Published in final edited form as:

Contemp Educ Psychol. 2017 October ; 51: 83–98. doi:10.1016/j.cedpsych.2017.06.008.

Bivariate Developmental Relations between Calculations and Word Problems: A Latent Change Approach

Jennifer K. Gilbert^a and Lynn S. Fuchs^a

^aVanderbilt University, 228 Peabody College, Nashville, TN 37203, USA

Abstract

The relation between 2 forms of mathematical cognition, calculations and word problems, was examined. Across grades 2–3, performance of 328 children (mean starting age 7.63 [$SD=0.43$]) was assessed 3 times. Comparison of *a priori* latent change score models indicated a dual change model, with consistently positive but slowing growth, described development in each domain better than a constant or proportional change model. The bivariate model including change models for both calculations and word problems indicated prior calculation performance and change were not predictors of subsequent word-problem change, and prior word-problem performance and change were not predictors of subsequent calculation change. Results were comparable for boys versus girls. The bivariate model, along with correlations among intercepts and slopes, suggest calculation and word-problem development are related, but through an external set of overlapping factors. Exploratory supplemental analyses corroborate findings and provide direction for future study.

Keywords

development; mathematical cognition; latent change score; bivariate relations

1. Introduction

Mathematics comprises a variety of related branches focused on the study of quantities as expressed in numbers or symbols. In the primary grades, the major curricular focus is whole numbers, usually conceptualized as understanding number, performing calculations (CALC), and solving word problems (WP); in the intermediate-middle grades, the focus shifts to rational numbers and algebraic thinking, and the high school curriculum includes algebra, geometry, trigonometry, and calculus (Common Core State Standards, 2010; Geary et al., 2008; National Mathematics Advisory Panel [NMAP], 2008). Yet, little is understood about how various aspects of mathematical cognition relate to: which aspects are shared or distinct and whether development in one domain influences development in others. Such

Inquiries should be sent to Jennifer K. Gilbert.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

understanding would provide insight into the nature of mathematics competence and guidance about how to optimize the design of curriculum and instruction.

The focus of the present study was connections between development of whole-number CALC and WP skill across grades 2–3. We focused on these grades because substantial progress occurs on multi-digit whole-number CALC and on WPs in this developmental period (Fuchs, Geary, et al., 2014; Geary, 2011). Both forms of early mathematical cognition are essential. Competence with whole-number CALC is foundational for success with every subsequent form of mathematics. WP solving, the best school-age predictor of adult employment and wages (Every Child a Chance Trust, 2009), represents a major emphasis within almost every strand of the mathematics curriculum throughout school. Yet, connections between these two essential forms of mathematical cognition are not well understood.

With typical development of CALC skill, children enter school with rudimentary understanding of number, insights into basic addition-subtraction concepts, and skill in counting objects to solve single-digit problems (Geary, 1994). In kindergarten and first grade, most children achieve a reliable, full set of associations among phonological, visual, and semantic representations of numerals, and addition-subtraction concepts and skilled use of counting for solving arithmetic problems improve (e.g., Groen & Resnick, 1977). Across grades 1–3, use of efficient counting strategies produces repeated associations of simple arithmetic problems with answers to secure representations in long-term memory (Fuson & Kwon, 1992; Siegler & Robinson, 1982; Siegler & Shrager, 1984). This permits direct retrieval, freeing attentional resources and providing the foundation for more complex CALC (Fuchs, Geary, et al., 2013; (LeFevre & Morris, 1999). Such multi-digit CALC involve knowledge of procedural steps that differ in sequence and complexity by problem type and rely on deliberate decision making, constrained by understanding of the numeration system and the ability to monitor strings of procedural actions (Fuchs, Geary, et al., 2014).

Less is known about how children develop facility with WPs. Moreover, prior work has not investigated how CALC and WP development is connected. What's known is that concurrent performance between domains is moderate to substantial (r s range from .35 to .71; e.g., Fuchs et al., 2008; Fuchs, Geary, et al., 2010a, 2010b; Seethaler et al., 2012). There are three explanations for this relation.

One possibility is that contextualizing CALC via stories supports understanding about CALC, and WP skill thus contributes to developing CALC skill. Some studies provide support for the hypothesis that contextualizing mathematics instruction with meaningful situations improves learning (Bottge, 1999; Cordova & Lepper, 1996; Maccini, Mulcahy, & Wilson, 2007), and this perspective is reflected in instruction designed to strengthen CALC in the primary grades (Bryant et al., 2011; Fuchs, et al., 2013; Fuchs, Powell, et al., 2014). Yet, the tenability of assumptions underlying this hypothesis have been questioned (Anderson, Reder, & Simon, 1996), and we identified no studies testing whether contextualizing CALC problems with stories facilitates CALC performance or whether improved WP skill is a leading indicator of subsequent CALC performance.

A second possibility for the moderate to strong relation between CALC and WP skill is that CALC skill is foundational for WP solving. Solving WPs relies on CALCs. However, whereas a CALC problem is set up for solution, a WP requires students to process a narrative to build a problem model and then construct number sentence(s) for calculating the unknown quantity. Levine, Jordan, and Huttenlocher (1992) provided some evidence that CALC skill precedes WP solving. Testers presented a set of objects, then moved the objects into a box and transformed this hidden set by visibly adding or removing objects. Children then constructed an array showing the number in the hidden set. Despite success with this CALC task at age 4, most children could not solve analogous WPs problems until age 5. This suggests that facility with CALC precedes WP solving and that CALC are foundational to and are a leading indicator of WP solving. This is consistent with longitudinal correlational work, where initial CALC skill predicts WP outcomes (Fuchs et al., 2006; Swanson & Beebe-Frankenberger, 2004).

It is nevertheless also possible that Levine et al.'s findings (1992) at age 4 reflect child-level variables necessary to support successful WP solving, not adequately developed until age 5. One likely candidate, which emerges in the literature as active in WP solving, is language comprehension (Fuchs et al., 2008; Fuchs, Geary, et al., 2010a, 2010b). Kintsch and Greeno (1985) described WP solving as demanding of language comprehension processes, and Cummins et al. (1988) provided corroborating data when they computationally modeled incorrect WP solving with incorrect math processes versus language comprehension processing errors. Results indicated that correct problem representation depended more on language comprehension; moreover, changing wording in minor ways dramatically affected accuracy. Some research identifies language comprehension as uniquely predictive of WP (not of CALC), while other external child-level factors, such as reasoning and working memory, are common across domains (Fuchs et al., 2010a; 2010b; Swanson, 2006). Common sources of variance in explaining individual differences across CALC and WP development speak to the third explanation for the moderate to strong relation between CALC and WP skill: CALC and WP development are related through an external set of overlapping factors.

The purpose of the present study was to deepen insight into the tenability of these explanations by directly investigating whether development in one domain influences the other. Based on Levine et al. (1992), along with studies indicating CALC skill predicts WP outcomes, our hypothesis was that CALC skill is foundational to WP solving (explanation #2). This hypothesis was, however, tentative because effects of WP solving on CALC development (explanation #1) have not been examined and because the literature reveals cognitive processes representing common external influences that may explain concurrent relations (explanation #3).

