DOI: 10.1111/cdev.13817

EMPIRICAL ARTICLE

CHILD DEVELOPMENT

Reading to learn? The co-development of mathematics and reading during primary school

Nicolas Hübner[1] | Christine Merrell[2] | Helen Cramman[2] | John Little[3] | David Bolden[2] | Benjamin Nagengast[4,5]

[1]Institute of Education, University of Tübingen, Tübingen, Germany
[2]School of Education, Durham University, Durham, UK
[3]Centre for Evaluation and Monitoring, Cambridge University Press/Cambridge Assessment, Durham, UK
[4]Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany
[5]Department of Education and the Brain & Motivation Research Institute (bMRI), Korea University, Seoul, Korea

Correspondence
Nicolas Hübner, Institute of Education, University of Tübingen, Münzgasse 22-30, 72074 Tübingen, Germany.
Email: nicolas.huebner@uni-tuebingen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: ZUK 63

Abstract

Understanding how early reading and mathematics co-develop is important from both theoretical and pedagogical standpoints. Previous research has provided mixed results. This paper investigates the development of reading and mathematics in a longitudinal sample of $N = 355{,}883$ students from the United Kingdom (2005–2019) aged 5 to 12 (49% girls). Results indicate a positive relation between the development of the two domains. In addition, a substantial statistically significant positive association between prior reading scores and subsequent changes in achievement in mathematics was found, whereas changes in reading were smaller for students with a higher prior performance in mathematics. The findings suggest that acquiring good reading skills is highly relevant for developing mathematics skills. Implications for theory and practice are discussed.

The development of abilities and skills in the first years of life lays the foundation for lifelong development (Shonkoff & Phillips, 2000), including academic achievement (Duncan et al., 2007), health behaviors (Pagani & Fitzpatrick, 2014), and college graduation (McClelland et al., 2013). Spanoudis and Demetriou (2020) proposed that the human mind comprises several domain-specific processes (verbal, quantitative, spatial, causal, social, and linguistic) and a central control, involving integrative and cognizance processes. They suggested that these mental functions and their interactions are served by overlapping brain structures and therefore do not work in isolation. Many studies have investigated associations between several of these domain-specific processes, most prominently between reading and mathematics as shown in a recent meta-analysis by Peng et al. (2020). Prior studies have used different methodological approaches, ranging from univariate or purely correlational, cross-sectional research to more complex multivariate longitudinal models. Such methodological variation might come at the

Abbreviations: BIC, Bayesian information criterion; BLGC, bivariate latent growth curve; CEM, Centre for Evaluation and Monitoring; CFI, comparative fit index; CLPM, cross-lagged panel model; InCAS, The Interactive Computer Adaptive System; LCS, latent change score; LGC, latent growth curve; POPS, Problems of Position; RI-CLPM, random intercept CLPM; RMSEA, root mean square error of approximation; RQ, research question; TLI, Tucker–Lewis index.

Nicolas Hübner is Assistant Professor of Education at the Institute of Education, fellow of the College for Interdisciplinary Educational Research (CIDER), and was Bradshaw Fellow at the University of Durham in 2019. Part of this research was conducted at the Hector Research Institute of Education Sciences and Psychology. Sadly, before the paper was submitted for publication, Christine Merrell passed away in January 2021. We acknowledge Christine as an author and are grateful for the substantial contributions she made to the conception and design of the study, interpretation of the data, and writing of the paper before she passed away.

cost of addressing seemingly similar but actually different research questions (RQs; e.g., Orth et al., 2021; Usami et al., 2019). Overall, these studies have produced mixed evidence on the co-development of reading and mathematics, with some studies indicating that mathematics is a stronger predictor of subsequent reading (e.g., Duncan et al., 2007), some indicating that the two domains are reciprocally related (e.g., Koponen et al., 2007; Little et al., 2021; Peng et al., 2020; Vanbinst et al., 2020), and some suggesting that reading is a stronger predictor of subsequent mathematics skills or growth (e.g., Erbeli et al., 2021; Shin et al., 2013).

In this paper, we used a large-scale dataset of over 350,000 students from the United Kingdom between the ages of 5 and 12 years and aimed to extend previous research on the co-development of reading and mathematics during primary school. Investigating this co-development is important because knowledge about it can inform policy and practice on how a hypothetical minimally invasive intervention on one skill (e.g., reading) that did not directly affect the other (e.g., mathematics) may transfer to the other skill in the coming period. More practically speaking, knowledge about whether improving reading achievement also leads to progress in mathematics or vice versa can help educators prioritize when deciding how to invest limited resources in fostering students' skill development. In this study, we propose three sets of RQs from a univariate perspective (e.g., How does mathematics achievement develop over time?), a bivariate perspective (e.g., How do mathematics and reading develop in concert?), and considering reading and mathematics subscores and potentially differential developmental paths across subgroups (e.g., Are there differences in associations across grades or for different ability levels?).

The co-development of mathematics and reading

The relation between the development of reading and the development of mathematics has long been an area of research (Monroe & Engelhart, 1931). There is still uncertainty about the shape of this relation although more recent studies have tended to apply increasingly sophisticated statistical approaches over and above simple correlations (e.g., Bailey et al., 2020; Duncan et al., 2007; Grimm, 2008; Peng et al., 2020; Shin et al., 2013).

As outlined by Erbeli et al. (2021) and based on different frameworks on the development of reading and mathematics (e.g., Ehri, 2005; Geary, 1994; Perfetti, 1985), it seems reasonable to believe that mathematics and reading co-develop because they go through similar stages: (a) First, children begin elementary school with *basic competencies*, for instance, knowledge about sounds and respective letters in reading and counting or the comprehension of quantities in mathematics. Following this, (b) a phase of *procedural strategies* follows, in which children are required to learn the alphabet and relations between different numerical representations (e.g., visual and phonological). (c) These strategies are *applied and updated* over time and saved in children's long-term memory and finally, (d) the strategies are *applied in scenarios, which are increasingly cognitively demanding*, for instance, because they require complex inferences or strategies. Erbeli et al. (2021) suggested that because reading and mathematics develop through these similar stages, it seems reasonable to assume that they might influence each other as they develop.

These theoretical considerations are also supported by different empirical studies. For instance, in a meta-analysis, Peng et al. (2020) considered data from more than 360,000 participants (between the ages of 2 and 81.24) and found a substantial partial correlation between reading and subsequent mathematics when holding prior mathematics achievement constant (r = .20, $p<.05$). In addition, they found evidence for the opposite direction in which mathematics predicted later reading while controlling for prior reading achievement (r = .22, $p<.05$). Besides this meta-analysis, there are several longitudinal studies that have focused in particular on the development of reading and mathematics in students over time (see Table 1). These studies were not considered in the abovementioned meta-analysis, used different methods for analyzing longitudinal data, and thus provide important additional evidence that can contribute to our understanding of the reciprocal relations between mathematics and reading achievement.

In one of these studies, Grimm (2008) investigated associations between growth in students' achievement in mathematics from age 9 through 14 (Grades 3 to 8 in the United States) and their reading and mathematics achievement in Grade 3, focusing on average yearly growth in three domains: Problem Solving and Data Interpretation, Mathematical Computation, and Mathematical Concepts and Estimation. His results suggest a positive relation between reading achievement and achievement growth in mathematics in all domains, even after student characteristics (e.g., gender and ethnicity) and prior mathematics achievement were controlled for. Grimm's study did not control for cognitive abilities; therefore, it remains unclear whether reading and mathematics are naturally related because of common underlying dimensions and similar cognitive processes, as suggested by other scholars (Bailey et al., 2020; Rhemtulla & Tucker-Drob, 2011; Wrigley, 1958). In another study, Shin et al. (2013) used bivariate latent growth curve (BLGC) models to investigate how reading and mathematics developed over time for students in midwestern school districts in the United States (from Grades 4 to 7). The authors found that reading growth was positively associated with growth in mathematics. In addition, their study provided particularly interesting results regarding intercept-slope associations: Whereas reading achievement in Grade 4 was positively associated with learning gains in mathematics from Grades 4 to 7, mathematics achievement in Grade 4 was negatively

TABLE 1 Exemplary overview of prior studies, modeling approaches, and findings on longitudinal associations between reading and mathematics

Study	Model	Findings
Bailey et al. (2020)	CLPM, RI-CLPM	CLPM: Higher cross-lagged coefficients from mathematics to reading than vice versa RI-CLPM: slightly larger cross-lagged coefficients from reading to mathematics
Duncan et al. (2007)	Multiple regression models	Larger regression coefficients when predicting reading from mathematics than vice versa
Erbeli et al. (2021)	Univariate and Bivariate LCS	Reading positively related to changes in mathematics
Peng et al. (2020)	Meta-analysis	Comparable reciprocal associations between reading and mathematics and vice versa
Shin et al. (2013)	Bivariate LGC	Reading positively related to subsequent mathematics, mathematics negatively related to subsequent reading

Note: We only considered studies that investigated models with paths from prior reading to subsequent mathematics and vice versa. Note that the meaning of cross-lagged coefficients can differ substantially when different modeling strategies are applied (e.g., Usami et al., 2019).

