynlanguage	-0.34	0.16	-0.66	-0.04
drift_dynmaths	-1.95	0.44	-2.85	-1.08
diff_dynlanguage	0.08	0.03	0.03	0.14
diff_dynmaths_dynlanguage	0.03	0.02	-0.01	0.07
diff_dynmaths	0.26	0.04	0.19	0.34
mvardles	-0.01	0.00	-0.01	0.00
mvardsif	-0.02	0.00	-0.03	-0.02
mvarmfur	-0.04	0.00	-0.04	-0.03
mvarmgfd	-0.03	0.00	-0.04	-0.02
mvarmzuv	-0.03	0.00	-0.03	-0.02
mm_dsif	0.03	0.00	0.02	0.03
mm_mgfd	0.01	0.00	0.00	0.01
mm_mzuv	0.04	0.00	0.03	0.04
T0var_language	0.00	0.01	-0.02	0.01
T0var_maths_language	-0.01	0.01	-0.03	0.01
T0var_maths	-0.01	0.01	-0.02	0.01
T0var_cint_language_language	0.02	0.03	-0.04	0.08
T0var_cint_language_maths	0.01	0.03	-0.05	0.07
T0var_cint_language	0.02	0.01	-0.01	0.04
T0var_cint_maths_language	0.11	0.03	0.05	0.16
T0var_cint_maths_maths	0.19	0.03	0.12	0.25
T0var_cint_maths_cint_language	-0.03	0.04	-0.10	0.04
T0var_cint_maths	0.03	0.01	0.01	0.06

are a necessary prerequisite for investigating this kind of dynamics.

A third meaning of the notion of dynamics emerges when referring to multiple developmental processes occurring at the same time. Related to this notion is the conceptualization of development as being multidimensional and multidirectional (e.g., Baltes, 1987). In other words, growth or decline may occur in different domains of functioning (e.g., physical and cognitive) or within different areas of the same domain (e.g., competence development in language and mathematics) *and* the different domains or areas change at a different pace or even in different directions. This understanding of dynamics requires multivariate study designs above the already mentioned prerequisite of longitudinal data collection.

Fourth, dynamics can be understood from a developmental systems perspective as the reciprocal relations between two or more developmental processes. These processes may be situated at the same level of functioning or comprise different levels. Researchers having this understanding of dynamics may want to investigate the repercussions of changes in one developmental process on changes in another one. Notably, understanding dynamics from this perspective typically includes all the other meaning outlined above. Long-term and short-term fluctuations on different variables that change over time in terms of pace and direction, that influence each other and that are subject to influence by other processes and variables, is how developmental systems can be conceptualized on a very abstract level.

In the present paper, we have put forward a perspective on the concept of dynamics that differs from the ones just mentioned not in the sense that it adds another layer to the complexity of the subject matter (because the fourth perspective mentioned already is the most complex one that can be conceived). Our perspective rather mathematically formalizes the aspects mentioned, in a framework that may reduce the sometimes imprecise use of the term "dynamics" in developmental and educational literature. We have applied this formalization to a specific substantive problem concerning competence development in educational settings. By doing so, we have demonstrated a viable approach to formalizing theory, adding some level of precision and quantification to research in this domain of dynamics and change.

Limitations

While the `Mindsteps` online-learning data are wonderfully rich in some dimensions, allowing scope for addressing questions of short-term change in educational domains, the unsurprisingly large amount of measurement uncertainty, as well as the relatively short observational period for most students (median age range is 1.73 years), does somewhat hinder the endeavour to distinguish long term trends from genuine short term variation. Another year or two of the `Mindsteps` system running will likely offer more 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.

Likely contributing somewhat to the measurement uncertainty is the fact that the data are from a 'low stakes' environment, meaning that there is little incentive for students to perform, ensuring that each score will surely be influenced by the students motivational level at the time of test taking. This aspect may have led to the high between domain correlations also. An approach to calculating the ability scores that takes into account response times and other contextual information about the assessment situation (such in-class mandatory assessments vs. free practicing at home) may help in future to somewhat disentangle the motivational aspects.

From a modeling perspective, the two-step approach, in which ability estimates for each occasion were first generated from raw data, is inevitably not ideal, but at present necessary for computational reasons. While we believe that including the age-moderated measurement model should have minimised the cost of such a two-step approach, we are investigating possibilities for combining both steps.

One direction that was not addressed here but would be of substantial interest to the online learning community, would be to include specific effects of each assessment session. That is, explicitly acknowledging that the assessment session is not purely a measurement but should have some incremental benefit on the students performance. The `ctsem` framework offers an obvious approach to this in terms of including exogenous time-varying covariates as input effects, discussed in detail in Driver and Voelkle (2018b).

