Jindra et al. Large-scale Assessments in Education (2022) 10:22
https://doi.org/10.1186/s40536-022-00136-6

Large-scale Assessments in Education

RESEARCH

Open Access

Dynamics between reading and math proficiency over time in secondary education – observational evidence from continuous time models

Christoph Jindra[1*], Karoline A. Sachse[1] and Martin Hecht[2]

*Correspondence:
Christoph Jindra
christoph.jindra@hu-berlin.de
[1]Institute for Educational Quality Improvement, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[2]Department of Psychology, Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany

Abstract

Introduction Reading and math proficiency are assumed to be crucial for the development of other academic skills. Further, different studies found reading and math development to be related. We contribute to the literature by looking at the relationship between reading and math using continuous time models. In contrast to previous studies, this allows us to (a) report estimates for autoregressive and cross-lagged effects for a range of possible time intervals while still only estimating one set of continuous time parameters and (b) identify peak effects for the relationship between the two. Using data from Starting Cohort 3 of the National Educational Panel Study, we find, in line with previous evidence, a larger effect of reading on math than the other way around. Furthermore, we identify peak standardized cross-lagged effects ($a_{reading \to math} \approx 0.30$, $a_{math \to reading} \approx 0.13$) for a time interval of approximately 6 months.

Introduction

Reading and math proficiency are considered crucial for later development of other academic skills (e.g., Koponen et al., 2020). While the domains are often treated independently in practice, research frequently considers them alongside each other, recognizing that beyond their individual developmental trajectories there is interplay (e.g., Cameron et al., 2019; Koponen et al., 2020; Korpipää et al., 2017; Vanbinst et al., 2020; Bailey et al., 2020; Erbeli et al., 2021; Purpura et al., 2017). Learning about these developmental dynamics through appropriate models is crucial for improved prediction and identification of starting points for potential interventions.

The contribution of this study is twofold. First, we contribute to our knowledge on the relationship between math and reading by describing the interplay between the two over time using continuous time models. By doing so, we provide further evidence on potential reciprocal effects and the nature of such effects using a novel modeling technique. This allows us to answer questions such as; do we find further evidence that reading is the leading competency, whose effect on math is stronger than vice versa (e.g., Bailey et al., 2020; Erbeli et al., 2021) and for which time interval do we find peak effects? However, we can also assess each constructs' persistence, that is, we can describe for which construct past deviations persist into the future for longer. Previous findings, for example, suggest that reading might be somewhat more persistent than math (Hecht et al., 2001; Korpipää et al., 2017). Furthermore, our application also demonstrates the potential of continuous time modeling for the analysis of the interplay between educational constructs in general and how these models increase comparability and thus facilitate the accumulation of knowledge. We do this by showcasing one of the main strengths of continuous time models, the ability to describe the entire dynamics between two constructs with one set of continuous time parameters, which can be used to calculate model implied discrete autoregressive and cross-lagged effects not just for the observed time interval but for a reasonable range.

The remainder of the paper is structured as follows. We firstly provide an overview over the current literature regarding the relationship between reading and math proficiency. Then, we discuss model requirements based on past evidence and data limitations, arguing for continuous time models as a solution for some of the issues. This section is followed by an introduction to continuous time models. Subsequently, we describe the data and measures. We then present the empirical results and conclude with a discussion of these as well as the limitations of our study.

Current evidence on the relationship between math and reading

Previous studies have frequently shown that proficiencies in math and reading are related (e.g., Cameron et al., 2019, Bailey et al., 2020, Erbeli et al., 2021, Hübner et al., 2022, Gnambs & Lockl, 2022). There are various possible reasons why we might observe an association:[1] (a) Reading skills could influence math skills, (b) math skills could influence reading skills, (c) both could be true, implying complex dynamics between the constructs, (d) the two constructs could share a set of time-constant and/or time-varying causes. Although previous empirical evidence is often based on vastly different types of empirical models which often lead to parameters that are not strictly comparable (e.g,. Orth et al., 2021), results are usually interpreted as providing evidence for one of the four mechanisms.

Thus, (a) reading proficiency could be important for math because language shapes the development of numbers concepts and learning rules of the number system is similar to mastering written language (as symbolic representational system) (LeFevre et al., 2010). Also, well developed phonological processing and fluency skills are a prerequisite for math. In line with this, Grimm (2008) found early reading skills to be a good predictor for success in mathematics and Jordan et al. (2002) reported that reading abilities influence children's growth in mathematics. The inverse pathway, (b) math influencing reading proficiency, is also theoretically and empirically grounded. Early math proficiency, such as fluent counting, potentially shapes formation and retrieval skills of visual-verbal associations in long-term memory, which are crucial for reading fluency (Koponen et

[1] The list is not exhaustive and we focus on simple mechanisms related to a potential causal understanding of the relationship between the constructs.

al., 2013). In line with this, Duncan et al. (2007) reported that early math skills have the greatest predictive power for later learning and that early math skills predicted reading better than reading predicted math. Further, Holenstein et al. (2020) found a transfer effect of mathematical literacy achievement on different school domains including reading for adolescents. Purpura et al. (2017) suggested mathematical language skills as a mediating mechanism for the effect of mathematical skills on later reading.

However, given that there exist studies that provide support for either (a) or (b), it seems difficult to exclude (c) as an option, and thus the possibility that there are complex dynamics between the variables. Thus the literature in (c) seems in most cases rather complementary than competitive to those cited in (a) and (b), unless a specific study explicitly seems to find support for one pathway only such as the one by Jordan et al. (2002), who find that in their case reading abilities do impact growth in math but not the other way around. On the other hand, Schmitt et al., (2017) for example reported a bidirectional relationship for certain time spans between kindergarten and preschool. Similarly, Cameron et al. (2019) concluded that proficiency in each domain contributes to the other in a reciprocal, supportive manner. Bailey et al. (2020) reported bidirectional effects between reading and math but found stronger effects for reading on math when using a cross-lagged panel model with random intercepts (RI-CLPM) compared to the simple cross-lagged panel model (CLPM). These results are somewhat in line with Erbeli et al., (2021) who found bidirectional effects with reading appearing to be a leading and math a lagging indicator. Hübner et al., (2022) also found a complex relationship between the constructs but concluded that reading skills might be particularly important for the development of math skills. Gnambs and Lockl (2022) on the other hand also demonstrated how sensitive results are towards modeling choices. While results from the CLPM and CLPM with lag 2 showed consistent positive bi-directional associations between the constructs, effects for the RI-CLPM changed over grades even turning negative for older students. However, one has to keep in mind that the methodological debate on the difficulties of comparing results from the different models is ongoing (for example Orth et al., 2021; Lucas, 2022). Recently, Lucas drew attention to the well-known problem that results from the CLPM are a potentially uninterpretable mix of between- and within-effects, whereas the RI-CLPM aims to disentangle both. This is why Bailey et al. (2020) argue that the observed reduction in their effect sizes, when using the RI-CLPM in contrast to the CLPM, is due to the former accounting for the effects of stable unmeasured factors, a source of between-effects, which leads to (d), confounders. It seems indisputable that both domains are likely to share a set of common causes. These could be a broadly defined *g*-factor, that is, general intelligence, as in the strata-theories (Carroll, 1993; Cattell, 1987; Horn, 1988). But also more specific common domain-general cognitive correlates were reported. Several studies discuss a variety of factors that seem to be correlated with both domains, such as working memory, attentive behavior, processing speed, listening comprehension, nonverbal reasoning, serial retrieval fluency, phonological awareness, processing speed, and numeral recognition (Fuchs et al., 2013; Korpipää et al., 2017; Koponen et al., 2020; Vanbinst et al., 2020; Cirino et al., 2018). Furthermore, there are likely other factors that cause the constructs to be related. Previous evidence, for example, suggests that there is a genetic component to the correlation between reading and mathematics ability (Davis et al., 2014). Additionally, standard confounders such as family or language background will lead to an empirical association between the variables unless accounted for. Thus, previous studies provide support for all the above stated reasons that could be behind the empirical association. However, depending on the underlying data generating process, different statistical strategies seem more appropriate, as will be discussed in the following section.