Abbreviations: CLPM, cross-lagged panel model; LCS, latent change score model; LGC, latent growth curve model; RI-CLPM, random intercept CLPM.

associated with reading growth from Grades 4 to 7. This suggests that students with a higher achievement in mathematics in Grade 4 had less positive growth in reading over time.

Finally, Erbeli et al. (2021) investigated longitudinal reciprocal relations between reading and mathematics. They focused on $N = 554$ academically at-risk students from Texas, who were assessed repeatedly over the course of elementary school (Grades 1 to 4). Applying growth curve models and dual latent change score (LCS) models, they found that particularly average and above-average reading performances of at-risk students were associated with larger subsequent changes in mathematics. The authors reported that this association was stronger for students with low mathematics performance.

To sum up, recent findings in this area have been mixed, with some studies indicating that mathematics is a stronger predictor of subsequent reading (e.g., Duncan et al., 2007), some indicating that the two domains are associated with one another (e.g., Little et al., 2021; Peng et al., 2020), and some suggesting that reading is a stronger predictor of subsequent mathematics skills or growth (e.g., Bailey et al., 2020; Shin et al., 2013).

Methodological considerations

Recently, Bailey et al. (2020) argued that many effects from prior studies on the relation between mathematics and reading may be confounded, as most studies have not disentangled within-person variation from stable between-person differences. This critique mirrors current methodological discussions on the value and use of the cross-lagged panel model (CLPM) versus the random intercept CLPM (RI-CLPM). In the context of CLPMs and RI-CLPMs, between-person variation can be understood as time-stable (trait-like) differences between individuals, whereas within-person variation can be understood as intraindividual (state-like) fluctuations over time (Hamaker et al., 2015; Mulder & Hamaker, 2020). The CLPM does not distinguish between these two sources of variance, whereas the RI-CLPM separates them (Usami et al., 2019). Obviously, the criticism that between- and within-person variation is confounded can be generalized to most prior studies in this area, including prior meta-analytical findings. Most interestingly, Bailey et al. (2020) found that results from traditional CLPMs and RI-CLPMs have not matched well: Statistically significant cross-lagged coefficients from CLPMs disappeared or even reversed when they were modeled with RI-CLPMs.

However, even findings from RI-CLPMs might not ultimately provide the best answer on reciprocal effects between reading and mathematics achievement. As outlined by Lüdtke and Robitzsch (2021), these models might not necessarily perform better than traditional CLPMs or full forward CLPMs (Marsh & Craven, 2006) in detecting causal reciprocal effects and might be suitable only in a limited set of practical scenarios and given strong assumptions (e.g., Andersen, 2021). In addition, as outlined in more detail by Usami et al. (2019), LCS models, such as the ones we applied in this paper, allow users to control for unobserved time-varying and time-invariant confounding variables and might therefore provide an even stronger inferential basis for addressing the question of whether there are reciprocal effects between reading and mathematics achievement than alternative models can provide (e.g., RI-CLPMs).

Keeping this in mind, it becomes evident that the decision to choose a specific (longitudinal) model critically depends on the exact RQ that is being addressed and assumptions about the structure and nature of the underlying (true) processes (e.g., the existence of trends or time-varying or invariant differences between individuals). When investigating reciprocal relations between cognitive variables, both time-varying (e.g., motivation) and time-invariant differences (e.g., cognitive abilities, socioeconomic background) between individuals seem likely to occur and to influence the interplay between the two, as is evident from more recent publications (e.g., Orth et al., 2021; Usami et al., 2019).

To enhance the understanding of longitudinal models, Usami et al. (2019) proposed a unified framework of longitudinal models that can be used to examine reciprocal relations. This framework shows that most of the models that are commonly used to investigate longitudinal reciprocal relations (e.g., RI-CLPMs, LCS models) can be placed under one common umbrella, but depending on the exact specification of the model, interpretations of seemingly similar path coefficients might differ (e.g., between the RI-CLPM and the CLPM; see also Orth et al., 2021). Table 1 provides an exemplary overview of prior studies that investigated longitudinal associations between reading and mathematics, the applied methods, and the central findings regarding the paths from reading to mathematics and mathematics to reading. Table 1 does not provide a complete overview of all studies but is instead intended to show the great range of methods applied to investigate associations between reading and mathematics. As can be seen, there is no common agreement on which model should be chosen to investigate the (e.g., reciprocal) relation between reading and mathematics, as reflected by the huge heterogeneity in the models that have been chosen to address this question (see Table 1).

In this study, we were interested in how prior achievement in one domain (reading or mathematics) is related to intraindividual changes in the other domain, while considering trends and dynamics simultaneously. As outlined by Grimm et al. (2017), this question can be answered by applying a dual LCS model. The advantage of the dual change score model lies in its combination of aspects of growth models and CLPMs: It captures within-person changes in students' reading and mathematics achievement, differences in these changes between students, and associations between the different variables across measurement occasions (Grimm et al., 2017).

Therefore, in the current study, we extend previous research on the co-development of reading and mathematics by using these models and data from a large-scale assessment of students in primary school, Grades 1 to 6 (termed Year 1 to Year 6 in England and P2 to P7 in Scotland and Northern Ireland).

The present study

As outlined above, findings on the association and direction of mathematics and reading achievement in longitudinal data are mixed and model-dependent (see Table 1; Bailey et al., 2020; Grimm, 2008; Usami et al., 2019). In addition, there is a lack of research that has applied statistical models that are able to separate between- and within-person processes. In this study, we address this gap in the literature and closely investigate the association between the two constructs from a more general perspective by using LGC models (see Figure A1 and A3 in Appendix S1) and more specifically by using bivariate dual LCS models (see Figure 1).

To do this, we considered data from a large set of students (N = 355,897), who were assessed several times during primary school. Bivariate LCS models are particularly useful for investigating how changes in reading and mathematics achievement are associated, as such models emphasize within-person change. In addition, they explicitly allow researchers to test for differences between how reading is associated with changes in mathematics and how math is associated with changes in reading after accounting for stable developmental processes and time-varying confounders (e.g., Grimm et al., 2016, 2017; Klopack & Wickrama, 2020; McArdle, 2009). Considering the rich set of different prior studies that have investigated longitudinal associations between reading and mathematics, our study extends this important literature in three specific ways. First, we considered a large set of N = 355,883 students from the U.K. education system and followed them from Grades 1 to 6. Our study can therefore help to clarify the extent to which prior findings are generalizable when considering (a) a substantially larger sample than most relevant prior studies with (b) more measurement time points from (c) students from the United Kingdom rather than the United States, including (d) a broad sample of students from general elementary school, and applying (e) a modeling technique that places a stronger focus on within-person processes.

Second, our data allowed us to more thoroughly investigate whether the associations found for global reading and mathematics scores can also be found when considering subscores. Specifically, we were able to consider subscores on *Comprehension*, *Word Decoding*, and *Word Recognition* for reading and subscores on *Counting and Informal Arithmetic* (Numbers 1); *Algebra and Formally Presented Arithmetic* (Numbers 2); *Measuring, Shapes, and Space; and Handling Data* for mathematics. The consideration of subscores can help to clarify whether the general pattern found for reading and mathematics is present for all subscores or whether it is mostly driven by some of these subscores. Finally, we also conducted a closer exploration of potentially differential developmental paths of reading and mathematics, for instance, by comparing how associations between mathematics and reading might change from Grades 1 to 6 and by investigating differences in dynamic relations for different levels of reading and mathematics achievement.

Specifically, we proposed the following RQs:

RQ 1a: How does academic achievement in mathematics develop over the course of primary education?