To consider these explanations, we assessed children's performance in both domains in fall and spring of second grade and spring of third grade. Relying on *a priori* latent change score models (McArdle, 2009), we first determined which of three models (constant change, proportional change, dual change) best describes development in each domain; we refer to these models as univariate models because only one outcome is considered at a time.

The constant change model, with a constant linear change term (slope), describes change as stable over time. In our specific case, this would mean that (on average) students grow at a steady rate over the course of grades 2 and 3 in terms of CALC and WP and that incremental change (during the next 6-month period as determined by our testing waves) is not dependent upon previous levels of performance.

By contrast, the proportional change model posits that change occurs incrementally, with change in one 6-month period proportional to performance at the preceding time. Substantively, this would mean that (on average) students' change in CALC and WP is related to their previous level of CALC and WP, respectively, with no constant or stable growth over the span of grades 2–3.

The dual change model describes development as a function of linear change and incremental change (see Table 1 for parameter/estimate interpretations). This would mean that change in CALC and WP has two influences, one that is constant or stable across the span of grades 2–3 and one that is dependent on (or proportional to) the previous level of performance. For instance, if the slope term and proportional change terms were both positive for CALC, that would mean that change in CALC skill is positive across grades 2–3 and that students who performed higher at a certain time point showed greater change in the subsequent period. In other words, change or growth is nonlinear, positive and accelerating.

Determining which model fits the data best is important for accurately describing development during this time period and may provide schools with guidance about the more efficient and efficacious timing of supplementary intervention for students struggling in these academic domains. It also is a necessary first step in determining bivariate developmental relations between CAL and WP.

Accordingly, after selecting the best model to describe development in each domain, we combined the univariate models into a bivariate change model describing the relation between both areas of development. A unique feature of the bivariate latent change score model is the exploration of time-sensitive developmental relations between domains, called coupling effects. Coupling effects detect whether development in one domain influences subsequent development in the other domain.

We considered five scenarios, the first involving no coupling between CALC and WP and the latter four involving various coupling effects: (1) correlated but uncoupled development, with no direct links between prior performance or change in one skill and subsequent change in the other (e.g., only the intercepts and slopes correlate, no other time-dependent relations exist); (2) prior CALC performance predicting subsequent change in WP (e.g., CALC level in fall of grade 2 predicting proportional change in WP from fall to spring of grade 2); (3) prior CALC change predicting subsequent change in WP (e.g., proportional change in CALC from fall to spring of grade 2 predicting proportional change in WP from spring of grade 2 to fall of grade 3); (4) prior WP performance predicting subsequent change in CALC (e.g., WP level in fall of grade 2 predicting proportional change in CALC from fall to spring of grade 2); and (5) prior WP change predicting subsequent change in CALC (e.g., proportional change in WP from fall to spring of grade 2 predicting proportional change in

CALC from spring of grade 2 to fall of grade 3). We used coefficients from the final model to construct a statistical field vector plot, used to visualize joint change trajectories.

As an exploratory supplemental analysis, we added fall of second-grade covariates to the best fitting model to test the robustness of findings and provide recommendations for future research. Covariates added to slopes and intercepts are fall of second-grade scores in word reading, listening comprehension, working memory, and nonverbal reasoning. We selected these domains as potentially overlapping factors because each has been shown in prior individual difference studies to predict development of CALC or WP or both (e.g., Geary & Widaman, 1992; Fuchs et al., 2006; Hitch, 1978; Fuchs et al., 2010). If significant coupling effects are detected, adding covariates would shed light on the extent to which our findings are robust when other factors are included in the prediction. If significant coupling effects are not detected, adding covariates may shed light on what other abilities are leading indicators of growth in these domains.

Finally, we conducted a within-sample robustness check on findings, as recommended in the literature (Duncan, Engel, Claessens, & Dowsett, 2014), by examining whether the pattern of effects was similar for boys versus girls. Although differences in the mathematics performance of boys versus girls may increase with grade level, with more complex mathematics that demands visual-spatial resources, and among higher-achievers (Stoet, Bailey, Moore, & Geary, 2016), we had no reason to expect differences in developmental patterns between relatively simple CALC and WP within a representative sample at this age, and we did not locate any relevant studies.

2. Method

2.1 Participants

Data for the present study are a subset from a parent study (Authors, 2014). The parent study was an efficacy trial using data from a single school year. Data from the parent study were also used in Authors (2014a), in which data across two school years were used in a mediation model to examine contributions of cognitive abilities and foundational mathematical competencies to numeration understanding. In the present study, we utilized data only from participants in the parent study who were in the business-as-usual condition. Further, the present study utilizes longitudinal data across two years to examine (co-)development in calculations and word problems, answering questions substantially different from either Authors (2014, 2014a).

The parent study was conducted in a U.S. southeastern school district of 81,000 students. The district encompasses a county of 668,000 residents from inner-city and surrounding suburban and rural communities, with middle-class, working-class, and high-poverty populations and 71% of students receiving subsidized lunch. The ethnic diversity of the district at the time data collection was 33% white non-Hispanic, 55% black, 10% white Hispanic, and 2.0% classified “other.” The county is home to the largest Kurdish community in the U.S. and large populations of Mexicans, Vietnamese, Cambodians, Laotians, Arabs, and Bantus, with smaller communities of Pashtuns from Pakistan and Afghanistan.

In the parent study, we had randomly sampled participants, stratifying by classroom and fall of second-grade mathematics performance, from middle- and low-income schools. Participants were selected from 1917 (85% of) children with parent consent in 127 second-grade classrooms taught by 96 teachers in 25 schools across four annual cohorts. For the present analysis, we excluded students who received instruction delivered by researchers in whole-class or small-group format during the parent study, because that instruction was designed to disturb the naturally occurring trajectory of CALC or WP development. Further, only the subsample of randomly selected students who were administered the CALC and WP measures was used in the present analysis.

The final sample for the present analysis was 328 students (age $M = 7.63$ [$SD = 0.43$]), from 33 second-grade classrooms in 17 schools, who dispersed to 153 third-grade classrooms in 46 schools. The large number of unique classroom and school sequences makes clustering irrelevant. In fall of second grade, standard scores for children in the present analysis were in the average range and representative of urban, diverse populations: $M = 93.52$ ($SD = 13.16$) on Wide Range Achievement Test-Arithmetic (WRAT-3; Wilkinson, 1993); $M = 103.34$ ($SD = 10.19$) on KeyMath-R Problem Solving (KM-R; Connolly, 1998); $M = 92.76$ ($SD = 12.24$) on Wechsler (1999) Abbreviated Scale of Intelligence. For children with complete demographics ($n = 309$), 84.59% received subsidized lunch; 57.05% were female; 40.66% were African American, 26.23% Caucasian, 25.90% Hispanic, 2.95% Kurdish, 4.26% other.

2.2 Measures

2.2.1 Outcomes—With WRAT-3, children write answers to CALC problems of increasing difficulty. At second and third grade for this sample, items largely address the four whole-number operations with and without regrouping with a smaller number of items addressing rounding and supplying a missing number in a sequence. In the parent study, α was .93. KM-R Problem Solving includes 18 WPs of increasing difficulty, involving all four operations and representing routine problems with and without transparent solution strategies. In addition to routine problems, other item types include word problem generation, supplying missing number in a sequence, word problem analysis, word problems containing irrelevant information, and currency knowledge and manipulation. Items are read aloud; responses are constructed. Testing is discontinued after three consecutive errors. In the parent study, α was .74.