Methodological challenges in the study of reading-math dynamics

It is well known that panel data and standard panel methods can offer some protection against the threat of time-constant unobserved heterogeneity (Allison, 2009; Bell & Jones, 2015; Halaby, 2004; Zyphur et al., 2020).[2] Thus, while we have to observe time-varying confounders to be able to control for them, in the absence of a valid identification strategy, we can at least mitigate the risk of time-constant confounding with panel data. On the other hand, if researchers are interested in the effect of reading on math but cannot exclude the possibility that math influences reading as well, one would need to go beyond standard models when simultaneously aiming to control for time-constant confounding (Allison et al., 2017; Arellano & Bond, 1991; Moral-Benito, 2013). However, it can be argued that if (c) is true, the statistical approach should account for this and any dynamics should be modeled explicitly. Thus, given that previous evidence implies that we cannot rule out reciprocal effects between math and reading skills and given that it seems highly unlikely that the constructs do not share common time-constant causes (d), a convincing statistical model should aim to account for both.

Different statistical approaches offer potential solutions, such as the random intercept cross-lagged panel model (RI-CPLM, Hamaker, 2015) or the general cross-lagged panel model (GCLM, Zyphur et al., 2020). Both incorporate dynamics in the form of autoregressive and cross-lagged parameters. Furthermore, both include correlated random intercepts to account for stable elements. While these elements are sometimes referred to as traits in the RI-CLPM, the correlated random intercepts approach in the GCLM corresponds more closely to what social scientists call a fixed effects approach to control for stable unobserved confounders (Allison, 2009; Halaby, 2004; Wooldridge, 2010; Zyphur et al., 2020; Bollen & Brand, 2011).[3] While we are not aware of an application that uses the GCLM to understand the dynamics between reading and math, Bailey et al., (2020) as well as Gnambs & Lockl (2022) have shown that results from the CLPM and the RI-CLPM differ substantially. The former attribute the reduction in effect sizes to controlling for time-constant traits. Thus, a high correlation between the random intercepts might indicate support for the idea of a substantial overlap between these time-constant factors. Besides the CLPM and the RI-CLPM, other longitudinal models to examine reciprocal relations exists, for example, the the stable trait autoregressive trait and state model (STARTS; Kenny & Zautra, 2001; also known as the trait-state-error (TSE) model, Kenny & Zautra, 1995), the latent curve model with structured residuals (LCM-SR; Curran et al., 2014), the autoregressive latent trajectory model (ALT; Bollen & Curran 2004; Curran & Bollen, 2001), and the latent change score model (LCS;

[2] We use the term time-constant unobserved heterogeneity for confounding factors that are constant over time and potentially unobserved, in contrast to time-varying confounding which can change over time. While standard models usually assume time-constant effects for these factors, research highlights that this assumption can be relaxed using specific modeling strategies (Bollen & Brand, 2011; Zyphur et al., 2020).

[3] There are substantive differences between the approaches. However, a full discussion is beyond the scope of the paper. One crucial difference is that the random-intercepts in the GCLM have indirect effects on future outcomes via autoregressive and cross-lagged paths, while the traits in the RI-CLPM only adjust the intercept of the observed variable at the time but do not allow for indirect effects (Usami, 2021).

Hamagami & McArdle 2001; McArdle & Hamagami, 2001). Usami et al.,(2019) provide a general framework to "facilitate the understanding of the strengths and weaknesses of these models" which "helps to clarify the conceptual and statistical differences" (p. 654). For longitudinal educational research, especially for large-scale assessment studies like PISA, Lohmann et al., (2022) discuss the advantages of the LCM-SR, one of which being that "systematic (linear) trends are disentangled from the autoregressive and cross-lagged parameters" (p. 8). In this vein, Curran and Bollen (2001) speak of the "best of both worlds" as linear growth and dynamics are combined into one model while retaining their typical interpretation. All these models share the property of being *discrete-time* models which is associated with some issues that we discuss next.

In practice, we observe a considerable amount of variation in the time intervals between measurements across studies. Codding et al. (2015), for example, analyze data from three measurement points distributed over a short time period, namely fall, winter, and spring of one academic year. Rinne et al. (2020) collect data for six measurement points distributed over three years, while Bailey et al. (2020) use data from four measurement points collected over four years. These variations represent a challenge for the interpretation and comparability of coefficients across studies, complicating the accumulation of knowledge as effects from standard discrete time models pertain to the specific time interval of the respective study only (Oud & Delsing, 2010). Furthermore, time intervals between measurements do not only vary between studies but frequently within studies as well. In large studies, such as the National Educational Panel Study (NEPS; Blossfeld et al., 2019), or the Early Childhood Longitudinal Study (ECLS; e.g., Tourangeau et al., 2018) this can already be due to the fact that field work is often taking place over a considerable amount of time (for examples in other fields see Steptoe et al., 2013; Sonnega et al., 2014). For the purpose of analyses, information on the exact timing of the measurements is sometimes discarded and measurements are subsumed under the year of the fieldwork for a specific wave. Similarly, studies often do not measure each construct across all waves, complicating the analysis as constructs are missing for some waves. While dynamic discrete time models struggle with these issues, continuous time models can handle different time intervals between and within studies (Oud & Delsing, 2010; Voelkle et al., 2012, 2018; Hecht et al., 2019; Hecht & Zitzmann, 2020) and can help to explore the unfolding of effects over time (Hecht & Zitzmann, 2021). Given that these models are further able to model complex dynamics between constructs and are potentially able to account for time-constant unobserved heterogeneity, they are well suited for research questions such as the interplay between reading and math proficiencies over time.