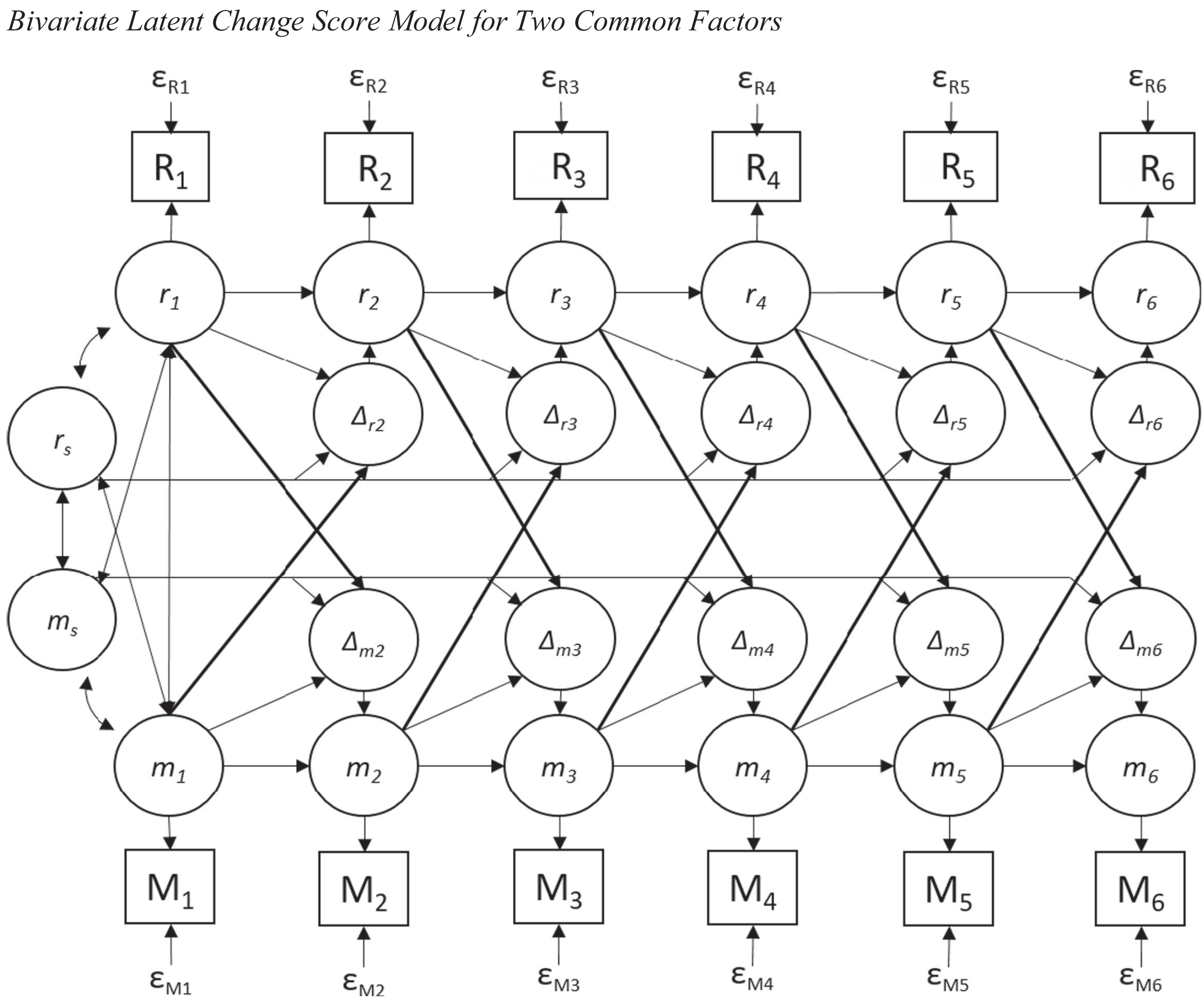

FIGURE 1 Bivariate latent change score model for two common factors. *Note*: Time-varying covariates not displayed for the sake of clarity

RQ 1b: How does academic achievement in reading develop over the course of primary education?

RQ 2a: How does academic achievement in reading and mathematics co-develop during primary education?

RQ 2b: How is the change in mathematics associated with previous reading achievement and vice versa, and which ability is the leading indicator in this dynamic process?

RQ 3a: Are there differences in co-developmental patterns between global mathematics and reading achievement scores and when considering different subscores in reading (i.e., *Comprehension, Word Decoding, and Word Recognition*) and mathematics (i.e., *Counting and Informal Arithmetic*, *Algebra and Formally Presented Arithmetic*, *Measuring, Shapes, and Space, and Handling Data*)?

RQ 3b: Are there differences across grades or different ability levels in the co-development of reading and mathematics?

Based on prior research, we were able to formulate four confirmatory (RQs 1a–2b) and two exploratory RQs (RQs 3a and 3b): For both RQs 1a and 1b, we expected to find positive growth in students' achievement. This expectation resulted from a large set of prior studies that found student achievement to increase over the course of primary school (e.g., Bailey et al., 2020; Bloom et al., 2008). Regarding RQs 2a and 2b, we were particularly interested in how growth in each construct is related to growth in the other (RQ 2a) and how changes in achievement in one construct (e.g., reading) is related to prior achievement in the other construct (e.g., mathematics) and vice versa. Regarding RQ 2a, based on prior research (Shin et al., 2013), we assumed to find positive associations between prior reading and subsequent growth in mathematics and negative associations between prior mathematics and subsequent growth in reading. For RQ 2b, we assumed to find positive coupling coefficients from reading to mathematics (Erbeli et al., 2021; Shin et al., 2013).

METHOD

Description of the study and sample

The data used in the current study came from the InCAS Assessment (The Interactive Computer Adaptive System), which was developed and conducted by the Centre for Evaluation and Monitoring (CEM; see www.cem.org). Due to legal issues and data protection rules, race of participants is generally not collected in studies by the CEM. Available population data from England, Northern Ireland, and Scotland in 2019 indicates that the majority of students in primary education are White British: 65.0% England, 81.9% Scotland, 94.7% Northern Ireland (Scottish Government, 2019; Toogood & Robinson, 2020; UK Government, 2020). Further detailed information about the ethnicity of pupils in schools in each nation is available in the references. The InCAS Assessment was developed to monitor students' achievement and progress in primary school (ages 5 to 12) and to provide teachers with diagnostic information about the strengths and weaknesses of individual children to inform their teaching. It consists of a suite of computer-adaptive tests, including mathematics, reading, spelling, and mental arithmetic (see Merrell and Tymms (2007) for an explanation of the rationale and development of the InCAS Assessment and Appendix S8 for further information on the sampling procedure). We used data provided by the CEM of $N = 884{,}826$ students from 3695 schools, who were assessed repeatedly during their years in primary school. We restricted the dataset to (a) students from schools in England, Northern Ireland, and Scotland; (b) data from 2005 to 2019 because the assessments from these years met standards of comparability; (c) students between the ages of 5 and 11 in year groups 1 through 6 (England), between the ages of 5 and 11 in Primary 2 through Primary 7 (Northern Ireland), and between the ages of 5 and 12 in Primary 2 through Primary 7 (Scotland); and (d) students who were assessed at a minimum of two time points, resulting in a sample of $N = 355{,}883$ students from 2614 schools (49% female students; see Table 2).

Instruments

Mathematics achievement

Achievement in mathematics was assessed with the InCAS general mathematics test, which included items relevant to the national curricula in England, Northern Ireland, and Scotland. Students were required to solve items from the areas of *Numbers 1* (covering counting and informal arithmetic), *Numbers 2* (covering algebra and formally presented arithmetic), *Measuring, Shapes, and Space* (covering the identification and understanding of the properties of 2D and 3D shapes and calculations involving time), and *Handling Data* (covering the interpretation and manipulation of information in tables, lists, and graphs). An overall mathematics score as well as scores for the individual areas were calculated. The test took 20 to 25 min, and all items were presented by an accompanying audio question prompt for the child generated through computer sound files to reduce reliance on the student's reading ability. Questions were presented as sentences asking the student to look at an image, problem, or calculation on the screen and then to select the correct answer from a choice of four (e.g., “What is the temperature on the thermometer?”; “Look at this sum: 2 + 1=? Now click on the answer!”; “In which list of fractions are all of the fractions equivalent?”). The reliability of mathematics achievement was high and ranged from .86 to .95 (Adams, 2005) across the different grade levels.

Reading achievement

The InCAS Assessment has three reading modules that are combined to produce an overall reading score: *Word Recognition*, *Word Decoding*, and *Comprehension*. In the *Word Recognition* module, students had to identify a single target word from a choice of five. The word was read aloud to the child using computer sound files, and then the child was given the word in the context of a sentence. For example, the item “so” (I am “so” tired) with the choice of answers: so, saw, sow, sew, os. At the easier levels, high- to medium-frequency words were given, whereas lower frequency words were used at more difficult levels. In the *Word Decoding* module, students had to identify the target word in a list of five unfamiliar or nonexistent words. Once again, the target word was spoken to the child using computer sound files. For example, selecting “frain” from a choice of fran, frin, frain, fain, fairn. In the *Comprehension* module, students had to read a passage of writing in which a choice of three plausible words was offered for approximately every fifth word. A number of different passages were available in the software, with the passage presented to the student determined by the student's scores in Word Decoding and Word Recognition. In each passage, the student had to choose the word that best fit that position in the sentence. Sometimes this involved choosing the word that was grammatically correct or was presented in the correct tense (e.g., “The children were/was/is playing with the toys”), and sometimes it involved choosing the word that had the correct meaning in the context of the sentence (e.g., “The space rocket was a white/while/long color”). Rasch scores were calculated and linearly transformed into age-equivalent scores. The test took 20 to 25 min. The reading achievement tests had high reliability, ranging from .95 to .98 (Adams, 2005) across the different grade levels.