2.2.2 Covariates

Wide Range Achievement Test 3 – Word Reading: (Wilkinson, 1993) progresses from letter identification to word identification. Children are asked to identify read 15 letters and read up to 55 words. No time limit is imposed. Each item is scored correct (1) or incorrect (0), and the total score is a simple sum. Testing is discontinued when a child earns 10 consecutive scores of 0. Cronbach's α as listed in the testing manual is .94.

Woodcock Diagnostic Reading Battery – Listening Comprehension: (Woodcock, 1997) measures the ability to understand sentences or passages. With 38 items, students supply the word missing at the end of sentences or passages that progress from simple verbal analogies and associations to discerning implications. Reliability is .80 at ages 5–18.

The Working Memory Test Battery for Children: (Pickering & Gathercole, 2001) – *Backward Digit Recall* subtest assesses central executive working memory. The subtest has six items at span levels from 2–7. For each item, the child repeats numbers spoken by the tester in the reverse order. Passing four items at a level moves the child to the next level. At each span level, the number of items to be remembered increases by one. Failing three items within a span terminates the subtest. (Reliability from the testing manual is not reported because of the small, unrepresentative sample.)

Wechsler Abbreviated Scale of Intelligence Matrix Reasoning: (Wechsler, 1999) measures nonverbal reasoning with pattern completion, classification, analogy, and serial reasoning tasks. For each of the 32 items, children selected one of five options that best completed a visual pattern. As reported in the testing manual, test-retest reliability for children ages 6–11 years is .76. The standard error of measurement reported for this instrument ranges from 2–4 for 6–11 year old children.

2.3 Procedure

Students were assessed in fall and spring of grade 2 and spring of grade 3. WRAT testing occurred in large groups in classrooms; all other testing occurred individually in other school locations. Testers were trained to criterion. Individual test sessions were audiotaped. A random sample was re-scored from tapes; agreement exceeded 97%. Data were double entered and verified.

3. Results

3.1 Scores

Each variable submitted to modeling was continuous and standardized using the mean and *SD* at the first time point. These *developmentally-scaled z-scores* are necessary for producing meaningful proportional change estimates. The latent score at time 3, although unmeasured, was included to scale measurement waves as equidistant. Means and *SDs* for CALC and WP at each measurement wave are shown in Table 2. Zero-order correlations among our outcomes of interest and potential external factors are shown in Table 4.

3.2 Missing values

Not all children in the sample had complete data at all test occasions, with 9.76% attrition by spring of grade 2 and additional 4.05% attrition by spring of grade 3. In spring of grade 2, children with and without complete data performed comparably on fall of second-grade CALC and WP ($ps > .05$). In spring of grade 3, these groups performed similarly on fall of second-grade CALC and WP ($ps > .05$) as well as spring of second-grade CALC and WP. To test differences in demographic variables, χ^2 tests were conducted on samples with and without complete data in the spring of grades 2 and 3. Among children with available demographic information, no significant differences were detected between children with and without complete data on subsidized lunch status, race, or sex. Further, variables representing missingness on CALC and WP variables were minimally correlated (rs ranged from $|.01|$ to $|.13|$) with the demographic variables as well as other CALC and WP scores. Thus, we consider our missing data “missing at random” and used maximum likelihood

estimation (with robust standard errors), which readily accommodates missing data by using all possible information in the submitted variables (i.e., cases contributing at least one outcome value are included). The MAR assumption was supported further in that maximum likelihood estimates of means, SDs, and correlations of the variables were nearly identical to the raw values which are based only on complete data (i.e., cases with incomplete data are dropped).

3.3 Sample size justification

As per Barker, Rancourt, and Jelalian's (2014), sample size requirements in longitudinal modeling depend on many factors such as magnitude of effect, measurement error, number of repeated measures, and study duration. In the present study, R^2 estimates from the final model output were .50–.79 for the six indicator variables, so latent variables accounted for a moderate to large portion of variance in the measures. In addition, assessments had sample alphas of .92 and .73, both acceptable (Bland & Altman, 1997). Three test occasions measurements is deemed the minimum but the 2-year study duration is theoretically satisfactory to detect meaningful changes. 100–200 participants are needed for estimating stable model estimates, and latent change score models with sample sizes equal to or smaller than the present study appear in peer-reviewed journals (e.g., $N = 316$, Ahmed, Wagner, & Lopez, 2014; $N = 118$, Barker et al.; $N = 253$, Ferrer & McArdle, 2003; $N = 316$, Quinn, Wagner, Petscher, & Lopez, 2015). So our sample of 328 appears sufficient for estimating unbiased parameters.

3.4 Data screening

Raw data for the outcome variables were checked for outliers (univariate and multivariate), distributional normality (univariate and multivariate), collinearity, outliers, and relative variances. Based on univariate distributions, two scores were omitted due to extreme (roughly 4 SD and 5 SD above the mean) and unusual (patterns discrepant from sample correlation) values. These scores were omitted to avoid biasing results. Mahalanobis distances were calculated to detect multivariate outliers. Three observations surpassed the critical χ^2 value; however, because there were no obvious remediations, they were retained. We checked multivariate outliers based on residuals from the most plausible model tested and none of the multivariate outliers was significant ($p < .001$). In distributional normality checks, no variables met cutoff values for extreme nonnormality. Still, we attempted a range of transformations to render the univariate distributions more normal, due to some skew or kurtosis; no single transformation in a construct over time corrected non-normality such that the null hypothesis of normal distribution was not rejected. Thus, we used the raw data for analyses, employing maximum likelihood estimation with robust standard errors to account for violations of nonnormality. In our final model, the scaling correction factor was 0.98. The closer the correction factor is to 1.00 the less the estimates have been corrected due to nonnormality. This would suggest that our data were not severely nonnormal and that the slight violations from normality observed in our variables are not likely to bias our findings. Using Stata's *collin* command, no evidence of extreme multicollinearity was detected; all variables had tolerance values $> .38$ (the cutoff for extreme collinearity is $< .10$). See Results for information on missing values. Finally, among the variables submitted for analysis, the ratio of largest to smallest variance was 5.4 which met the criterion of < 10 .

3.5 Model estimation

To test the connection between development of CALC and WP, we compared *a priori* latent change score models (McArdle, 2009) (Mplus 7.31; Muthén & Muthén, 1998–2015), first univariate then bivariate. We tested the generalization of our findings by testing bivariate models separately for boys versus girls. Finally, we conducted an exploratory supplemental analysis to assess whether fall of second-grade covariates affected the robustness of findings or provided guidance on future research. Alpha was set to .05 for all hypothesis testing. Because the MLR estimator was used, Satorra-Bentler (2001) scaled (SB) χ^2 tests were used to inform model selection decisions.