The advantages of continuous-time models over discrete-time models have been excellently described and illustrated in other works (e.g., Voelkle et al., 2012; Ryan et al., 2018; Hecht et al., 2019; Hecht & Zitzmann, 2021; Lohmann et al., 2022; Hecht et al., 2022). To shortly reiterate: Continuous-time modeling conceptualizes longitudinal data as "snapshots" that inform the estimation of continuously evolving processes. Thus, data from all time points can be used for estimation. This lifts the usual spacing restrictions in discrete-time models and allows the usage of flexible longitudinal designs with intra- and interindividual varying spacing between measurement occasions. Once the continuous-time model is estimated, it can be used to calculate corresponding discrete-time model parameters for any desired time interval length. This helps to compare discrete-time model results from studies with differently spaced measurement occasions. For instance, if one programme uses 1-year intervals and another programme 3-year intervals, both programmes could estimate a continuous-time model, then calculate discrete-time parameters for the same time interval (e.g., 2 years or any other desired interval length) and hence arrive at comparable results. Another advantage of continuous-time models which is frequently discussed in the literature (and particularly highlighted by Hecht & Zitzmann, 2021) is that the unfolding and dissipation of dynamic effects can be explored. This is usually done by plotting the dynamic parameter of interest (y axis) against the discrete-time interval length (x axis). We use this technique later in the present work to identify for which time interval length the reciprocal relationship of students' reading and math proficiencies is maximal (which could be termed "peak cross-lagged effects", Hecht & Zitzmann, 2021). This exploration with the help of continuous-time modeling is interesting and relevant because researchers often (implicitly or explicitly) search for these effects, but might not be able to identify these effects when applying discrete-time models.

Continuous time modeling

We develop the main idea behind continuous time modeling following Voelkle et al. (2012), starting with a simple discrete time multivariate autoregressive model. We then move from an intuitive approach of dealing with variations in time intervals between studies in discrete time models to the continuous treatment of time. The starting point for our discussion is an autoregressive model of the following form:

$$x_j(t) = A(\Delta t) \times x_j(t - \Delta t) + w_j(\Delta t) \tag{1}$$

All variables are assumed to be in deviation form, which allows us to ignore intercepts for now. In Eq. 1, $x_j(t)$ and $x_j(t - \Delta t)$ each represent a $K \times 1$ vector of the same K variables for individual j, once observed at discrete time point t and previously at time $t - \Delta t$. Given that they appear on both sides of the equation, each variable can be an outcome and explanatory variable at the same time. The variables are linked to each other over time via the $K \times K$ matrix $A(\Delta t)$, containing the autoregressive parameters on its diagonal and cross-lagged effects on the off-diagonals. $w_j(\Delta t)$ is a $K \times 1$ vector of stochastic error terms, which are assumed to be uncorrelated over time. Both, $A(\Delta t)$ and $w_j(\Delta t)$, are a function of the time between measurements, indicated by Δt. However, while they are a function of time, the underlying data generating process is nonetheless assumed to be constant over time. Hence, the equation just highlights that, while actual processes most often evolve continuously over time, our measurements are usually observed in discrete time intervals at specific time points and therefore, results based on discrete time methods will result in parameters that are a function of the study specific time intervals (Δt). Consequently, they are not directly comparable.

An *intuitive approach* to make parameters from studies with varying time intervals comparable would be to predict normalized changes, $(x_j(t) - x_j(t - \Delta t))/\Delta t$, instead of levels. Thus, ignoring error terms, we can relate the normalized changes to previous levels as follows:

$$\frac{\Delta x_j(t)}{\Delta t} = A^* \times x_j(t - \Delta t) \tag{2}$$

Voelkle et al. (2012) call A^{*} a crude approximation of the underlying continuous time process. It is important to note that A^{*} is independent of the study specific time interval. However, the parameters are related and can be transformed into each other via the following equation:

$$A\left(\Delta t\right) = A^{*} \times \Delta t + I \tag{3}$$

Thus, once we have either A^{*} or $A\left(\Delta t\right)$ and the time interval, we can easily move between the two. While simple and intuitive, Voelkle et al. (2012) highlight two main shortcomings of this approach. Firstly, while it represents an approximation of the underlying continuous process, it is just a crude approximation of the so-called drift matrix that represents the data generating process in continuous time modeling. Secondly, the approach can only be used if the time intervals between measurements are constant over time. Thus, while intuitive, the structure of some data prevents us from using this simple approach, which brings us to the exact approach. The intuitive approach still relies on a discrete understanding of time as we predict change between discrete measurement points. In continuous time models, change is still the dependent variable. However, given that time is treated as continuous variable, instead of predicting the normalized change between discrete measurement points, we now describe the relationship between $x_j\left(t\right)$ and the first derivative of $x_j\left(t\right)$ with respect to time:

$$\frac{dx_j\left(t\right)}{dt} = A \times x_j\left(t\right) \tag{4}$$

Thus, the *change* in $x_j\left(t\right)$ over an infinitesimally small time interval, the first derivative, is a function of $x_j\left(t\right)$ itself and all the relevant parameters describing this relationship are contained in drift matrix A. The solution of the differential equation from above is given by

$$x_j\left(t\right) = e^{A(t-t_0)}x_j\left(t_0\right), \tag{5}$$

where $x_j\left(t_0\right)$ represents the value of the variables in the model at the initial time t_0. Taking the first derivative of Eq. 5 with respect to time (t) yields Eq. 4. In order to distinguish parameters in A from those in $A\left(\Delta t\right)$, their naming conventions differ. Those on the diagonal are called *auto*-effects, while those off the diagonal are called *cross*-effects (in contrast to *autoregressive* and *cross-lagged* effects in the discrete time case; see Table 1 in the work of Hecht & Voelkle, 2021, for an overview of discrete-time and continuous-time terms). Like in the case of the intuitive approach, parameters in $\boldsymbol{A}$ can be transformed into discrete time parameters. Voelkle et al. (2012) show that the nonlinear relationship is given by the matrix exponential

$$A\left(\Delta t\right) = e^{A \times \Delta t} \tag{6}$$

Thus, as in the intuitive approach, we can move between discrete and continuous time parameters. However, unlike in the intuitive approach, the relationship in Eq. 6 represents the exact relationship between continuous and discrete time parameters, not just an approximation, and can be used to adequately describe the relationship between the variables over time.

Until now, we have ignored error terms, intercepts and time-constant unobserved heterogeneity in our discussion.[4] Including those leads to the following model

$$\frac{dx_j(t)}{dt} = A \times x_j(t) + \xi_j + G\frac{dW_j(t)}{dt} \quad (7)$$

Briefly, $W_j(t)$ represents the continuous time error process, a so-called Wiener process or Brownian motion. $dW_j(t)$ is the stochastic error term, an infinitesimally small increment of the Wiener process, while **G** represents the effect of the stochastic error term on the change in $x_j(t)$. Together, these terms can be used to describe the noise in the continuous time process.