TABLE 2 Cross-sectional descriptive statistics

Variable	*N*	*M*	*SD*	Min	Max
England	355,883	0.23	0.42	0	1.00
Northern Ireland	355,883	0.32	0.47	0	1.00
Scotland	355,883	0.45	0.50	0	1.00
Gender (1 = female)	355,309	0.49	0.50	0	1.00
GenMaths_T_1	53,701	7.01	0.91	3	13.35
GenMaths_T_2	127,203	7.84	1.05	3	14.84
GenMaths_T_3	96,096	8.67	1.32	3	16.00
GenMaths_T_4	212,799	9.05	1.50	3	16.00
GenMaths_T_5	199,470	9.58	1.58	3	16.00
GenMaths_T_6	191,987	10.43	1.54	3	16.00
Reading_T_1	53,592	6.45	1.62	4	14.10
Reading_T_2	127,372	7.83	1.80	4	14.17
Reading_T_3	96,222	8.96	1.93	4	14.46
Reading_T_4	211,281	9.42	2.00	4	16.00
Reading_T_5	198,966	9.82	2.04	4	16.00
Reading_T_6	190,780	10.86	1.96	4	16.00
DevAbil_T_1	52,711	6.44	1.87	3	14.89
DevAbil_T_2	125,148	7.92	1.87	3	16.00
DevAbil_T_3	91,107	9.27	2.00	3	16.00
DevAbil_T_4	182,835	10.11	2.03	3	16.00
DevAbil_T_5	128,579	10.81	2.23	3	16.00
DevAbil_T_6	138,638	11.83	2.06	3	16.00

Note: *N* = sample size. T_ indicates the year group (e.g., T_1_ = year group 1). We oriented on official enrollment cut-offs and included students who were legally allowed to be enrolled in Grade 1 (England: at least 5 years 0 months, Northern Ireland: at least 5 years 2 months, Scotland: at least 5 years 5 months) and students whose age would fit these enrollment policies in Grade 6 (England: not older than 11 years 10 months, Northern Ireland: not older than 11 years 11 months, Scotland: not older than 12 years 4 months).

Abbreviations: DevAbil, developed abilities; GenMaths, general mathematics; *M*, mean; *SD*, standard deviation.

Developed abilities

We also controlled for students' developed abilities. The InCAS Assessment provides an overall developed ability score that is based on a picture vocabulary test and a nonverbal ability test (Luyten et al., 2017; Merrell & Tymms, 2007). For the picture vocabulary test, students heard a word and then had to choose one of five pictures that represented the word. On the nonverbal ability test, students had to identify a specific pattern of dots from within a larger more complicated pattern. The nonverbal ability test is based on the Moseley (1976) Problems of Position (POPS) test. The test took 20 to 25 min. The Rasch scores from these tests had high reliability, ranging from .91 to .93 (Adams, 2005) across the different grades.

Month of assessment

Schools could choose to undertake the InCAS Assessment at any point during the academic year. We, therefore, controlled for variation in the months of assessment during the academic year. We defined 2-month assessment windows (e.g., JAN–FEB, MAR–APR) to control for the time of assessment. The choice of 2-month windows resulted from considerations of the ease of the interpretation of the results and for practical reasons because some cells/months had only a few cases (particularly toward the end of the school year when the summer holidays were beginning). We decided to code the beginning of the school year (SEP–OCT) as 0 and the remaining 2-month blocks in ascending order.

Statistical analyses

To address the different RQs, we first examined descriptive statistics. In doing this, we considered the cross-sectional achievement of the students in mathematics and reading. Afterward, we performed an in-depth analysis of the data by applying a variety of models to conduct longitudinal analyses (Grimm et al., 2017). As we were particularly interested in developmental processes, we used latent linear growth curve models (RQs 1a and 1b), bivariate LGC models (RQs 2a and 3a), and bivariate dual LCS models (RQs 2b, 3a, and 3b) to investigate the dynamic development of reading and mathematics. We expand upon these models in the following sections.

LGC models

To investigate growth in students' learning of mathematics and reading (RQs 1a and 1b), we specified first-order linear LGC models (see Figure A1 in Appendix S1) based on the scores for the six grades of students in our study (e.g., Bollen & Curran, 2006). The correlations between the latent intercept and slope factor indicate potential differences in the growth over time, depending on achievement at the first measurement occasion (e.g., a negative correlation suggests that students with higher starting values will have shallower growth compared with students with lower starting values and vice versa). LGC models were specified separately for mathematics and reading (see Figure A1 in Appendix S1).

BLGC model

To address RQs 2a and 3a, we specified a joint model with both growth curve models for reading and mathematics and took a closer look at the correlations of the latent intercept and slope factors across the two domains (see Figure A3 in the Appendix S1). This was done to obtain initial evidence of potential dual-process growth trends. A positive correlation between the latent slopes of the two domains would indicate that individual growth over time in the two constructs develops rather uniformly (more/less growth in one domain goes along with more/less growth in the other domain), whereas a negative correlation would indicate differential growth patterns (more/less growth in one domain goes along with less/more growth in the other domain). In addition, the correlation between the intercept and slope factors of opposing domains (e.g., reading intercept with mathematics slope and vice versa) would indicate whether higher or lower achievement at the first measurement occasion in one domain goes along with steeper or shallower growth trends in the other domain.

Bivariate dual LCS model

To address RQs 2b, 3a, and 3b, we specified bivariate dual LCS models (McArdle, 2009; see Figure 1). As outlined above, we were particularly interested in how one construct (e.g., reading) is related to change in the other construct (e.g., mathematics) and vice versa. As outlined in Equations (1 and 2), in these models, change (e.g., Δf_{mit}) is considered to be a function of a *constant change parameter* (α_{mt}), the preceding score on the same construct (the *proportional change* parameter; $\beta_m f_{mi,t-1}$), the preceding score on the other construct (the *coupling parameter*; $\gamma_m f_{ri,t-1}$), and a residual term (ϵ_{mit}):

$$\Delta f_{mit} = \alpha_{mt} + \beta_m f_{mi,t-1} + \gamma_m f_{ri,t-1} + \epsilon_{mit}, \tag{1}$$

$$\Delta f_{rit} = \alpha_{rt} + \beta_r f_{ri,t-1} + \gamma_r f_{mi,t-1} + \epsilon_{rit}. \tag{2}$$

Note that these models often include a so-called *intercept factor* in addition to the constant change factor (McArdle, 2009). However, this intercept factor actually coincides with the latent T1 measure of the respective construct (Usami et al., 2019). Comparable to CLPMs, LCS models are used to compare the coupling parameters across the two opposing constructs (Klopack & Wickrama, 2020). LCS models typically assume time-invariant proportional change and coupling parameters (Usami et al., 2019). Exemplary code for this model can be found in Appendix S2.

Additional specifications

In all cases, we decided to specify adjusted and unadjusted models. Adjusted models also controlled for developed abilities, whereas we controlled for the time of the assessment during the year in all models. We addressed the nested data structure (i.e., students nested in schools) using cluster-robust standard errors (Snijders & Bosker, 2012). In addition, we applied FIML, which is implemented in Mplus (Muthén & Muthén, 1998–2017), in all our analyses to deal with missing data. Additional information on the statistical analyses can be found in Appendix S3.

RESULTS

The development of academic achievement in mathematics (RQ 1a)

We first took a closer look at the development of mathematics achievement over time. Descriptive statistics (see Table 2) showed a general, cross-sectional improvement in test scores over time. On average, students in Grade 1 had a score of M = 7.01 points (SD = 0.91), and this score increased over the course of primary school to M = 10.43 points (SD = 1.54) in Grade 6 (see Figure A2 in Appendix S1).

Next, we specified LGC models for reading and mathematics. The model fits are displayed in Table 3 (Models 1 to 4). We specified unadjusted models, which did not include developed abilities, and adjusted models, which included this time-varying covariate. Both adjusted and unadjusted models had sufficient fit with regard to the comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA), above the traditional cut-off criteria (Hu & Bentler, 1999). The unadjusted model in mathematics showed a good fit to the data, χ^2 = 4119.81, p<.001, df = 46, Bayesian

TABLE 3 Model fits of the linear latent growth curve models

Model	Additional adj.	χ^2	*df*	BIC	CFI	TLI	RMSEA
1. Mathematics	No	4119.814***	46	5,184,032.032	.97	.97	.02
3. Reading	No	7942.514***	46	5,621,686.349	.97	.97	.02
5. Mathematics and reading	No	16,278.866***	190	7,866,715.458	.98	.98	.02
2. Mathematics	Yes	30,531.86***	76	7,493,295.36	.97	.96	.03
4. Reading	Yes	53,471.01***	76	8,021,612.57	.97	.96	.04
6. Mathematics and reading	Yes	73,072.86***	250	10,196,532.28	.96	.95	.03

Note: Adjusted models controlled for developed abilities, and all models controlled for the time of assessment.

Abbreviations: BIC, Bayesian information criterion; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation.

***$p < .05$.

information criterion (BIC) = 5,184,032.032, CFI = .97, TLI = .97, RMSEA = .02. The fit statistics for the adjusted model were comparable.