Conclusion

Probably the most striking empirical result from this work is the high correlation between short term fluctuations in performance of maths and language. Unfortunately, there are a variety of possible reasons (and hence explanations) for this high correlation that will take further research to disentangle – is it an artifact of our low-stakes measurement procedure or measurement model, or perhaps genuine fluctuations in a common cause like motivation? With respect to the three theories relating maths and language considered, the only directional results we have speak *against* the medium function hypothesis () in which language facilitates mathematics performance, as our results suggests that an upwards fluctuation in language is followed by declines (or faster dissipation of gains) in maths. The high correlation in system noise between maths and language would seem to support the thinking-function hypothesis (), at least to the extent that one is comfortable assuming that the relatively short term fluctuations in performance addressed here are analogous to the sources of longer-term growth. Although this high correlation in system noise would seem to speak *against* the idea that learning of 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 specialisation, does fit nicely with the ideas of competition discussed by ().

In terms of the modeling framework put forward here, while it is certainly not the only approach to considering such questions, the flexibility of combining stochastic differential equations with measurement models and individual differences does offer quite some opportunities. The basics of the models discussed in this paper can easily be extended using `ctsem` or other software to include additional latent processes (cognitive performance such as working memory would be invaluable) as well as more sophisticated dynamics and measurement as needed. The cost of such opportunities as offered in this framework is sometimes found in unexpected dependencies in parameterisation, some of which we have discussed and explained in the hopes of mitigating problems. However, given that the parameterisation employed here is, arguably, closer to the way we often think about developmental processes, such dependencies are not necessarily just a 'modeling' problem but may also reflect fundamental difficulties at the theoretical level, wherein certain concepts are difficult to distinguish. By taking the step to explicitly formalise developmental theories in such a way, we also take steps to interrogate and expand our own understanding of the implications of the theories, even before any contact with data. We believe this to be a good thing.

References

Aalen, O. O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, *1987*(3-4), 177–190.

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.

Ajello, A. M., Caponera, E., & Palmerio, L. (2018). Italian students' results in the PISA mathematics test: Does reading competence matter? *European Journal of Psychology of Education*, *33*(3), 505–520. https://doi.org/10.1007/s10212-018-0385-x

Andin, J., Fransson, P., Rönnberg, J., & Rudner, M. (2015). Phonology and arithmetic in the language–calculation network. *Brain and Language*, *143*, 97–105. https://doi.org/10.1016/j.bandl.2015.02.004

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.

Baltes, P. B., Lindenberger, U., & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In R. M. Lerner (Ed.), *Handbook of child psychology: Vol. 1 Theoretical models of human development* (5. ed., pp. 1029–1143). Wiley.

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual review of psychology*, *50*(1), 471–507.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge University Press.

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*.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition* (Second). Psychology Press.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brandmaier, A. M., Driver, C. C., & Voelkle, M. C. (2018). Recursive partitioning in continuous time analysis. In *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.

Bruner, J. S., et al. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P., Sims, B. C., Jewkes, A. M., & Morrison, F. J. (2008). Age-and schooling-related effects on executive functions in young children: A natural experiment. *Child neuropsychology*, *14*(6), 510–524.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge 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.

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, *5*(2), 208–225. https://doi.org/10.1037/1528-3542.5.2.208

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. *Math. Cognit*, *1*, 83–120.

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

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Craik, F. I., & Salthouse, T. A. (2011). Lifespan cognitive development: The roles of representation and control. In *The handbook of aging and cognition* (pp. 563–607). Psychology Press.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic bulletin & review*, *3*(4), 422–433.

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.

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.

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. (2018a). Hierarchical Bayesian continuous time dynamic modeling. *Psychological methods*, *23*(4), 774–799. https://doi.org/http://dx.doi.org/10.1037/met0000168

Driver, C. C., & Voelkle, M. C. (2018b). Understanding the time course of interventions with continuous time dynamic models. In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), *Continuous time modeling in the behavioral and related sciences*. Springer International Publishing.

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://doi.org/10.1016/B978-0-12-813995-0.00034-0

Eid, M., Holtmann, J., Santangelo, P., & Ebner-Priemer, U. (2017). On the definition of latent-state-trait models with autoregressive effects. *European Journal of Psychological Assessment*.

Eysenck, H. J. (1987). Thomson's "Bonds" or Spearman's "Energy": Sixty Years On. *Mankind Quarterly*, *27*(3), 259–274.

Featherman, D. L., & Sorensen, A. (1983). Societal transformation in Norway and change in the life course transition into adulthood. *Acta Sociologica*, *26*(2), 105–126.

Fetzer, M., & Tiedemann, K. (2018). The interplay of language and objects in the process of abstracting. In *Language and communication in mathematics education* (pp. 91–103). Springer.