The random variables ξ_j are assumed to follow a normal distribution with means μ_ξ and variance-covariance matrix Ψ. Thus, the vector μ_ξ denotes the continuous time-intercepts and variance and covariances across subjects are contained in the matrix Ψ. Together they account for nonzero mean trajectories. By modeling random intercepts for both domains, we allow individuals to deviate from the overall intercepts. More importantly, permitting a covariance between the random intercepts is argued to account for unobserved time-constant heterogeneity (compare Zyphur et al., 2020).

So far, our model is suitable for describing stable equilibrium processes, that is, processes that converge to final means over time. In educational research, such a model might be implausible, because development often occurs over sustained periods. To incorporate such trends, Lohmann et al. (2022, in this same issue) developed the continuous-time latent change score model with structured residuals (CT-LCM-SR). The basic idea is to add an additional continuous-time process (for each variable) to the model that is specified in such a way that linear trends are captured. Thus, this model combines central components from both dynamic and descriptive models, the dynamics of the variables and the linear growth in the variables. As the focus of the present work is on the dynamics, we use ideas of the CT-LCM-SR to control for linear trends in the data so that the targeted dynamics can be properly estimated. An alternative option would be to detrend the data beforehand, but this might come with other disadvantages (e.g., the common problem of 2-step analyses of how to transfer the uncertainty of parameter estimates from the step-1 to the step-2 model). With respect to the interpretation of discrete-time dynamic effects, with controlling for the linear trends in the LCM-SR (or in its continuous-time version, the CT-LCM-SR) we fabricate an equivalent interpretation as in corresponding dynamic models that assume no trends, except that the reference line around which the state values fluctuate is the linear trend and not a flat line (which is, however, actually also just a linear trend with a zero slope).

Method

Sample

We use data from the German NEPS study (Blossfeld & Roßbach, 2019). The study has a multicohort sequence design, consisting of several representative cohorts either defined by age or by specific points in the educational system. We analyse data from Starting Cohort 3 (SC3), a representative sample of students attending grade 5 in school

[4] Given that we do not include additional time-varying or time-constant predictors in our model, we do not further discuss these here. However, both types of predictors can be incorporated (see for example Oud and Delsing 2010; Driver et al. 2017).

Table 1 Competence measures in NEPS SC3 by wave

Wave	Field work	Reading	Math
1	Autumn 2010 -Spring 2011	✓	✓
2	Autumn 2011 - Spring 2012	-	-
3	Autumn 2012 - Spring 2013	✓	✓
4	Autumn 2013 - Spring 2014	-	-
5	Autumn/Winter 2014	✓	
6	Spring 2015	-	✓
7	Autumn 2015 - Spring 2016	-	-
8	Autumn/Winter 2016	-	-
9	Autumn/Winter 2017	✓	✓

year 2010/2011 who were subsequently followed over time. Students were selected using a stratified multistage sampling design (Skopek et al., 2012). Schools were sampled in the first stage, stratified by school type. Subsequently two classes in each school were selected in the second stage and all students in those classes were eligible for interviews. We use the five waves for which competency scores in math and reading are available (see Table 1). In waves 1 (grade 5), 3 (grade 7), and 9 (grade 12), students took tests in both math and reading. In wave 5 (grade 9), students took math tests only, while reading tests were administered around five months later, in wave 6 (grade 9). We include all students with at least one non-missing value on the competency scores in our analysis, leading to an effective sample size of 7,639 students.[5] Mean age at wave 1 wave is 10.5 (SD=0.64, min.=8, max.=15), while mean age at wave 9 is 17.4 (SD=0.59, min.=16, max.=19). The share of female students at wave 1 is 0.48.

As described above, time intervals vary between waves and furthermore within each wave, testing took place over a period of several months, resulting in unequal time intervals between measurement points (see Table 2). We use the month with the first observed competence score as our baseline (t=0) and express subsequently elapsed time between measurements in years and months from baseline.

Competence scores

Reading (Gehrer et al., 2013) and mathematics proficiency (Neumann et al., 2013) were measured by the NEPS competence tests, which are constructed to adequately measure the respective construct in all age cohorts assessed. Students were tested by paper-based competence tests. The tests were scaled and linked based on IRT models. Individual scores (WLEs) based on the linked tests are available for different waves as described above and were placed on a common scale to facilitate meaningful mean-level comparisons across time. The first wave serves as a reference scale and values can thus be interpreted as developmental trajectories across measurement points (see Kutscher et al., 2020, and Petersen et al., 2020, for a detailed description). WLE reliabilities range between 0.721 and 0.812 across time and between domains.

Data analysis

The continuous time model is estimated via Structural Equation Modeling (Bollen, 1989) in R (R Core Team, 2021) by imposing restrictions on the relevant parameters (Oud & Delsing, 2010; Voelkle et al., 2012). Specifically, the R package ctsemOMX (Driver

[5] As recommended by Skopek et al. (2012), we exclude the subsample of students with special education needs in our analyses.

Table 2 Fieldwork for competence tests in NEPS SC3 by wave

Wave	N	Timing	N	Timing	N	Timing	N	Timing	N	Timing
1	5	2010-10								
1	2177	2010-11								
1	2889	2010-12								
1	130	2011-1								
3			3223	2012-11						
3			2229	2012-12						
3			21	2012-6						
3			721	2013-1						
5					1583	2014-11				
5					3207	2014-12				
5					98	2015-1				
6							1237	2015-4		
6							2358	2015-5		
6							818	2015-6		
6							3	2015-7		
9									401	2017-11
9									4248	2017-12
9									75	2018-1

Note. Timing refers to the year and month of measurement. In case information on the timing of measurement was not available, we used the mode of the respective wave.

et al., 2021), which is based on OpenMX (Boker et al., 2021), was used for the analyses. In order to account for growth over time, we added a linear trend component to the continuous time model. In our case, growth intercepts and slopes are assumed to be "fixed", that is, they do not vary over persons and thus have no random component. Random intercepts are however included in the models as trait variables to control for time-constant unobserved heterogeneity (Driver et al., 2017). The model specification can be found as a path diagram in Figure S1 in the Supplementary Materials. The model is estimated using Full Information Maximum Likelihood (FIML), thus missing values are accounted for. Descriptive statistics were computed using Stata 16.1 SE (StataCorp, 2019). Growth curve models for validating the continuous time model trend component were estimated in Stata.

Results

Descriptive statistics can be found in Table 3. We report results for each domain by wave alongside the correlations between the domains across waves. Additionally, we provide summary statistics on the time between measurement points. All summary statistics are calculated using the maximum number of cases available for the specific statistic. Descriptive means show an increase of average proficiency in both domains over time in our sample. Correlations show strong associations between proficiency scores in the same domain across waves, though decreasing with increasing time intervals between measurement occasions. However, compared to reading, descriptive results show stronger correlations across waves for math. Furthermore, we also find strong correlations across domains with the similar pattern of decreasing associations for increasing time intervals.