Table 4 presents the results of the LGC parameters. We found an intercept mean of $M = 6.08$ and a positive mean of the slope of $M = 0.86$ (both $ps < .001$) in the unadjusted models, and these were strongly comparable to the results from the adjusted models. These findings suggest a positive growth in mathematics achievement scores of 0.86 points per year, which can be interpreted as an average growth of $d = .65$. In addition, we found variation in both parameters (i.e., intercepts and slopes) between students (Intercept: $s^2 = 0.86$, $p < .001$; Slope: $s^2 = 0.03$, $p < .001$), suggesting that students' academic achievement varied with respect to the first measurement occasion and also with respect to its growth. When controlling for developed abilities, the means of the intercept and slope remained comparable (Intercept: $M = 6.72$, Slope: $M = 0.76$, both $ps < .001$). In this model, the variation in slopes and intercepts turned out to be smaller and decreased to $s^2 = 0.30$, $p < .001$ (intercept) and $s^2 = 0.02$, $p < .001$ (slope), which might be interpreted as evidence that the covariate explained some of the individual differences. In addition, we found a positive correlation between the growth over time and the initial level of mathematics achievement in the unadjusted models ($r = .34$, $p < .001$), suggesting that students with higher mathematics achievement at T1 had steeper growth in learning, compared with students with lower achievement at T1. This correlation was not statistically significantly different from zero after we adjusted for developed abilities ($r = -.02$, $p = .549$).

The development of academic achievement in reading (RQ 1b)

When we more closely investigated students' reading achievement across different grade levels, we found an average score of $M = 6.45$ points ($SD = 1.62$) in Grade 1, which increased over the course of primary school to $M = 10.86$ points ($SD = 1.96$; see Table 2 and Figure A2 in Appendix S1). A closer look at the fit statistics of the LGC models (see Table 2) suggested a sufficient fit of the adjusted and unadjusted models for reading achievement. The model fits for reading achievement in the unadjusted model were good, $\chi^2 = 7942.51$, $p < .001$, $df = 46$, BIC = 5,621,686.349, CFI = .97, TLI = .97, RMSEA = .02. The fit statistics for the adjusted model were comparable and above the suggested cut-off criteria (Hu & Bentler, 1999).

Parameter estimates for the growth curve models are presented in Table 4. The intercept for the unadjusted model for reading was $M = 5.42$, and the average growth in reading across time was positive $M = 1.11$ (both $ps < .001$). The average growth in reading achievement amounted to $d = .59$ per year. These findings were largely comparable to the results for the adjusted models. We found statistically significant fvariation in both intercepts and slopes between students (Intercept: $s^2 = 3.74$, $p < .001$; Slope: $s^2 = 0.03$, $p < .001$). When we controlled for developed abilities, the variability in the slope and intercept decreased to $s^2 = 1.96$, $p < .001$ (intercept) and $s^2 = 0.03$, $p < .001$ (slope). In contrast to mathematics, we found a negative correlation between growth over time and initial reading achievement in the unadjusted model and the adjusted model (both $r = -.37$, $p < .001$), suggesting that students with higher abilities at T1 had shallower growth in their achievement, compared with students with lower achievement at T1.

The joint development of competencies in mathematics and reading (RQ 2a)

To investigate the co-development of reading and mathematics, we first specified bivariate LGC models, which jointly considered mathematics and reading (see Figure A3 in the Appendix S1). The fit statistics for the unadjusted model were good ($\chi^2 = 16{,}278.87$, $p < .001$, $df = 190$, BIC = 7,866,715.458, CFI = .98, TLI = .98, RMSEA = .02). The same held for the adjusted model (see Table 3). Results for the bivariate LGC model were in line with the proposed cut-off criteria (see Table 3; Hu & Bentler, 1999). Results for the means and variances of the latent intercept and slope factors were largely comparable to the results from the univariate LGC models

TABLE 4 Results of the latent growth curve models for mathematics

Model	Additional adj.	Intercept				Linear slope				Correlation	
		M	*p*	s^2	*p*	*M*	*p*	s^2	*p*	*r*	*p*
Mathematics	No	6.08	<.001	0.86	<.001	0.86	<.001	0.03	<.001	.34	<.001
Mathematics	Yes	6.72	<.001	0.30	<.001	0.76	<.001	0.02	<.001	−.02	.549
Mathematics and reading: mathematics	No	6.03	<.001	0.99	<.001	0.89	<.001	0.03	<.001	.28	<.001
Mathematics and reading: mathematics	Yes	6.66	<.001	0.33	<.001	0.77	<.001	0.02	<.001	.05	.048
Reading	No	5.42	<.001	3.74	<.001	1.11	<.001	0.03	<.001	−.37	<.001
Reading	Yes	6.26	<.001	1.96	<.001	0.94	<.001	0.03	<.001	−.37	<.001
Mathematics and reading: reading	No	5.40	<.001	3.75	<.001	1.12	<.001	0.03	<.001	−.36	<.001
Mathematics and reading: reading	Yes	6.20	<.001	2.05	<.001	0.77	<.001	0.02	<.001	−.36	<.001

Note: Adjusted models controlled for developed abilities, and all models controlled for the time of assessment. Correlations display the correlation of the random intercept and random slope of the respective model. Correlations between the intercepts and slopes for mathematics with reading amounted to: Intercepts: $r = .82$, $p < .001$ and Slopes: $r = .33$, $p < .001$, in the unadjusted model. In the adjusted model, they were as follows: Intercepts: $r = .62$, $p < .001$, and Slopes: $r = .34$, $p < .001$. The correlation between the intercept in mathematics and the slope in reading in the unadjusted model amounted to $r = -.31$, $p < .001$ ($r = -.30$, $p < .001$ in the adjusted model). The correlation between the intercept in reading and the slope in mathematics in the unadjusted model amounted to $r = .37$, $p < .001$ ($r = .23$, $p < .001$ in the adjusted model).

(see Table 4). In these models, we were particularly interested in the correlations between the intercept and slope factors, as initial evidence of the nature of the co-developmental process.

In the unadjusted models, we found intercept-slope correlations that amounted to $r = .28$ ($p < .001$) for mathematics and $r = -.36$ for reading ($p < .001$). For mathematics, this suggests that, comparable to the univariate model, a higher score at T1 was associated with steeper achievement growth from T1 to T6. For reading, higher reading achievement at T1 was, similar to the univariate models, associated with shallower achievement growth over time. Results for the adjusted models were comparable but decreased to $r = .05$ ($p = .048$) for mathematics and remained similar for reading.

As outlined above, bivariate LGC models were used to investigate relations between achievement growth and achievement at T1 across the two constructs. First, we found a positive association between the growth in both constructs, as indicated by a positive correlation between slopes in both the unadjusted ($r_s = .33$, $p < .001$) and adjusted ($r_s = .34$, $p < .001$) models. Achievement in reading and mathematics was strongly associated as indicated by strong positive intercept correlations (unadjusted model: $r_i = .82$, $p < .001$, adjusted model: $r_i = .62$, $p < .001$).

Most interestingly, in line with previous research (Shin et al., 2013), we found a negative relation between the mathematics intercept and growth in reading in the unadjusted model $r = -.31$ ($p < .001$), whereas the association between the reading intercept and growth in mathematics was positive $r = .37$ ($p < .001$). These results were strongly comparable to the results found in the adjusted models. Students with higher mathematics achievement in Grade 1 showed a somewhat shallower growth in reading achievement from Grade 1 to Grade 6 compared with students with lower mathematics achievement in Grade 1 who showed a somewhat steeper growth in reading from Grade 1 to Grade 6. By contrast, students with higher reading achievement in Grade 1 had a steeper growth in mathematics from T1 to T6 compared with students with lower reading achievement at the first measurement occasion who showed a shallower growth in mathematics.

Associations between changes in reading and mathematics and prior achievement (RQ 2b)

Model fit

Finally, we specified bivariate dual LCS models for mathematics and reading (see Figure 1). As suggested by Grimm et al. (2017) and Klopack and Wickrama (2020), we specified four types of adjusted and unadjusted models: (a) models in which all coupling parameters were constrained to zero, (b) models in which only coupling parameters from reading to changes in mathematics were constrained to zero, and (c) models in which only coupling parameters from mathematics to reading were constrained to zero. Finally, in the fourth model, (d) both coupling coefficients were estimated. Model fits for these models are presented in Table 5. Note that all models adequately represented the data, indicated by good global model fits (e.g., RMSEA < .05; CFI and TLI > .95; see Table 5).