3.6 Univariate Models of Change

We assessed which of three models of change described each skill best. All models include a latent intercept, representing average performance in the fall of grade 2. The constant rate of change is represented by the α parameter whereas the proportional rate of change is represented by the β parameter. The constant change model includes a latent slope that represents a constant rate of change (α set to 1 with variance freely estimated; β set to 0). With proportional change, the rate of change depends on performance level at each prior time point (α set to 0 and β freely estimated). The dual change model includes constant linear change (α set to 1 with variance freely estimated) and proportional change (β freely estimated). Figure 1 depicts the dual change model. Again, setting $\alpha = 0$ produces the proportional change model; setting $\beta = 0$ produces the constant change model. Measurement invariance should be explored as part of model building. Grimm (2007) states that although measurement invariance is not necessary for model identification, researchers should consider why it is occurring. To satisfy measurement invariance when a single indicator represents a latent factor, the within-domain time-specific residual scores should be the same. Thus, residuals of indicator scores were estimated separately and fixed to equality in post-hoc manner as statistically necessary. Both restricted models were nested in and compared to the dual change model via Satorra-Bentler (2001) scaled-SB χ^2 tests.

The dual change model for CALC fit the data well (see Table 4 for model fit information): χ^2 exact-fit hypothesis was not rejected, $p = .258$; RMSEA = .03 was below the .05 threshold (although 90% CI = .00–.12); CFI and TLI each $> .99$; SRMR = 0.04 (with each cell in the normalized residual covariance matrix < 1.96). This model imposed equality constraints on residuals of the first two test occasions, which appeared optimal after comparing 3-, 2-, and 1- residual models. The larger residual variance at the latest time point may reflect the diverging performance of the student population as students have the opportunity to learn and complete more complex calculations over the course of schooling. Constraining proportional change parameter (β) to 0 (constant change model) produced significantly worse fit than the dual change model, SB-scaled $\chi^2(1) = 74.99$, $p < .001$. Constraining the constant change parameter (α) to 0 along with its variance and covariance with the intercept (proportional change model) also resulted in significantly worse fit, SB-scaled $\chi^2(3) = 726.40$, $p < .001$. Thus, the dual change model for CALC, indicating CALC development has a constant change component and a change component dependent on previous performance level, was retained for bivariate analyses.

The dual change model for WP fit the data well: χ^2 exact-fit hypothesis was not rejected, $p = .920$; RMSEA = .03 was below the .05 threshold (90% CI = .00 – .04); CFI and TLI were each = or > 1.00; SRMR = 0.01 (with each cell in the normalized residual covariance matrix < 1.96). This model imposed equality constraints on residuals of the last two measurement occasions; the smallest variance observed at the first time point likely reflects the curricular progression in which word problem solving has not yet been taught. Constraining the proportional change parameter (β) to 0 (constant change model) resulted in significantly worse fit than the dual change model, SB-scaled $\chi^2(1) = 48.74$, $p < .001$. Constraining the constant change parameter (α) and the variance and covariance with the intercept (proportional change model) to 0 resulted in significantly worse fit, SB-scaled $\chi^2(1) = 422.20$, $p < .001$. Thus, the dual change model for WP was retained for bivariate analyses.

3.7 Bivariate Models of Change

We combined the univariate dual change models into a bivariate model of change. The first bivariate model simply combined the two change models by allowing the intercepts and slopes between the domains to correlate. Performance or change in one domain was not allowed to predict subsequent change in the other. Thus, this model we refer to as the model of correlated but uncoupled development (see Figure 2). To this model, we compared four alternative models. In Figure 3, we test whether prior WP level predicts subsequent CALC change (beyond CALC slope). In Figure 4, we test whether prior WP change predicts subsequent CALC change (beyond CALC slope). In Figure 5, we test whether prior CALC level predicts subsequent WP change (beyond WP slope). In Figure 6, we test whether prior CALC change predicts subsequent WP change (beyond WP slope). For clarity and transparency, we have supplied Mplus syntax in the supplemental file for the model depicted in Figure 5. In each model, the coupling parameter, γ (see bolded lines), were set to equality. The exact fit hypothesis could not be rejected, $\chi^2(8) = 4.620$, $p = .797$. Indices of CFI, TLI, RMSEA and 90% CI, SRMR, and individual normalized residual values also indicated good fit (Kline, 2011). See SB-scaled χ^2 tests at bottom of Table 4. So, the uncoupled development model did not fit significantly worse than any of the four less restrictive model. Thus correlated but uncoupled development, which represents the most parsimonious and well-fitting model, indicates CALC and WP development are related but neither is a leading indicator (a significant predictor) of the other.

For results of the post-hoc sensitivity analysis on boys versus girls (as reported by teachers), see Table 5 for SB-scaled χ^2 tests. These models yielded the same conclusion as the full sample, thereby providing a within-sample robustness check that supports the generalizability of findings.

3.8 Specific effects in the most parsimonious and well-fitting model

Model-based estimates of parameters from the correlated but uncoupled bivariate dual change model are in Figure 7. In these models, due to the conversion of raw scores to developmentally-scaled z-scores, the intercept was set to 0, representing the average predicted score in fall of grade 2. Significant variability was detected in CALC level in fall of grade 2 (0.63) and CALC growth (0.18). Interpretation of the slope estimate in a dual change model differs from slope in a regular linear growth model; here, slope is the unique

effect beyond the proportional change parameter. After accounting for proportional change, the CALC slope ($p < .001$) indicates children improved, on average, 1.30 *SDs* per 6 months. The proportional change estimate of -0.37 along with positive slope indicates slowing of positive growth over time (higher scores at one time are associated with less change the next time). Model-estimated means of latent change scores declined: 1.30, 0.83, and 0.53 *SD* units, indicating the largest change in the middle of grade 2. The positive correlation between intercept and slope (.57) reveals children with initially higher CALC skill had higher CALC growth rates than children with lower beginning skill.

WP intercept and slope varied significantly (0.78 and 0.37). WP slope of 0.82 ($p < .001$) indicates improvement of 0.82 *SD* per 6 months. The negative proportional change estimate, (-0.47) along with positive slope signifies a decreasing growth rate. Model-estimated means of change scores declined: 0.82, 0.43, and 0.23 *SD* units, with the largest change in mid-grade 2. The correlation between intercept and slope was .69.

The correlations between CALC and WP intercepts (.61) and slopes (.57) were moderate. A follow-up test of mean slopes between outcomes using the MODEL TEST command indicated children grow faster in CALC (relative to fall of grade-2 CALC) than in WP (relative to fall of grade-2 WP), Wald $\chi^2(1) = 60.14$, $p < .001$. However, the proportional changes (deceleration) over time between the outcomes were not significantly different, Wald $\chi^2(1) = 3.29$, $p = .070$. Positive and moderate correlations between CALC intercept and WP slope and between WP intercept and CALC slope signify children who start with higher skill in one domain tend to grow faster in the other domain, compared to children who start with lower skill.

To graphically represent the joint estimated change trajectories defined by the final estimates of the correlated but uncoupled dual change model, we generated a statistical field vector plot (Figure 8). The tail of the arrow represents a specific starting value for WP and CALC (starting CALC and WP values at any time point during the study, given that our model assumes a consistent change process throughout the study period). The head of the arrow represents the expected values of WP and CALC after the subsequent time period. Thus, the length of the arrow signifies amount of expected change. For meaningful interpretation, interpret arrows that only fall within the range of actual scores, represented by circles.