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.

Freund, A. M., & Tomasik, M. J. (2021). Managing conflicting goals through prioritization? The role of age and relative goal importance. *Plos one*, *16*(2), e0247047.

Freund, Alexandra 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

Geary, D. C. (2011). Cognitive Predictors of Achievement Growth in Mathematics: A Five Year Longitudinal Study. *Developmental psychology*, *47*(6), 1539–1552. https://doi.org/10.1037/a0025510

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Heckhausen, J., & Buchmann, M. (2019). A multidisciplinary model of life-course canalization and agency. *Advances in Life Course Research*, *41*, 100246. https://doi.org/10.1016/j.alcr.2018.09.002

Ho, V. M., Lee, J.-A., & Martin, K. C. (2011). The cell biology of synaptic plasticity. *Science (New York, N.Y.)*, *334*(6056), 623–628.

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

Jamieson, R. K., Johns, B. T., Taler, V., & Jones, M. N. (2022). The importance of formal modelling for the development of cognitive theory. *Bilingualism: Language and Cognition*, *25*(2), 218–219. https://doi.org/10.1017/S1366728921000420

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger/Greenwood.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A Longitudinal Study of Mathematical Competencies in Children With Specific Mathematics Difficulties Versus Children With Comorbid Mathematics and Reading Difficulties. *Child development*, *74*(3), 834–850.

Kim, M. (2018). The dynamics of moral development: Rethinking what develops and how. In *Routledge international handbook of schools and schooling in Asia* (pp. 591–599). Routledge.

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*(7), 977.

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.

Lickliter, R., & Honeycutt, H. (2003). Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychological Bulletin*, *129*(6), 819.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in cognitive sciences*, *10*(10), 464–470.

Marks, G. N. (2016). Explaining the substantial inter-domain and over-time correlations in student achievement: The importance of stable student attributes. *Educational Research and Evaluation*, *22*(1-2), 45–64. https://doi.org/10.1080/13803611.2016.1191359

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and Achievement: Just how Correlated are they? *Journal of Psychoeducational Assessment*, *21*(3), 244–260. https://doi.org/10.1177/073428290302100302

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 _eprint: 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.

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.

Riediger, M., & Freund, A. M. (2004). Interference and facilitation among personal goals: Differential associations with subjective well-being and persistent goal pursuit. *Personality and Social Psychology Bulletin*, *30*(12), 1511–1523. https://doi.org/10.1177/0146167204271184

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Ruissen, G. R., Zumbo, B. D., Rhodes, R. E., Puterman, E., & Beauchamp, M. R. (2021). Analysis of dynamic psychological processes to understand and promote physical activity behaviour using intensive longitudinal methods: A primer. *Health Psychology Review*, 1–34.

Ryan, O., & Hamaker, E. L. (2021). Time to Intervene: A Continuous-Time Approach to Network Analysis and Centrality. *Psychometrika*. https://doi.org/10.1007/s11336-021-09767-0

Sampson, D., & Fytros, D. (2008). Competence Models in Technology-Enhanced Competence-Based Learning. In H. H. Adelsberger, Kinshuk, J. M. Pawlowski, & D. G. Sampson (Eds.), *Handbook on Information Technologies for Education and Training* (pp. 155–177). Springer.

Shavelson, R. J. (2008). Guest editor's introduction. *Applied Measurement in Education*, *21*(4), 293–294. https://doi.org/10.1080/08957340802347613

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). American Council on Education.

Shepard, L. A., & Dwyer, C. A. (2008). Formative Assessment: Caveat Emptor. In *The Future of Assessment* (pp. 279–303). Routledge.

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

Sternberg, R. J., & Grigorenko, E. L. (2002). *The general factor of intelligence*. Erlbaum.

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*(5), 389–408.

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*.

Tomasik, M. J., Knecht, M., & Freund, A. M. (2017). Some evidence for the usefulness of an optimal foraging theory perspective on goal conflict and goal facilitation. *Journal of Personality and Social Psychology*, *113*(6), 962.

Uttal, D. H., & Perlmutter, M. (1989). Toward a broader conceptualization of development: The role of gains and losses across the life span. *Developmental Review*, *9*(2), 101–132.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

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.

van Geert, P. (2003). Dynamic systems approaches and modeling of developmental processes. *Handbook of developmental psychology*, *640672*.

van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Development Perspectives*, *5*(4), 273–278.

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*, 1–24. https://doi.org/10.1080/00273171.2018.1496813

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

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.

Waddington, C. H. (1957). *The strategy of the genes*. Routledge.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., & Gong, Q. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, *35*(2), 871–880. https://doi.org/10.1016/j.neuroimage.2006.12.017