We compared the adjusted no coupling model with the model in which coupling parameters from reading to changes in mathematics were constrained to zero using Satorra- Bentler-scaled chi-square difference tests (Satorra & Bentler, 2010). We found that the model with paths from mathematics to changes in reading had a statistically significantly better model fit than the no coupling model (χ^2 = 327.64, df = 1, p < .001). When comparing the no coupling model with the model with paths from reading to changes mathematics, we found a statistically significant better model fit of the latter model (χ^2 = 198.09, df = 1, p < .001). Finally, the full coupling model showed a statistically significant better model fit, compared with a model with paths from mathematics to changes in reading (χ^2 = 381.55, df = 1, p < .001) and compared with a model with paths from reading to changes in mathematics (χ^2 = 18.94, df = 1, p < .001; see Table 5). Regarding unadjusted models, we found a similar pattern of results (see Table 5). In summary, both the adjusted solution and the unadjusted solution suggest a superior fit of the full coupling model. Practically, this can be interpreted as evidence that both reading and mathematics are relevant for developmental processes on the opposing construct (Grimm et al., 2017).

Proportional change parameters

Next, we conducted a closer examination of the coupling parameters in the full coupling model (see Table 6): The results were quite stable across all models regarding the directions of the parameter estimates and their statistical significance. Regarding the unadjusted full coupling model, the proportional change parameters were β = .26 for reading and β = −.53 for mathematics (both ps < .001), suggesting a higher change in reading for students with higher previous reading scores and a lower change in mathematics for students with higher previous achievement in mathematics. In the full coupling model with adjustment, the proportional change parameter for reading was no longer statistically significant (β = −.04, p = .056) and the parameter for mathematics decreased to β = −.18 (p < .001).

Coupling parameters

Regarding the coupling parameters, we found a similar pattern for both the adjusted and unadjusted full coupling models. Here, changes in reading achievement were negatively related to prior achievement in mathematics (β_{uadj} = −.27 and β_{adj} = −.06, both ps < .001), whereas changes in mathematics achievement were positively related to prior reading achievement (β_{uadj} = .72 and β_{adj} = .24, both ps < .001). These findings are in line with findings from the partial coupling models in suggesting that a higher level of previous reading achievement leads to larger changes in subsequent mathematics achievement (after previous mathematics achievement is controlled for), whereas a higher level of previous mathematics achievement leads to shallower changes in reading achievement (after previous reading achievement is controlled for).

Associations between reading and mathematics subscores (RQ 3a)

To address RQ 3a, we specified adjusted and unadjusted growth curve and bivariate dual LCS models

TABLE 5 Model fits of the bivariate dual latent change score models

Model	Additional adj.	χ^2	df	BIC	CFI	TLI	RMSEA
LCS no coupling	No	13,977.575***	131	7,010,825.560	.98	.98	.02
LCS partial coupling: M → ΔR	No	13,286.159***	130	7,008,460.798	.99	.98	.02
LCS partial coupling: R → ΔM	No	12,050.990***	130	7,002,424.389	.99	.99	.02
LCS full coupling	No	10,057.269***	129	6,994,477.863	.99	.99	.02
LCS no coupling	Yes	70,490.87***	191	9,341,781.69	.97	.97	.03
LCS partial coupling: M → ΔR	Yes	71,023.18***	190	9,341,794.41	.97	.97	.03
LCS partial coupling: R → ΔM	Yes	70,319.96***	190	9,341,106.61	.97	.97	.03
LCS full coupling	Yes	70,775.36***	189	9,340,984.03	.97	.97	.03

Note: Adjusted models controlled for developed abilities, and all models controlled for the time of assessment. LCS full coupling free = no equality constraints on coupling and proportional change parameters. Note that the chi-square values reported here cannot be used for chi-square difference testing. Instead, we used the recommended Satorra-Bentler-scaled chi-square difference test (Satorra & Bentler, 2010).

Abbreviations: BIC, Bayesian information criterion; CFI, comparative fit index; LCS, latent change score model; M, mathematics; R, reading; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation.

***p < .05.

TABLE 6 Results for the bivariate dual latent change score models with full coupling and partial coupling

			Proportional change				Coupling parameters		
Model	**Additional adj.**		***β***	***SE***	***p***		***β***	***SE***	***p***
LCS partial coupling: M → ΔR	No	ΔR_{ti} on R_{ti-1}	.13	.01	<.001	ΔR_{ti} on M_{ti-1}	−.17	.01	<.001
		ΔM_{ti} on M_{ti-1}	.01	.01	.079	ΔM_{ti} on R_{ti-1}	—	—	—
LCS partial coupling: M → ΔR	Yes	ΔR_{ti} on R_{ti-1}	−.12	.02	<.001	ΔR_{ti} on M_{ti-1}	.00	.01	.926
		ΔM_{ti} on M_{ti-1}	−.08	.01	.119	ΔM_{ti} on R_{ti-1}	—	—	—
LCS partial coupling: R → ΔM	No	ΔR_{ti} on R_{ti-1}	−.09	.01	<.001	ΔR_{ti} on M_{ti-1}	—	—	—
		ΔM_{ti} on M_{ti-1}	−.32	.01	<.001	ΔM_{ti} on R_{ti-1}	.46	.02	<.001
LCS partial coupling: R → ΔM	Yes	ΔR_{ti} on R_{ti-1}	−.12	.01	<.001	ΔR_{ti} on M_{ti-1}	—	—	—
		ΔM_{ti} on M_{ti-1}	−.15	.01	<.001	ΔM_{ti} on R_{ti-1}	.20	.02	<.001
LCS full coupling	No	ΔR_{ti} on R_{ti-1}	.26	.02	<.001	ΔR_{ti} on M_{ti-1}	−.27	.02	<.001
		ΔM_{ti} on M_{ti-1}	−.53	.02	<.001	ΔM_{ti} on R_{ti-1}	.72	.03	<.001
LCS full coupling	Yes	ΔR_{ti} on R_{ti-1}	−.04	.02	.056	ΔR_{ti} on M_{ti-1}	−.06	.01	<.001
		ΔM_{ti} on M_{ti-1}	−.18	.02	<.001	ΔM_{ti} on R_{ti-1}	.24	.02	<.001

Note: These results are based on models in which achievement scores were standardized on the basis of their respective mean and standard deviation at T1. The fit of these models can be found in Table 5.

Abbreviations: LCS, latent change score model; M, mathematics; R, reading; SE, standard error.

for all combinations of reading (i.e., three) and mathematics (i.e., four) subscores (see the Instruments section). The results from these models can be found in Supporting Information (Tables E1–E10 in Appendix S5). The model fit suggested that LGC models and LCS models fit the data well (see Tables E1 and E2 in Appendix S5) with all models showing CFI/TIL ≥ .96 and RMSEA ≤ .03. When inspecting results for the BLGC models (see Tables E3–E6 in Appendix S5), we were particularly interested in the intercept-slope correlations across the different subscores. We found statistically significant negative correlations between the intercepts of all the mathematics subscores with all the reading subscore slopes, whereas the intercepts of the reading subscores were positively associated with the mathematics slopes in the adjusted models (12 out of 12). This pattern was also found for eight of 12 unadjusted models. Similar to our findings for the global scores, this suggests that higher reading subscores in Grade 1 were associated with steeper subsequent growth in mathematics, whereas higher mathematics subscores in Grade 1 were associated with shallower growth in reading from Grades 1 to 6.

When inspecting the results for the LCM models (see Tables E7–E10 in Appendix S5), we found similar results for the majority of the models, in line with the reported findings when using the global mathematics and reading scores and our findings from the LGCMs: The coupling parameters between prior reading subscores and the change in mathematics subscores were positively associated, whereas we found a negative association between prior mathematics subscores and the change in reading subscores. This pattern was consistent across all the subscores in the unadjusted models. For the adjusted models, we found this pattern of results for eight of the 12 models, whereas for the four models considering the reading subscore decoding (i.e., Number 1 and Decoding, Number 2 and Decoding, MSS and Decoding, and Data and Decoding) we found a positive association between prior mathematics subscores and subsequent changes in reading subscores. This suggests that for changes in decoding, mathematics achievement might be more relevant than for other reading subscores.

Subgroup differences in the co-development of reading and mathematics (RQ 3b)

To address this RQ, we re-specified the adjusted LCS and allowed the coupling parameters to vary freely from time point to time point. Doing this required us to constrain two correlations that were not statistically significantly different from each other to be equal (r1 with sr and m1 with sm; see Appendix S2). We found the following coupling parameters when predicting changes in mathematics from prior reading: β_{T1} = .18 (p < .001), β_{T2} = .20 (p < .001), β_{T3} = .02 (p = .646), β_{T4} = .08 (p = .018), β_{T5} = .09 (p = .017). Of these, all the coefficients were statistically significantly different from each other (all ps < .01) except for β_{T1} and β_{T2} (p = .336) and β_{T4} and β_{T5} (p = .267). For coupling parameters from prior mathematics to subsequent changes in reading, we found: β_{T1} = −.54 (p < .001), β_{T2} = −.55 (p < .001), β_{T3} = −.50 (p < .001), β_{T4} = −.40 (p < .001), β_{T5} = −.28 (p < .001). When comparing the different coefficients, we found that all coefficients were statistically significant (all ps < .05) except for β_{T1} and β_{T2} (p = .696) and

β_{T2} and β_{T3} (p = .483). Overall, these findings suggest that associations between prior reading scores and changes in mathematics were stronger in earlier grades (i.e., G1 and G2) than in later grades. The negative associations between prior mathematics scores and subsequent changes in reading were stronger in earlier years and diminished over time.