The figure reveals that few students perform in the average or high range in one skill ($z > 2$) while performing poorly in the other skill ($z < -2$), which was expected from the .61 correlation of latent intercepts. Greatest expected change (longest arrows) within the range of actual data is in the lower left-hand quadrant, where beginning WP and CALC scores are simultaneously below average ($z < 0$). The least expected change occurs for combinations of scores with relatively high starting values. This relation between starting skill and amount of expected change is consistent with the negative proportional change estimates from the dual change score model. The plot features straight arrows, to signify the uncoupled nature of development of these two forms of mathematical cognition. In the presence of significant coupling, arrows would have been curved, suggesting a dynamic relation in which one domain pulls the other in a specific direction.

3.9 Exploratory supplemental analysis

To the correlated but uncoupled model, we added to the intercepts and slopes of CALC and WP four covariates, representing fall of second-grade skill in word reading, listening comprehension, working memory, and nonverbal reasoning. Our purpose was to provide insight on other factors driving CALC and WP development. The fit statistics for the exploratory supplemental model were acceptable. The exact fit hypothesis could not be rejected, $\chi^2(16) = 25.088$, $p = .068$. Indices of CFI, TLI, RMSEA and 90% CI, SRMR, and individual normalized residual values also were within the acceptable range as defined by Kline (2011). Results of this model are presented in Table 6. Two of the covariates, word reading and nonverbal reasoning, were significantly related ($ps < .01$) to all four growth parameters: CALC and WP intercepts and slopes. Listening comprehension was significantly related ($ps < .001$) to both the intercept and slope of WP but not to the intercept and slope of CALC. Finally, working memory was significantly related ($ps < .05$) to the WP parameters and the CALC slope ($p < .001$) but not to CALC intercept.

The path estimates of the best fitting model changed minimally even after the covariates were entered (see Figure 9), indicating that the findings regarding the correlated but uncoupled development held. The proportional change estimates were nearly identical, and the manifest variable residual variances differed by a maximum of 0.03. Expectedly, the variances and covariances of the latent intercepts and slopes did change in a negative direction due to the covariates explaining variability in the parameters. After accounting for the covariates, the correlation between CALC and WP intercepts dropped from .61 to .39 and their respective slopes fell from .57 to .49. These residual correlations suggest there are additional shared factors, beyond those considered here, which contribute to the development of CALC and WP. The pattern of relations among the covariates to the CALC and WP domains corroborates the correlated but uncoupled development model. That is, CALC and WP do not directly influence the development of one another; rather, an overlapping but separate set of factors underlies each skill domain.

4. Discussion

The present study corroborates prior work showing a moderate to strong concurrent relation between two essential forms of primary-grade mathematical cognition: CALCs and WPs. In fall of second grade, the correlation was .43; in spring of second grade, .50; in spring of third grade, .56. Beyond corroborating previous estimates of concurrent performance, results document moderate to strong correlations between CALC and WP intercepts (.61) and more interestingly between CALC and WP slopes (.57). Clearly, children with strong performance and improvement in one domain enjoy strong performance and growth in the other, and children who struggle tend to experience continued difficulty in both domains. We also extended prior work by documenting that development in each domain is best characterized by a dual change model, including a stable growth component and an incremental growth component (level at one point accounts for additional influence at the next point). Together, this indicates that across grades 2–3, children manifest positive but slowing growth in each domain.

Moreover, with raw scores converted to developmentally-scaled z -scores and intercept set to 0 for each domain in the dual change model, we compared developmental patterns for CALC versus WP. After accounting for proportional change, the mean CALC slope indicated 6-month increases of 1.30 SD s, with the proportional change revealing higher scores at one time are associated with less change at the next. The same was true for WP. Although decelerating proportional change was comparable, slopes were greater for CALC than WP, suggesting children learn more about CALC than WP in this timeframe. This may be due to a more ambitious CALC curriculum in grades 2–3 (which spans various forms of multi-digit operations), compared to WPs (where problems remain similar to the basic change, compare, and total schemas introduced at first grade). Alternatively, the slower pace of WP improvement may be due to teachers' more explicit instructional emphasis on CALC than WP, a point we return to later.

The major contribution of the present study, however, resides in deepening insight into the nature of the relation between CALC and WP performance and growth. We considered three possibilities: strength with WPs facilitates (or at least predicts) CALC development, or strength in CALCs is foundational to (or at least predictive of) WP development, or CALC and WP development are related through an external set of overlapping factors. Accordingly, we contrasted four models of coupled development (involving directional influence of prior performance or incremental change from each domain to change in the other) to a model of correlated but uncoupled development. None of the coupled development models improved fit, and this finding was robust for boys and girls. This bolsters support for the third explanation, in which CALCs and WPs have correlated intercepts and slopes, but CALC development is not a leading indicator of WP development, nor is WP development a leading indicator of CALC development.

In other words, it appears that outside variables, which somewhat overlap across CALC and WP, are the best explanation for correlated performance and incremental change in CALC and WP. Some variables external to CALCs and WPs may occur at the child level. For example, in prior work, individual differences in reading, reasoning, and working memory predict development in both domains (Fuchs et al., in press; Swanson, 2006), while language comprehension predicts development in WP but not CALC (Fuchs et al., 2010). As these skills and cognitive processes become more sophisticated, each may simultaneously bolster development of both forms of mathematical cognition. Other potentially active overlapping sources of influence are environmental: the quality of home or community interactions in supporting mathematical thinking (Sheldon & Epstein, 2005) and teachers' instructional practices.

Alternatively, as common sources of child-level variance develop, they may increase a child's capacity to respond to environmental influences on CALC and WP learning. Although reasoning and working memory are established predictors of each mathematics outcome, future research should contrast bivariate models of change while modeling potential contributions of child-level and environmental sources of variance in correlated performance and incremental change across the two domains.

In the meantime, present findings, based on direct examination of bivariate models, indicate CALC and WP skill represent distinct components of mathematical cognition. This conclusion is consistent with results from three other, less direct forms of study. First, correlational research documents important distinctions in some cognitive processes contributing to CALC and WP development (Fuchs et al., 2006, 2008; Fuchs et al., 2010; Swanson, 2006). For example, Fuchs, Geary, Compton, Fuchs, Hamlett, and Bryant (2010) found that for CALC, only working memory explained additional variance beyond basic number knowledge; by contrast, language comprehension, attentive behavior, reasoning, and working memory offered additional explanatory value for WP. Second, WPs can be a persistent deficit even when CALC skill is adequate (Fuchs et al., 2008; Swanson, Jerman, & Zheng, 2008). Third, in a large randomized control trial (Fuchs, Powell, et al., 2014), students who received research-based CALC instruction experienced stronger CALC outcomes than students who received research-based WP intervention, but these effects did not transfer to WPs. Parallel results occurred for WP intervention.

Understanding that CALC and WP may represent distinct components of mathematical cognition has implications for instructional practice. Teaching students to perform accurate CALC is more straightforward than teaching successful WP solving. Accordingly, research and classroom instruction in the U.S. has focused disproportionately on CALC over WP. Recent shifts in curricular standards attempt to reallocate focus to contextualized problem solving (CCSS, 2010). Yet, the most widely used approach, as reflected in textbooks, involves a general metacognitive framework for WP solving (i.e., reading problem, underline important information, draw the problem, plan solution, predict and estimate answer, compute answer, check answer), when evidence for this approach is limited (Babakhani, 2011; Montague, 1992, 2008).