Parameters from the continuous time model are displayed in Table 4. The parameters accounting for a linear trend (tr_{math} and $tr_{reading}$) can be interpreted in a similar fashion to those in standard growth curve models. Both indicate, as expected, growth in each of

Table 3 Descriptive statistics

	M	*SD*	min	max	Math Wave 1	Math Wave 3	Math Wave 6	Math Wave 9	Reading Wave 1	Reading Wave 3	Reading Wave 5	Reading Wave 9
Math Wave 1	-0,01	1,17	-4,68	4,00	1							
Math Wave 3	0,82	1,23	-4,65	4,81	0,74	1						
Math Wave 6	1,60	1,19	-2,53	6,88	0,71	0,74	1					
Math Wave 9	2,60	1,11	-1,69	6,81	0,62	0,65	0,69	1				
Reading Wave 1	-0,02	1,27	-4,71	3,96	0,64	0,58	0,57	0,45	1			
Reading Wave 3	0,79	1,36	-4,34	5,78	0,55	0,60	0,57	0,44	0,61	1		
Reading Wave 5	1,33	1,12	-2,12	6,33	0,53	0,52	0,58	0,45	0,59	0,63	1	
Reading Wave 9	2,10	1,01	-1,62	7,47	0,48	0,47	0,49	0,45	0,53	0,57	0,52	1
Time between waves 1 and 3	23,91	0,86	18	26								
Time between waves 3 and 5	24,12	0,85	22	30								
Time between waves 5 and 6	5,21	0,80	3	7								
Time between waves 6 and 9	30,85	0,87	29	33								

Notes: Statistics are calculated using listwise deletion.

the domains over time. A comparison of our estimates to those from growth curve models with random intercepts shows that estimates from the continuous time model are nearly identical to those from the simpler models (see Section S2 in the Supplementary Materials).

The main parameters of interest, describing the dynamics, the auto- and cross-effects, are all statistically significant at the 5% level against the null hypothesis that the respective parameter is equal to 0. Due to the complex nature of these parameters, we provide discrete-time equivalents for varying time intervals for interpretation in Fig. 1. One way of interpreting Fig. 1 is that the plots show the autoregressive and cross-lagged effects we would expect based on our model for a range of plausible time intervals in hypothetical studies. For example, in case of the discrete-time cross-lagged parameters, a hypothetical study with a time-interval between measurements that approaches zero would provide us with estimates close to zero. This is a standard result as the effects would still have to develop.

Discrete-time autoregressive parameters for time intervals between zero and four years are shown in the upper panel of Fig. 1. For our model, we find a strong decline in the size of the autoregressive parameters for reading (dashed line). In contrast, the decline in the effect size for the discrete-time autoregressive parameters for math is less rapid and it takes longer time intervals for the coefficients to go to zero (solid line). Thus, results for our data suggest that any deviations from mean math proficiency levels in earlier waves transmit to more distant later waves. Discrete-time versions of the cross-effects (i.e., cross-lagged effects), are displayed in the bottom panel of Fig. 1. Effects of math on reading are displayed as a dashed line, while those for reading on math are displayed as a dotted line. The general shape of the relationship between time interval length and the discrete versions of the parameters is largely comparable, with effects peaking at a time interval of around six months ($a_{reading \rightarrow math} \approx 0.30$, $a_{math \rightarrow reading} \approx 0.13$), followed by a subsequent strong decline in effect sizes. Thus, our model suggests that studies with time-intervals of around six months between measurements would find the strongest effects for the cross-lagged parameters. For our data, the effect of reading on math is larger than the effect of math on reading for nearly all time intervals before finally converging and dissipating.

Discussion

In the present article we investigated the dynamics between reading and mathematics achievement in 10- to 19-year-olds. Reading and math proficiency are important factors for successful participation in society and understanding their interplay allows prediction and identification of starting points for prevention and intervention. In order to understand the interplay of reading and mathematics development, we estimated a continuous time model including linear trend components. In this approach, central components from both, dynamic and descriptive models, are combined: the dynamics of the variables and the linear growth in the variables. The model provided us with a set of continuous-time coefficients that describe the underlying processes' persistence and their cross-effects, controlled for linear trends and time-constant unobserved confounders.

Applying the model to the NEPS SC3 data revealed that the interplay between math and reading followed patterns that were similar to recently reported findings: We found evidence for both paths, effects of reading on math and effects of math on reading. The

Table 4 – Continuous time model parameter estimates for math and reading

Parameter		Estimate	SE
Auto effect	$a_{reading}$	-4.54	1.405
	a_{math}	-1.47	0.230
Cross effect	$a_{reading \longrightarrow math}$	1.84	0.680
	$a_{math \longrightarrow reading}$	1.14	0.413
Diffusion variance	$\sigma^2_{reading}$	5.49	1.688
	σ^2_{math}	0.91	0.093
Diffusion covariance	$\sigma^2_{reading \leftrightarrow math}$	−0.99	0.428
Interceptst_0	$b_{reading}$	0.27	0.014
	b_{math}	0.28	0.014
Linear trend	$tr_{reading}$	0.25	0.002
	tr_{math}	0.33	0.002
Variance traits	$\psi_{reading}$	0.98	0.021
	ψ_{math}	1.06	0.022
Covariance traits	$\psi_{reading \leftrightarrow math}$	0.86	0.019

Notes: N = 7639.