We investigated differences in associations for different achievement levels in reading and mathematics by using a statistical vector plot, which visualizes dynamic relations between different levels of the variables under investigation (e.g., Grimm et al., 2017; McArdle, 2009). The statistical vector field plot that resulted from our adjusted LGCM with fully standardized reading and mathematics scores (z-standardized) is presented in Appendix S7. We found several interesting patterns. First, the largest change in mathematics achievement predicted by reading was found for average and above-average reading achievement levels in combination with low mathematics achievement levels. This can be seen in the steep arrows in the middle and bottom right of the figure. Second, the lower a student's reading achievement, given a low level of mathematics achievement, the smaller the changes in mathematics. This is indicated by the arrows that are more horizontal when moving from the bottom right of the figure toward the bottom left of the figure. Finally, the figure suggests that improvements in the average level of mathematics achievement seem more likely to happen for students with at least slightly above-average reading achievement. This is indicated by the horizonal or downward arrows in the middle left of the figure, which get steeper when moving to the right of the figure. Overall, our results are strongly in line with findings from prior studies (e.g., Erbeli et al., 2021) and suggest that the association between reading and changes in mathematics and vice versa can differ between different ability levels.

DISCUSSION

Findings

In this study, we investigated the co-development of reading and mathematics. Previous research has produced mixed evidence on whether reading or mathematics is the leading indicator in this co-developmental process. We first specified univariate LGC models, followed by bivariate LGC models and bivariate dual LCS models. Particularly, the latter allows for a focus on within-person changes, which was called for in previous research (Bailey et al., 2020).

Regarding univariate LGC models, our findings suggest that positive development occurs in both mathematics and reading from Grade 1 to Grade 6. In addition, we found variation in reading and mathematics at the first measurement occasion as well as in their growth trajectories. Next, we investigated the co-development of reading and mathematics using bivariate LGC models. Here, our findings were largely in line with prior research in suggesting that initial achievement (intercepts) in mathematics and reading were positively related to one another, and so were the growth trajectories (slopes) in the two opposing constructs (e.g., Aiken, 1971; Grimm, 2008). Furthermore, our findings also revealed a pattern previously reported by Shin et al. (2013): Students with higher mathematics achievement in Grade 1 showed shallower growth in reading achievement from Grade 1 to Grade 6, even after we controlled for cognitive abilities and initial reading achievement. By contrast, students with higher reading achievement in Grade 1 showed steeper growth in mathematics from T1 to T6 compared with students with lower reading achievement at the first measurement occasion (who showed shallower growth in mathematics).

In order to more closely investigate the question of which variable is the leading indicator, we specified adjusted and unadjusted bivariate dual LCS models. Here, we found that the full coupling model with paths from reading to changes in mathematics and vice versa (see Table 5) had the best fit to the data. Regarding the coupling parameters, the pattern of results suggests that whereas prior reading achievement was positively associated with subsequent changes in mathematics, higher mathematics achievement was negatively associated with subsequent changes in reading achievement. Students with higher achievement in reading, therefore, showed a somewhat higher subsequent change in mathematics achievement compared with students with lower previous reading achievement. By contrast, students with higher achievement in mathematics showed a shallower subsequent change in reading achievement compared with students with lower prior mathematics achievement.

When we analyzed associations between reading and mathematics on the more fine-grained level of subscores, we found that we were largely able to replicate the result patterns found for global mathematics and reading scores. Practically, this suggests that our findings for reading and mathematics scores were most likely not driven unevenly by some subscores more than others. Most interestingly, when inspecting how coupling parameters changed over the course of primary school, we found that associations between prior reading scores and changes in mathematics were stronger in earlier grades (i.e., G1 and G2) than in later grades and that negative associations between prior mathematics scores and subsequent changes in reading were stronger in earlier years and diminished over time. Practically, however, the directions of these associations (see Table 6) remained largely comparable across timepoints. Finally, our visualization of differential associations between reading and mathematics given different ability levels suggests that, in line with prior studies (e.g., Erbeli et al., 2021), particularly students with average to high reading achievement and low mathematics achievement made the greatest progress in mathematics and that with decreasing reading achievement, changes in mathematics became smaller.

In summary, our findings are largely in line with recent prior findings in the field in suggesting that reading constitutes a relevant predictor of subsequent mathematics (e.g., Erbeli et al., 2021; Shin et al., 2013). More specifically, we found a positive coefficient for the association between prior reading and subsequent changes in mathematics but a negative coefficient for the path from prior mathematics achievement to changes in reading achievement, which might be explained by the fact that our study consisted of elementary school students who were mostly concerned with foundational mathematics. Future longitudinal studies are required to test whether the negative path from mathematics to changes in reading might vanish or even reverse for older students when mathematics "heavily involves high-level cognition" (Peng et al., 2020, p. 23).

Limitations

There are some limitations that need to be mentioned before the implications are outlined. First, although we considered a very large set of students using the CEM database, we were not able to repeatedly assess all students every year. This sometimes resulted in low covariance coverage across grades. Specifically, for some grades (e.g., Grades 1 and 6), depending on the statistical model, there were only about $n = 4271$ cases (i.e., 1.2%) with information at both the very first and the very last assessments (e.g., in Grades 1 and 6). Although this might still seem to be a large number, this drop in cases might have resulted in less precise estimates.

Another limitation resulted from the limited set of covariates that we were able to consider. For instance, due to limited availability and model complexity, we were not able to consider students' socioeconomic status or gender in our models. Furthermore, school-level variables that might be related to the developmental processes were not available. In the adjusted models, we were able to consider overall developed ability scores that were based on a *picture vocabulary* test and a *nonverbal ability* test. These tests assess learning aspects that are not specifically taught in the school curriculum. Future studies should more closely investigate differences in these developmental processes between boys and girls, consider students' socioeconomic background, and consider school-level variables in order to better explain variation in the co-development of these taught skills.

In addition, it seems important to keep in mind that our data stems from England, Scotland, and Northern Ireland. Generalizations of our findings should not be carried out without great caution and consideration of the specificities of the learning background and environment. Notably, studies from other countries like the United States or studies considering different student populations such as academically at-risk children (e.g., Erbeli et al., 2021; Shin et al., 2013), have provided comparable results. However, more studies are needed that explicitly investigate potential cross-cultural and cross-country differences in the co-development of reading and mathematics.

Finally, it is important to understand that in order for our results to be interpreted in a causal sense, specific assumptions have to hold (e.g., strong ignorability), which might be more or less plausible in the specific context of our study. In order to address this threat, we made use of an additional time-varying covariate and LCS models, which have been attested a stronger basis for estimating causal effects than other models used in prior research on this topic, particularly when trends are present in the data (e.g., the CLPM or the RI-CLPM; Usami et al., 2019). Furthermore, our models are able to yield causal effects only if the underlying data-generating process is adequately reflected. This means that if confounding variables are present in our study and these behave as our LCS models assume, our models will yield causal effects. However, as for all prior studies using longitudinal models from Usami et al.'s (2019) unified framework, this remains an untestable assumption.

Implications

The major finding of our study is that reading seems to be a particularly important skill for subsequent changes in mathematics from early on. There are several implications of our study for research on the co-development of reading and mathematics in the age range of students in primary school. First, it seems very important to investigate abilities not only from a unidimensional but also from a multidimensional perspective. This means that in order to understand the development of mathematics, it seems important to also consider development in reading. Specifically, our findings from unidimensional models suggest a continuous strong (average) development in both reading and mathematics. Only when investigating bivariate within-person processes were we able to detect the interdependencies of the two constructs, which suggest a reciprocal relation between reading and mathematics, with positive paths from reading to mathematics but negative paths from mathematics to reading. As outlined above, this finding might result from the specific group of students considered in this study (elementary school students) and might change as students get older. Considering links between initial abilities and subsequent achievement growth and changes seems important for developing a more complete understanding of these relations across the lifespan.

Related to this, it seems beneficial to make use of more advanced methods along with large-scale datasets. Recently, there has been an increase in scientific publications and freely available resources on why and how to apply (e.g., within-person process) models in statistical software (e.g., Bailey et al., 2020; Berry & Willoughby, 2017; Grimm et al., 2017; Hamaker et al., 2015; McNeish & Hamaker, 2020). Unfortunately, there is still a lack of applications of these more advanced methods, which allow for a narrower but also a much more complex investigation. Going beyond these methods might also include using identification strategies, such as regression-discontinuity designs to investigate effects of mathematics and language interventions on the development of achievement (e.g., Gilraine & Penney, 2021).