So, as reflected in textbooks and classroom practice, the dominant perspective is that teaching explicit strategies for conceptualizing and solving WP, shown to be effective in research (Fuchs et al., 2004; Jitendra et al., 2011), is not required. Such thinking is based on the assumption that competence with CALC will transfer to WP solving. Our findings suggest otherwise. Explicit WP instruction may be required to meet the needs of students with limitations in the cognitive processes associated with CALC and WP development or whose opportunities for mathematical enrichment are compromised. Before closing, we note that this study has important limitations.

4.1 Limitations

Before closing, we note that this study has important limitations. As with most structural equation models, the number of potential models exceeds the number feasible to test. Thus, we chose a set of *a priori* models aimed at answering our specific questions and selected our final model as the most parsimonious model that fit the data. Further, more frequent data collection would have permitted inclusion of additional growth parameters or relaxation of model constraints. To extend beyond the present study, researchers should consider alternative modeling procedures and more frequent data collection. Further, latent variables comprising multiple measures of each domain would provide a more robust picture of the relations among CALC and WP that would be less influenced by the idiosyncrasies of any

one test and possibly result in more powerful statistical tests, due to the reduction in measurement error. Finally, to adequately address the external factors driving development in CALC and WP, we would need to assess the external factors repeatedly (at the same time intervals we assessed CAL and WP) rather than measuring the external factors only at the start of the second grade. Future research with longitudinal measurements of important underlying skills is necessary to uncover leading indicators of CALC and WP development.

Acknowledgments

This research was supported by Award Number R01HD053714, R24HD075443, and Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

References

- Authors. 2014
- Authors. 2014a
- Ahmed Y, Wagner RK, Lopez D. Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of Educational Psychology*. 2014; 106:419–434. DOI: 10.1037/a0035692 [PubMed: 24954951]
- Anderson JR, Reder LM, Simon HA. Situated learning and education. *Educational Researcher*. 1996; 25(4):5–11. DOI: 10.3102/0013189X025004005
- Babakhani N. The effect of teaching the cognitive and meta-cognitive strategies (self-instruction procedure) on verbal math problem-solving performance of primary school students with verbal problem-solving difficulties. *Procedia Social and Behavioral Sciences*. 2011; 15:563–570.
- Barker DH, Rancourt D, Jelalian E. Flexible models of change: Using structural equations to match statistical and theoretical models of multiple change processes. *Journal of Pediatric Psychology*. 2014; 39:233–245. DOI: 10.1093/jpepsy/jst082 [PubMed: 24190912]
- Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *British Medical Journal*. 1997; 314(7080): 572. [PubMed: 9055718]
- Bottge B. Effects of contextualized math instruction on problem solving of average and below-average achieving students. *Journal of Special Education*. 1999; 33:81–92.
- Bryant DP, Bryant BR, Roberts G, Vaughn S, Pfannenstiel KH, Porterfield J, Gersten R. Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*. 2011; 78:7–23.
- Common Core State Standards Initiative. National Governors Association Center for Best Practices. Washington, DC: Council of Chief State School Officers; 2010.
- Connolly, AJ. KeyMath – Revised. Circle Pines, MN: American Guidance Service; 1998.
- Cordova DI, Lepper MR. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*. 1996; 88:715–730. <http://dx.doi.org/10.1037/0022-0663.88.4.715>.
- Cummins DD, Kintsch W, Reusser K, Weimer R. The role of understanding in solving word problems. *Cognitive Psychology*. 1988; 20:405–438.
- Duncan GJ, Engel M, Claessens A, Dowsett CJ. Replication and robustness in developmental research. *Developmental Psychology*. 2014; 50:2417–2425. DOI: 10.1037/a0037996 [PubMed: 25243330]
- Every Child a Chance Trust. The long-term costs of numeracy difficulties. 2009. Retrieved. August 14, 2009, <http://www.everychildachancetrust.org/counts/index.cfm>
- Ferrer E, McArdle J. Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. 2003; 10:493–524. DOI: 10.1207/S15328007SEM1004_1

- Fuchs LS, Fuchs D, Compton DL, Powell SR, Seethaler PM, Capizzi AM, Fletcher JM. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*. 2006; 98:29–43. DOI: 10.1037/0022-0663.98.1.29
- Fuchs LS, Fuchs D, Prentice K, Hamlett CL, Finelli R, Courey SJ. Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology*. 2004; 96:635–647. DOI: 10.1037/0022-0663.96.4.635
- Fuchs LS, Fuchs D, Stuebing K, Fletcher JM, Hamlett CL, Lambert WE. Problem-solving and computation skills: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*. 2008; 100:30–47. DOI: 10.1037/0022-0663.100.1.30 [PubMed: 20057912]
- Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Bryant JV. The contributions of numerosity and domain-general abilities to school readiness. *Child Development*. 2010a; 81:1520–1533. DOI: 10.1111/j.1467-8624.2010.01489.x [PubMed: 20840238]
- Fuchs LS, Geary DC, Compton DL, Fuchs D, Schatschneider C, Hamlett CL, Changas P. Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology*. 2013; 105:58–77. DOI: 10.1037/a0030127 [PubMed: 24065865]
- Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Bryant JV. The contributions of numerosity and domain-general abilities to school readiness. *Child Development*. 2010a; 81:1520–1533. DOI: 10.1111/j.1467-8624.2010.01489.x [PubMed: 20840238]
- Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Seethaler PM, Bryant JV, Schatschneider C. Do different types of school mathematics development depend on different constellations of numerical and general cognitive abilities? *Developmental Psychology*. 2010b; 46:1731–1746. DOI: 10.1037/a0020662 [PubMed: 20822213]
- Fuchs LS, Geary DC, Fuchs D, Compton DL, Hamlett CL. Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology*. 2014; 106:482–498. DOI: 10.1037/a0034444 [PubMed: 25284885]
- Fuchs LS, Gilbert JK, Powell SR, Cirino PT, Fuchs D, Hamlett CL, Seethaler PM, Tolar TM. The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus pre-algebraic knowledge. *Developmental Psychology*. 2016; 52:2085–2098. [PubMed: 27786534]
- Fuchs LS, Powell SR, Cirino PT, Schumacher RF, Marrin S, Hamlett CL, Changas PC. Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? *Journal of Educational Psychology*. 2014; 106:990–1006. <http://dx.doi.org/10.1037/a0036793>. [PubMed: 25541565]
- Fuson KC, Kwon Y. Korean children's single-digit addition and subtraction: Numbers structured by ten. *Journal for Research in Mathematics Education*. 1992; 23:148–165.
- Geary, DC. *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association; 1994.
- Geary, DC., Boykin, AW., Embretson, S., Reyna, V., Siegler, R., Berch, DB., Graban, J. The Final Report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education; 2008. Report of the task group on learning processes. <http://www.ed.gov/about/bdscomm/list/mathpanel/report/learning-processes.pdf>
- Geary DC, Widaman KF. Numerical cognition: On the convergence of componential and psychometric models. *Intelligence*. 1991; 16:47–80. DOI: 10.1016/0160-2896(92)90025-M
- Grimm KJ. Multivariate longitudinal methods for studying developmental relationships between depression and academic achievement. *International Journal of Behavioral Development*. 2007; 31:328–339. DOI: 10.1177/0165025407077754
- Groen G, Resnick LB. Can preschool children invent addition algorithms? *Journal of Educational Psychology*. 1977; 69:645–652.
- Hitch GJ. The role of short-term working memory in mental arithmetic. *Cognitive Psychology*. 1978; 10:302–323. DOI: 10.1016/0010-0285(78)90002-6
- Jitendra AK, Star JR, Rodriguez M, Lindell M, Someki F. Improving students' proportional thinking using schema-based instruction. *Learning and Instruction*. 2011; 21:731–745. DOI: 10.1016/j.learninstruc.2011.04.002