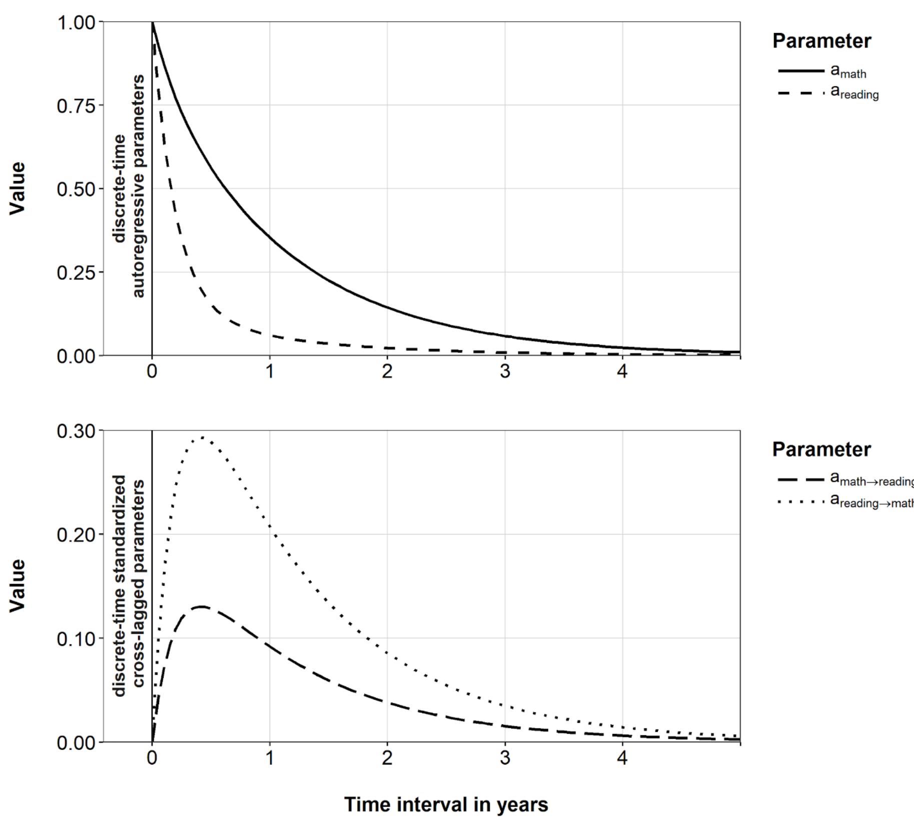

Fig. 1 - Discrete-time autoregressive and cross-lagged parameters for varying time intervals

former appears to be somewhat stronger than the latter, which is in line with Erbeli et al. (2021), results from the RI-CLPM in the study by Bailey et al. (2020) as well as the conclusions in the study by Hübner et al. (2022). Further, math appeared to be the more persistent construct. However, our findings also extend previous research in several respects. First, our continuous time modeling strategy allows us to discretize our auto- and cross-effects for different time intervals and identify for which time intervals we expect peak effects between the domains. For example, Bailey et al. (2020) and Erbeli et al. (2021) reported effects for time intervals of one year, whereas Rinne et al. (2020) examined time intervals of approximately six months and Korpipää et al. (2017) of six years. Discretizing the continuous time parameters to these intervals (see Fig. 1) suggests that autoregressive effects of reading decrease rapidly, leveling off at an interval above six months. This does not hold for math, where, based on our results, we would expect to observe substantial effects of past changes even after two years. Thus, our results suggest that any deviations from mean reading levels (which may, e.g., be due to an intervention aimed at improving reading skills) might dissipate rather quickly. Similarly, for the cross-effects we estimated the maximum mutual influence of both constructs on each other at a time interval around six months. For all time intervals, we found the effect of reading on math to be stronger than the effect of math on reading. Thus, our results also suggest that interventions to improve reading skills might have larger positive spillover effects on math skills than the other way around. We would like to highlight some potential practical implications of our results. Firstly, any changes in one of the variables due to an exogenous one-off intervention targeted at one of the domains would, based on our model, not lead to long lasting effects as can be seen by the decreasing effects sizes for larger time-intervals. Secondly, any intervention targeted on reading would seem to have larger, potentially unintended, effects on math than the other way around. But, given that the effects disappear in our model, another implication would be that interventions might have to be sustained for long term effects. However, our approach does not have a valid identification strategy and we cannot claim to have identified any potential causal effects. Thus, until further research has confirmed our results, any implications need to be viewed with caution.

Second, we can compare our results to findings stemming from different analyses of different age groups. Very many studies examining the interplay of reading and math focus on children in primary school or preschoolers (e.g., Cameron et al., 2019; Koponen et al., 2020; Bailey et al., 2020; Erbeli et al., 2021; Purpura et al., 2017; Vukovic et al., 2013), whereas only a few studies examine development up to the end of lower secondary age, like Chen and Chalhoub-Deville (2016), Codding et al. (2015) and Grimm (2008) examining up to 14-year-olds or Korpipää et al. (2017) examining up to 15-year-olds. But none investigated the full secondary track age up to 18-year-olds as was the case in our sample.

There are several limitations to our study. NEPS has a complex survey design including unequal selection probabilities, clustering, stratification and a refreshment sample. Additionally, like in other panel studies, we observe non-response as well as attrition. While NEPS provides weights as a way to potentially account for some of these issues, the use of sampling weights is currently not implemented in the ctsemOMX package and hence we report results for the unweighted sample. Consequently, we cannot exclude bias in our estimates due to effect heterogeneity or endogenous sampling (Solon et al., 2015). On the other hand, given that the models are estimated using FIML, non-response as well as drop out should be accounted for. However, it is currently not possible to include auxiliary variables to improve the performance of FIML (Collins et al., 2001) for this type of model. Hence, the benefits of this approach might be somewhat limited in our case. Furthermore, not considering the complex survey design, including the clustering of students in schools, also implies that we are likely to underestimate the standard errors (Heeringa et al., 2010). However, accounting for complex survey designs in continuous time modeling is an ongoing area of research.

One of the major appeals of models such as the RI-CLPM, the GCLM and the continuous time model for applied research is the potential to control for time-constant confounding by including correlated random intercepts. There are, however, substantive differences between the approach in the RI-CLPM and the one implemented in ctsemOMX, the latter being more comparable to the GCLM. Given that recent research discusses scenarios under which the RI-CLPM might not perform as expected (Lüdtke & Robitzsch, 2021), further research might be necessary to fully understand the conditions under which the correlated random intercepts, as implemented in ctsemOMX, control for time-constant confounding. Crucially, we do not include any potential time-varying confounders, such as changes in classroom composition. Thus, independent of the methodological debate surrounding the effectiveness of the RI-CLPM approach to account for time-constant confounding, we cannot claim to have identified causal effects and the remaining effects might disappear when adding further controls. Thus, although we find significant cross-effects between the constructs which may indicate potential for interventions, our results alone are not sufficient to draw any strong recommendations for policies aimed at improving proficiencies. Similarly, we use a linear trend to account for changes over time. However, if the underlying change over time is non-linear our model would not account properly for this, which might affect the estimation of the parameters in our model. Future research using data with more measurement points should ideally explore if another model for the trend is more appropriate.

With the presented continuous-time model we were able to calculate discrete-time dynamic model parameters (autoregressive and cross-lagged effects) for any arbitrary time interval (and hence explore the dependency of these parameters on the length of the time interval, see Fig. 1), although the assessments were conducted with specific time interval lengths (of around one year, see Table 2). However, Voelkle et al.-(2012) warn that one has to be careful when inter- or extrapolating to discrete time points that have not been observed empirically. In a similar fashion, Hecht and Zitzmann (2021) caution that peak cross-lagged effects might be located in regions with no or sparse data and that the quality of inter- or extrapolations into such regions might depend on design characteristics. These issues can be seen as providing some support for the call for potentially different survey designs that, instead of having repeated measures at equidistant intervals, maximise the potential to accurately estimate the effects for different time intervals by varying the length between measurements. Thus, further research should confirm our results by applying continuous time models to survey data with different time intervals than the ones observed in NEPS.

One of the assumptions we have to make is that the nature of the process itself does not change over time. There is some recent evidence that this might not hold for our sample (Gnambs & Lockl, 2022). Hence, future research should look into the potential consequences for our model and ways of relaxing the assumption.