Related to this, what currently seems to be missing are more fine-grained developmental theories that can be tested using the different models that are available. For instance, assumptions about the (im)plausibility of accumulating factors as inherent to LCS models or autoregressive latent trajectory models seem difficult to judge from a theoretical perspective. The same holds true for other features of these models, for instance, linear trend assumptions, which are a common feature of latent curve models with structured residuals. A comprehensive, applied literature that links theoretical models more strongly with these different types of "new" longitudinal models and methods has great potential to enhance our knowledge and understanding of reciprocal relations in developmental processes.

In the same vein, most of these models do not allow for an easy causal interpretation and rely on specific assumptions. From a more general perspective, when considering our results along with many of the previously reported results, for instance, from Erbeli et al. (2021) or Shin et al. (2013), the results have been somewhat coherent: Associations between prior reading and subsequent mathematics are substantial and positive, unlike the reverse path from prior mathematics to reading. These are important findings that have been replicated in different samples. Notably, before being able to judge the causality of these findings, and given the huge heterogeneity in available modeling options (see Table 1), it seems important to discuss the underlying theoretical model of reality more thoroughly in future studies. In our study, we decided to apply LCS models and LGC models because these models were well suited for addressing our RQs on the bidirectional development within students and the between-student differences in this within-student development (e.g., Grimm et al., 2017; Klopack & Wickrama, 2020). However, beyond the application of these models, it seems promising to more critically discuss (a) how the different models that are currently being applied to study longitudinal relations between the two constructs compare with one another, (b) which models are more or less reasonable choices for studying the co-development of the two constructs, given the models' assumptions, (c) which results different models produce, and, most importantly, (d) to what extent these or other models are helpful for drawing causal inferences. In addition to multiverse- or meta-analyses, future studies might also more thoroughly consider new weighting methods for continuous treatment variables (e.g., Fong et al., 2018; Hübner et al., 2022; Imai & Ratkovic, 2014) or marginal structural models (Robins et al., 2000) that focus specifically on modeling strategies, interpretations, and results to expand the understanding of longitudinal associations between reading and mathematics.

CONCLUSION

The present study investigated the co-development of reading and mathematics using a large set of data from students in England, Scotland, and Northern Ireland. Our findings from bivariate LGC models generally suggest that the development of the two domains is positively related. When examining associations more closely, we found that students with higher reading achievement in Grade 1 showed a steeper average growth in mathematics from T1 to T6 compared with students with lower reading achievement in Grade 1. On the other hand, higher achievement in mathematics at T1 was negatively associated with growth in reading. Using bivariate dual LCS models, we found a statistically significantly positive association between prior reading scores and subsequent changes in achievement in mathematics, whereas changes in reading were substantially smaller for students with a higher prior performance in mathematics. The findings suggest that acquiring good reading skills is highly relevant for developing mathematics skills.

FUNDING INFORMATION

This work was supported by the University of Durham and the Institutional Strategy of the University of Tübingen (ZUK 63, funded by the German Research Foundation 2012–2019) as well as by the Federal Ministry of Education and Research (BMBF) and the Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments (since 2019).

ORCID

Nicolas Hübner https://orcid.org/0000-0003-3528-8086
David Bolden https://orcid.org/0000-0002-2717-7266
Benjamin Nagengast https://orcid.org/0000-0001-9868-8322

REFERENCES

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, *31*(2–3), 162–172. https://doi.org/10.1016/j.stueduc.2005.05.008

Aiken, L. R. (1971). Verbal factors and mathematics learning: A review of research. *Journal for Research in Mathematics Education*, *2*(4), 304. https://doi.org/10.2307/748485

Andersen, H. K. (2021). Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve model with structured residuals and the random intercept cross-lagged panel model. *Psychological Methods*. https://doi.org/10.1037/met0000285

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

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328. https://doi.org/10.1080/19345740802400072

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Wiley-Interscience.

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

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook* (1st ed., pp. 135–154). Blackwell.

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

Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, *12*(1), 156–177. https://doi.org/10.1214/17-AOAS1101

Geary, D. C. (1994). Children's mathematical development: Research and practical applications. *American Psychological Association*. https://doi.org/10.1037/10163-000

Gilraine, M., & Penney, J. (2021). *Cramming: Short- and long-run effects*. EdWorkingPaper, 21-444. https://doi.org/10.26300/94pe-5j18

Grimm, K. J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, *33*(3), 410–426. https://doi.org/10.1080/87565640801982486

Grimm, K. J., Mazza, G. L., & Mazzocco, M. M. M. (2016). Advances in methods for assessing longitudinal change. *Educational Psychologist*, *51*(3–4), 342–353. https://doi.org/10.1080/00461520.2016.1208569

Grimm, K. J., Ram, N., & Estabrook, R. (2017). Growth modeling: Structural equation and multilevel modeling approaches. In *Methodology in the social sciences*. The Guilford Press.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hübner, N., Wagner, W., Zitzmann, S., & Nagengast, B. (2022). How causal is a reciprocal effect? Contrasting traditional and new methods to investigate the reciprocal effects model of self-concept and achievement. *PsyArXiv*. https://doi.org/10.31234/osf.io/f3e8w

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 243–263. https://doi.org/10.1111/rssb.12027

Klopack, E. T., & Wickrama, K. K. A. S. (2020). Modeling latent change score analysis and extensions in Mplus: A practical guide for researchers. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 97–110. https://doi.org/10.1080/10705511.2018.1562929

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J.-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal of Experimental Child Psychology*, *97*(3), 220–241. https://doi.org/10.1016/j.jecp.2007.03.001

Little, C. W., Lonigan, C. J., & Phillips, B. M. (2021). Differential patterns of growth in reading and math skills during elementary school. *Journal of Educational Psychology*, *113*(3), 462–476. https://doi.org/10.1037/edu0000635

Lüdtke, O., & Robitzsch, A. (2021). A critique of the random intercept cross-lagged panel model. *PsyArXiv*. https://doi.org/10.31234/osf.io/6f85c

Luyten, H., Merrell, C., & Tymms, P. (2017). The contribution of schooling to learning gains of pupils in years 1 to 6. *School Effectiveness and School Improvement*, *28*(3), 374–405. https://doi.org/10.1080/09243453.2017.1297312

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25 educational outcomes. *Early Childhood Research Quarterly*, *28*(2), 314–324. https://doi.org/10.1016/j.ecresq.2012.07.008

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychological Methods*, *25*(5), 610–635. https://doi.org/10.1037/met0000250

Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. *Journal of Computer Assisted Learning*, *23*(1), 27–35. https://doi.org/10.1111/j.1365-2729.2007.00196.x

Monroe, W. S., & Engelhart, M. D. A. (1931). *A critical summary of research relating to the teaching of arithmetic*. University of Illinois.

Moseley, D. (1976). *Personality and learning: Helping with learning difficulties*. Open University Press.

Mulder, J. D., & Hamaker, E. L. (2020). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *1–11*, 638–648. https://doi.org/10.1080/10705511.2020.1784738

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

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*(4), 1013–1034. https://doi.org/10.1037/pspp0000358

Pagani, L. S., & Fitzpatrick, C. (2014). Children's school readiness: Implications for eliminating future disparities in health and education. *Health Education & Behavior*, *41*(1), 25–33. https://doi.org/10.1177/1090198113478818

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

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Rhemtulla, M., & Tucker-Drob, E. M. (2011). Correlated longitudinal changes across linguistic, achievement, and psychomotor domains in early childhood: Evidence for a global dimension of development. *Developmental Science*, *14*(5), 1245–1254. https://doi.org/10.1111/j.1467-7687.2011.01071.x

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Scottish Government. (2019). *Pupil census 2019 supplementary tables*. https://www.gov.scot/publications/pupil-census-supplementary-statistics/

Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, *23*, 92–100. https://doi.org/10.1016/j.lindif.2012.10.002

Shonkoff, J. P., & Phillips, D. (2000). *From neurons to neighborhoods: The science of early child development*. National Academy Press. http://site.ebrary.com/lib/academiccompletetitles/home.action

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Spanoudis, G., & Demetriou, A. (2020). Mapping mind-brain development: Towards a comprehensive theory. *Journal of Intelligence*, *8*(2), 19. https://doi.org/10.3390/jintelligence8020019

Toogood, J., & Robinson, L. (2020). *Annual enrolments at schools and in funded pre-school education in Northern Ireland 2019–2020*. https://www.education-ni.gov.uk/sites/default/files/publications/education/revised%203rd%20March%202020%20-%20Annual%20enrolments%20at%20schools%20and%20in%20pre-school%20…pdf

UK Government. (2020). *Schools, pupils and their characteristics*. https://explore-education-statistics.service.gov.uk/data-tables/fast-track/af49fff7-b004-462b-97f4-7a337df581bc

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

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

Wrigley, J. (1958). The factorial nature of ability in elementary mathematics. *British Journal of Educational Psychology*, *28*(1), 61–78. https://doi.org/10.1111/j.2044-8279.1958.tb01426.x

SUPPORTING INFORMATION

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

How to cite this article: 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