- Kline, RB. Principles and practice of structural equation modeling. 3. New York: Guilford; 2011.
- Kintsch W, Greeno JG. Understanding and solving word arithmetic problems. *Psychological Review*. 1985; 92:109–129. [PubMed: 3983303]
- Levine SC, Jordan NC, Huttenlocher J. Development of calculation abilities in young children. *Journal of Experimental Child Psychology*. 1992; 53:72–103. [PubMed: 1545190]
- Maccini P, Mulcahy CA, Wilson MG. A follow-up of mathematics interventions for secondary students with learning disabilities. *Learning Disabilities Research & Practice*. 2007; 22:58–74. DOI: 10.1111/j.1540-5826.2007.00231.x
- McArdle JJ. Latent variable modeling of longitudinal data. *Annual Review of Psychology*. 2009; 60:577–605. DOI: 10.1146/annurev.psych.60.110707.163612
- Montague M. The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities*. 1992; 25:320–248. DOI: 10.1177/002221949202500404
- Montague M. Self-regulation strategies to improve mathematical problem -solving for students with learning disabilities. *Learning Disability Quarterly*. 2008; 3:37–44.
- Muthén, LK., Muthén, BO. Mplus user's guide (version 7.31). Los Angeles, CA: Muthén & Muthén; 1998–2015.
- National Mathematics Advisory Panel. Foundations for Success: Final Report of the National Mathematics Advisory Panel. Washington, DC: United States Department of Education; 2008. <http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- Pickering, S., Gathercole, S. Working Memory Test Battery for Children. London: The Psychological Corporation; 2001.
- Quinn JM, Wagner RK, Petscher Y, Lopez D. Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*. 2015; 86:159–175. DOI: 10.1111/cdev.12292 [PubMed: 25201552]
- Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001; 66:507–514. Retrieved from <http://link.springer.com/article/10.1007/BF02296192>.
- Seethaler PM, Fuchs LS, Fuchs D, Compton DL. Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology*. 2012; 104:224–234. <http://dx.doi.org/10.1037/a0024968>. [PubMed: 22347725]
- Sheldon SB, Epstein JL. Involvement counts: Family and community partnerships and mathematics achievement. *Journal of Educational Research*. 2005; 98(4):196–206.
- Siegler, RS., Robinson, M. The development of numerical understandings. In: Reese, H., Lipsitt, LP, editors. *Advances in child development and behavior*. Vol. 16. New York: Academic; 1982. p. 241–312.
- Siegler, RS., Shrager, J. Strategy choice in addition and subtraction: How do children know what to do?. In: Sophian, C., editor. *Origins of cognitive skills*. Hillsdale, NJ: Erlbaum; 1984. p. 229–293.
- Stoet G, Bailey DH, Moore AM, Geary DC. Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *PLOS One*. 2016; doi: 10.1371/journal.pone.0153857
- Swanson HL. Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology*. 2006; 98:265–281. DOI: 10.1037/0022-0663.98.2.265
- Swanson HL, Beebe-Frankenberger M. The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious mathematics difficulties. *Journal of Educational Psychology*. 2004; 96:471–491. DOI: 10.1037/0022-0663.96.3.471
- Swanson HL, Jerman O, Zheng X. Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*. 2008; 100:343–379. DOI: 10.1037/0022-0663.100.2.343
- Wechsler, D. Wechsler AS2reviated Scale of Intelligence. San Antonio, TX: Psychological Corporation; 1999.
- Wilkinson, GS. Wide Range Achievement Test. Vol. 3. Wilmington, DE: Wide Range; 1993.
- Woodcock, RW. Woodcock Diagnostic Reading Battery. Itasca, IL: Riverside; 1997.

Highlights

- Development of skill in calculations and word problems across grades 2–3 is positive but slowing.
- At each time point, the greatest expected change was associated with the lowest beginning scores.
- Development between domains is correlated but distinct, signifying no automatic transfer.
- An overlapping but separate set of child-level factors underlies each skill domain.

Figure 1.

Path analysis diagram for univariate dual growth model. I = intercept. L = level. S = slope. β = autoregressive, or proportional change, path predicting change as a function of the previous level of performance. Δ = delta, change score. α = loading for latent slope factor which was set to 1 (i.e., constant change). Paths without a parameter were set to 1 for model identification. To properly scale time, L3 was added as a place holder for the fall of grade 3, in which assessments were not administered.

Figure 2.

Bivariate latent change score model depicting correlated but uncoupled development between calculation and word problems. I = intercept. L = level. S = slope. β = autoregressive, or proportional change, path predicting change as a function of the previous level of performance. Δ = delta, change score. α = loading for latent slope factor which was set to 1 (i.e., constant change). Paths without a parameter were set to 1 for model identification. Curved lines represent correlations. To properly scale time, L3 was added as a placeholder for the fall of grade 3, in which assessments were not administered.

Figure 3.

Bivariate latent change score model depicting coupling between word problem level and subsequent proportional calculation change (bolded lines). I = intercept. L = level. S = slope. β = autoregressive, or proportional change, path predicting change as a function of the previous level of performance. Δ = delta, change score. α = loading for latent slope factor which was set to 1 (i.e., constant change). Paths without a parameter were set to 1 for model identification. Curved lines represent correlations. To properly scale time, L3 was added as a placeholder for the fall of grade 3, in which assessments were not administered.

Contemp Educ Psychol. Author manuscript; available in PMC 2018 October 01.

Figure 5.

Bivariate latent change score model depicting coupling between calculation level and subsequent proportional word problem change (bolded lines). I = intercept. L = level. S = slope. β = autoregressive, or proportional change, path predicting change as a function of the previous level of performance. Δ = delta, change score. α = loading for latent slope factor which was set to 1 (i.e., constant change). Paths without a parameter were set to 1 for model identification. Curved lines represent correlations. To properly scale time, L3 was added as a placeholder for the fall of grade 3, in which assessments were not administered.

Figure 6.

Bivariate latent change score model depicting coupling between calculation proportional change and subsequent proportional word problem change (bolded lines). I = intercept. L = level. S = slope. β = autoregressive, or proportional change, path predicting change as a function of the previous level of performance. Δ = delta, change score. α = loading for latent slope factor which was set to 1 (i.e., constant change). Paths without a parameter were set to 1 for model identification. Curved lines represent correlations. To properly scale time, L3 was added as a place holder for the fall of grade 3, in which assessments were not administered.

Figure 7.

Final correlated but uncoupled development model of calculation (CALC) and word-problem (WP) change. Paths without estimates were fixed to 1 for model identification. Path values are estimates in within-construct developmentally-scaled z-scored units, all $p < .001$. Curved arrows between variables represent correlations. For intercept and slope factors, the first number is the mean; the second, variance.