Continuous-time modeling provides a natural way of integrating longitudinal data with differently spaced measurement occasions. Alternatives might be the Mplus' TSCORE option, which can, however, only be used to define growth models (Muthén & Muthén, 2017, p. 614) or the "phantom variable approach". Oud and Voelkle (2014) and Voelkle and Oud (2015) discuss this approach's potential to account for unequal time intervals with the conclusion that it is rather limited. Instead, they argue for using continuous-time modeling.

In sum, this paper is the first to analyze the interplay of math and reading proficiency in a large representative sample of German students using continuous time models. We find evidence that math is the more persistent domain and further evidence math and reading are positively coupled over time with reading having a stronger effect on math than math on reading. The effects are most pronounced for a time interval of approximately six months.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40536-022-00136-6.

Supplementary Material 1

Acknowledgements
Dynamics between reading and math proficiency over time in secondary education – Observational evidence from continuous time models.

Author contributions
Conceptualization, C.J, and K.S.; data acquisition, C.J., methodology, C.J., K.S. and M.H.; formal analysis, K.S. and C.J.; data curation, K.S. & C.J.; writing—original draft preparation, C.J. and K.S.; writing—review and editing, C.J., K.S. and M.H.; visualization, M.H. All authors have read and agreed to the published version of the manuscript.

Funding
None.

Data Availability
This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 3–5th Grade, doi:https://doi.org/10.5157/NEPS:SC3:1.0.0. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. The anonymized data are available for the scientific community at https://www.neps-data.de.

Declarations

Competing interests
None.

Received: 28 February 2022 / Accepted: 6 October 2022

References

Allison, P. D. (2009). Fixed effects regression models. *Quantitative Applications in the Social Sciences* (160 vol.). SAGE

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. Socius: Sociological Research for a Dynamic World, 3, 1–17. https://doi.org/10.1177/2378023117710578

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, *58*(2), 277. https://doi.org/10.2307/2297968

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

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, *3*(01), 133–153. https://doi.org/10.1017/psrm.2014.7

Blossfeld, H. P., & Roßbach, H. G. (Eds.). (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS). Edition ZfE* (2nd ed.). Springer VS

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., Mehta, P., von Oertzen, T., Gore, R. J., Hunter, M. D., Hackett, D. C., Karch, J., Brandmaier, A. M., Pritikin, J. N., Zahery, M., Kirkpatrick, R. M., Wang, Y., & Niesen, J. (2021). *OpenMx: Extended Structural Equation Modelling* (2.19.8) [Computer software]. https://CRAN.R-project.org/package=OpenMx

Bollen, K. A. (1989). *Structural equations with latent variables. A Wiley-interscience publication*. New York: Wiley

Bollen, K. A., & Brand, J. E. (2011). A General Panel Model with Random and Fixed Effects: A Structural Equations Approach. *Social Forces*, *89*(1), 1–34. https://doi.org/10.1353/sof.2010.0072

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research*, *32*, 336–383. https://doi.org/10.1177/0049124103260222

Cameron, C. E., Kim, H., Duncan, R. J., Becker, D. R., & McClelland, M. M. (2019). Bidirectional and co-developing associations of cognitive, mathematics, and literacy skills during kindergarten. *Journal of Applied Developmental Psychology*, *62*, 135–144. https://doi.org/10.1016/j.appdev.2019.02.004

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press

Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. Elsevier

Chen, F., & Chalhoub-Deville, M. (2016). Differential and long-term language impact on math. *Language Testing*, *33*(4), 577–605. https://doi.org/10.1177/0265532215594641

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

Codding, R. S., Petscher, Y., & Truckenmiller, A. (2015). CBM reading, mathematics, and written expression at the secondary level: Examining latent composite relations among indices and unique predictions with a state achievement test. *Journal of Educational Psychology*, *107*(2), 437–450. https://doi.org/10.1037/a0037520

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. https://doi.org/10.1037/1082-989x.6.4.330

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107–135). American Psychological Association. https://doi.org/10.1037/10409-004

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*, 879–894. https://doi.org/10.1037/a0035297

Davis, O., Band, G., Pirinen, M., et al. (2014). The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nature Communications*, *5*, 4204. https://doi.org/10.1038/ncomms5204

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), 1–35. https://doi.org/10.18637/jss.v077.i05

Driver, C., Voelkle, M., & Oud, H. (2021). *ctsemOMX: Continuous Time SEM - „OpenMx" Based Functions* (1.0.4) [Computer software]. https://CRAN.R-project.org/package=ctsemOMX

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Erbeli, F., Shi, Q., Campbell, A. R., Hart, S. A., & Woltering, S. (2021). Developmental dynamics between reading and math in elementary school. *Developmental Science*, *24*(1), e13004. https://doi.org/10.1111/desc.13004

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

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, *5*(2), 50–79

Gnambs, T., & Lockl, K. (2022). Bidirectional effects between reading and mathematics development across secondary school. *Zeitschrift für Erziehungswissenschaft*. https://doi.org/10.1007

Grimm, K. J. (2008). Longitudinal Associations Between Reading and Mathematics Achievement. *Developmental Neuropsychology*, *33*(3), 410–426. https://doi.org/10.1080/87565640801982486

Halaby, C. N. (2004). Panel Models in Sociological Research: Theory into Practice. *Annual Review of Sociology*, *30*(1), 507–544. https://doi.org/10.1146/annurev.soc.30.012703.110629

Hamagami, F., & McArdle, J. J. (2001). Advanced studies of individual differences: Linear dynamic models for longitudinal data analysis. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 203–246). Psychology Press. https://doi.org/10.4324/9781410601858

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hecht, M., Hardt, K., Driver, C. C., & Voelkle, M. C. (2019). Bayesian continuous-time Rasch models. *Psychological Methods*, *24*, 516–537. https://doi.org/10.1037/met0000205

Hecht, M., Horstmann, K. T., Arnold, M., Sherman, R. A., & Voelkle, M. (2022). Modeling dynamic personality theories in a continuous-time framework: An illustration [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/q97pz

Hecht, M., & Voelkle, M. C. (2021). Continuous-time modeling in prevention research: An illustration. *International Journal of Behavioral Development*, *45*(1), 19–27. https://doi.org/10.1177/0165025419885026

Hecht, M., & Zitzmann, S. (2020). A computationally more efficient Bayesian approach for estimating continuous-time models. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 829–840. https://doi.org/10.1080/10705511.2020.1719107

Hecht, M., & Zitzmann, S. (2021). Exploring the unfolding of dynamic effects with continuous-time models: Recommendations concerning statistical power to detect peak cross-lagged effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–9. https://doi.org/10.1080/10705511.2021.1914627

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*(2), 192–227. https://doi.org/10.1006/jecp.2000.2586

Heeringa, S., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis. *Chapman & Hall / CRC statistics in the social and behavioral sciences series*. Taylor & Francis. http://www.gbv.eblib.com/patron/FullRecord.aspx?p=555702

Holenstein, M., Bruckmaier, G., & Grob, A. (2020). Transfer effects of mathematical literacy: an integrative longitudinal study. *European Journal of Psychology of Education*, 1–27. https://doi.org/10.1007/s10212-020-00491-4

Horn, J. (1988). Thinking about Human Abilities. In J. R. Nesselroade & R. B. Cattell (Hrsg.), Handbook of Multivariate Experimental Psychology (pp. 645–685). Springer US. https://doi.org/10.1007/978-1-4613-0893-5_19

Hübner, N., Merrell, C., Cramman, H., Little, J., Bolden, D., & Nagengast, B. (2022). Reading to learn? The co-development of mathematics and reading during primary school. *Child Development*, *00*, 1–17. https://doi.org/10.1111/cdev.13817

Jordan, N. C., Kaplan, D., & Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: findings of a two-year longitudinal study. *Journal of Educational Psychology*, *94*(3), 586–597. https://doi.org/10.1037//0022-0663.94.3.586

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52–59. https://doi.org/10.1037/0022-160
Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243–263). American Psychological Association
Koponen, T., Eklund, K., Heikkilä, R., Salminen, J., Fuchs, L., Fuchs, D., & Aro, M. (2020). Cognitive Correlates of the Covariance in Reading and Arithmetic Fluency: Importance of Serial Retrieval Fluency. *Child Development*, *91*(4), 1063–1080. https://doi.org/10.1111/cdev.13287
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
Kutscher, T., & Scharl, A. (2020). NEPS Technical Report for Reading: Scaling Results of Starting Cohort 3 for Grade 12. NEPS Survey Papers, volume 67. Leibniz Institute for Educational Trajectories, National Educational Panel Study, Bamberg, Germany
LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to Mathematics: Longitudinal Predictors of Performance. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x
Lohmann, J. F., Zitzmann, S., Voelkle, M. C., & Hecht, M. (2022). A primer on continuous-time modeling in educational research: An exemplary application of a continuous-time latent curve model with structured residuals (CT-LCM-SR) to PISA data. *Large-Scale Assessments in Education*, *10*, 1–32. https://doi.org/10.1186/s40536-022-00126-8
Lüdtke, O., & Robitzsch, A. (2021). A critique of the random intercept cross-lagged panel model. https://doi.org/10.31234/osf.io/6f85c
Lucas, R. E. (2022, February 14). It's Time To Abandon The Cross-Lagged Panel Model. https://doi.org/10.31234/osf.io/pkec7
McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–175). American Psychological Association
Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business and Economic Statistics*, *31*(4), 451–472. https://doi.org/10.1080/07350015.2013.818003
Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide: 8th Edition (Version 8, April 2017)*. Los Angeles, CA: Muthen & Muthen. https://www.statmodel.com/download/MplusUserGuideVer_8.pdf
Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., & Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online*, *5*(2), 80–109
Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*, 1013–1034. https://doi.org/10.1037/pspp0000358
Oud, J. H. L., & Delsing, M. J. M. H. (2010). Continuous Time Modeling of Panel Data by means of SEM. In K. Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 201–244). Springer
Oud, J. H. L., & Voelkle, M. C. (2014). Do missing values exist? Incomplete data handling in cross-national longitudinal studies by means of continuous time modeling. *Quality & Quantity*, *48*, 3271–3288. https://doi.org/10.1007/s11135-013-9955-9
Petersen, L. A., Litteck, K., & Rohenroth, D. (2020). NEPS Technical Report for Mathematics: Scaling Results of Starting Cohort 3 for Grade 12. NEPS Survey Paper, Volume 75. Leibniz Institute for Educational Trajectories, National Educational Panel Study, Bamberg, Germany
Purpura, D. J., Logan, J. A. R., Hassinger-Das, B., & Napoli, A. R. (2017). Why do early mathematics skills predict later reading? The role of mathematical language. *Developmental Psychology*, *53*(9), 1633–1642. https://psycnet.apa.org/buy/2017-32731-001
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org
Rinne, L. F., Ye, A., & Jordan, N. C. (2020). Development of arithmetic fluency: A direct effect of reading fluency? *Journal of Educational Psychology*, *112*(1), 110–130. https://doi.org/10.1037/edu0000362
Ryan, O., Kuiper, R. M., & Hamaker, E. L. (2018). A continuous time approach to intensive longitudinal data: What, why and how? In K. van Montfort, J. H. L. Oud, & M. C. Voelkle (Eds.), Continuous time modeling in the behavioral and related sciences (pp. 27–57). Springer International Publishing. https://doi.org/10.1007/978-3-319-77219-6
Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology*, *109*(8), 1120–1140. https://doi.org/10.1037/edu0000193
Skopek, J. S., Pink, & Bela, D. (2012). *Data Manual. Starting Cohort 3 – From Lower to Upper Secondary School. NEPS SC3 1.0.0. NEPS Research Data Paper*. University of Bamberg
Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For? *The Journal of Human Resources*, *50*(2), 301–316. http://www.jstor.org/stable/24735988
Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort Profile: the Health and Retirement Study (HRS). *International Journal of Epidemiology*, *43*(2), 576–585. https://doi.org/10.1093/ije/dyu067
Steptoe, A., Breeze, E., Banks, J., & Nazroo, J. (2013). Cohort Profile: The English Longitudinal Study of Ageing. *International Journal of Epidemiology*, *42*, 1640–1648. https://doi.org/10.1093/ije/dys168
StataCorp. (2019). *Stata Statistical Software: Release 16. College Station*. TX: StataCorp LLC
Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2018). *User's manual for the ECLS-K: 2011 kindergarten-third grade data file and electronic codebook, public version*. Washington, DC: National Center for Education Statistics
Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24, 637–657*. https://doi.org/10.1037/met0000210
Usami, S. (2021). On the Differences between General Cross-Lagged Panel Model and Random-Intercept Cross-Lagged Panel Model: Interpretation of Cross-Lagged Parameters and Model Choice. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(3), 331–344. DOI: https://doi.org/10.1080/10705511.2020.1821690
Vanbinst, K., van Bergen, E., Ghesquière, P., & De Smedt, B. (2020). Cross-domain associations of key cognitive correlates of early reading and early arithmetic in 5-year-olds. *Early Childhood Research Quarterly*, *51*, 144–152. https://doi.org/10.1016/j.ecresq.2019.10.009

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The Role of Time in the Quest for Understanding Psychological Mechanisms. *Multivariate Behavioral Research*, *53*(6), 782–805. 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. L. (2015). Relating latent change score and continuous time models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 366–381. https://doi.org/10.1080/10705511.2014.935918

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, *115*(2), 227–244. https://doi.org/10.1016/j.jecp.2013.02.002

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From Data to Causes I: Building A General Cross-Lagged Panel Model (GCLM). *Organizational Research Methods*, *23*(4), 651–687

Publisher's Note