Figure 8.

Statistical vector field plot illustrating the joint estimated change trajectories of WP (solving word problems) and CALC (performing calculations). Circles represent actual data points. Arrows indicate expected change. The tail of the arrow represents a specific starting value for WP and CALC (at any time point) and the head of the arrow represents the expected values of WP and CALC after the next time period. Thus, the length of the arrow signifies amount of expected change, which varies according to different starting values.

Figure 9.

Final correlated but uncoupled development model of calculation (CALC) and word-problem (WP) change with covariates. Covariates and corresponding paths to intercepts and slopes are not pictured here for clarity sake but estimate values are listed Table 6. Paths without estimates were fixed to 1 for model identification. Path values are estimates in within-construct developmentally-scaled z -scored units, all $p < .001$. Curved arrows between variables represent correlations. For intercept and slope factors, the first number is the mean; the second, variance.

Table 1

Summary of Parameter and Estimate Interpretations

Parameter/Estimate	Interpretation
σ_Y^2	Residual variance of time-specific Y scores
S_Y	Residual of time-specific Y scores
Y	Observed indicators of latent Y performance
L1–L4	Latent true scores
σ_I^2 , intercept variance	Variability among students in predicted fall of grade 2 performance
I, intercept mean	Predicted level of performance at the fall of grade 2
β	Autoregressive/proportional change path predicting latent change as a function of the previous level of performance (above and beyond stable change)
1– 3, latent change scores	Latent change between testing waves influenced by a constant/stable component and the previous latent score
$\sigma_{I,S}$, intercept and slope covariance	Relation between fall performance and rate of growth
σ_S^2 , slope variance	Variability among students in rates of growth across grades 2 and 3
S, slope mean	Stable, constant change over the course of the study (above and beyond the proportional change); rate of growth
α	Loading for the latent slope factor, set to 1 for model identification

Table 2

Means and Standard Deviations of Outcomes (N = 328)

Variable	n	Raw				Standard			
		M	SD	M	SD	M	SD	M	SD
Fall of grade 2									
Calculation skill	324	19.12	2.55	93.52	13.16				
Word-problem solving skill	325	2.29	1.56	103.34	10.19				
Spring of grade 2									
Calculation skill	296	22.47	2.48	98.41	11.85				
Word-problem solving skill	296	3.60	1.91	105.57	10.63				
Spring of grade 3									
Calculation skill	283	25.88	3.64	98.34	14.54				
Word-problem solving skill	284	4.62	2.21	101.12	10.94				

Calculation skill is Wide Range Achievement Test-3 Arithmetic (Wilkinson, 1993). Word-problem solving skill is KeyMath-R Problem Solving (Connolly, 1998).

Table 3

Correlations of External Factors and Outcomes (N = 328)

Variable	1	2	3	4	5	6	7	8	9	10
<i>External factors, fall of grade 2</i>										
1. Word reading	—									
2. Listening comprehension	.35	—								
3. Working memory	.34	.29	—							
4. Nonverbal reasoning	.25	.15	.29	—						
<i>Outcomes</i>										
5. Calculation skill, fall grade 2	.38	.18	.32	.27	—					
6. Calculation skill, spring grade 2	.42	.19	.33	.35	.58	—				
7. Calculation skill, spring grade 3	.38	.13	.24	.32	.42	.55	—			
8. Word-problem solving skill, fall grade 2	.52	.44	.37	.34	.43	.49	.37	—		
9. Word-problem solving skill, spring grade 2	.38	.54	.35	.36	.43	.51	.41	.63	—	
10. Word-problem solving skill, spring grade 3	.43	.44	.34	.32	.47	.50	.56	.55	.67	—

Word reading is Wide Range Achievement Test-3 Word Reading (WRAT-3; Wilkinson, 1993). Listening comprehension is Woodcock Diagnostic Reading Battery Listening Comprehension (Woodcock, 1997). Working memory is Working Memory Test Battery for Children Backward Digit Recall (Pickering & Gathercole, 2001). Nonverbal reasoning is Wechsler Abbreviated Scale of Intelligence Matrix Reasoning (Wechsler, 1999). Calculation skill is WRAT-3 Arithmetic (Wilkinson, 1993). Word-problem solving skill is KeyMath-R Problem Solving (Connolly, 1998).

Table 4

Model Fit Indices and Results of Satorra-Bentler Scaled χ^2 Difference Tests

Model	$\chi^2(df)$	<i>p</i>	CFI/TLI	RMSEA(90% CI)	SRMR	SB-Scaled χ^2 Difference vs. Dual Change (<i>df</i>)	<i>p</i>
Univariate CALC							
Dual Change	2.714(2)	.258	0.997/0.995	.033(.000–.120)	0.040		
Constant Change	75.854(3)	<.001	0.691/0.691	.272(.221–.327)	0.165	74.986(1)	<.001
Proportional Change	683.098(5)	<.001	0.000/–0.725	.643(.603–.684)	1.408	726.396(3)	<.001
Univariate WP							
Dual Change	0.167(2)	.920	1.000/1.009	.000(.000–.039)	0.009		
Constant Change	49.973(3)	<.001	0.847/0.847	.218(.168–.274)	0.093	48.742(1)	<.001
Proportional Change	344.789(5)	<.001	0.000/0.334	.455(.415–.497)	0.581	422.203(3)	<.001
	$\chi^2(df)$	<i>p</i>	CFI/TLI	RMSEA(90% CI)	SRMR	SB-Scaled χ^2 Difference vs. Uncoupled (<i>df</i>)	<i>p</i>
Bivariate CALC & WP							
Correlated but Uncoupled	4.620(8)	.797	1.000/1.009	.000(.000–.042)	0.023		
CALC level→WP change	4.108(7)	.767	1.000/1.008	.000(.000–.047)	0.023	0.511(1)	.475
CALC change→WP change	4.108(7)	.767	1.000/1.008	.000(.000–.047)	0.023	0.511(1)	.475
WP level→CALC change	1.446(7)	.984	1.000/1.016	.000(.000–.000)	0.011	2.728(1)	.098
WP change→CALC change	1.446(7)	.984	1.000/1.016	.000(.000–.000)	0.011	2.728(1)	.098

CFI=comparative fit index. TLI=Tucker-Lewis Index; RMSEA=root-mean-square error of approximation; SRMR=standardized root-mean-square residual; SB=Satorra-Bentler; WP=word-problem skill, KeyMath-R Problem Solving (Connolly, 1998); CALC=calculation skill, Wide Range Achievement Test-3 Arithmetic (Wilkinson, 1993).

Table 5

Results of Alternate Bivariate Models vs. Correlated but Uncoupled Model by Sex

	Female Model (<i>n</i> = 174)			Male Model (<i>n</i> = 131)		
	SB Scaled χ^2 Difference Test (vs. Uncoupled)	<i>df</i>	<i>p</i>	SB Scaled χ^2 Difference Test (vs. Uncoupled)	<i>df</i>	<i>p</i>
CALC level → WP change	0.831	1	.362	0.014	1	.906
CALC change → WP change	0.832	1	.362	0.013	1	.909
WP level → CALC change	1.634	1	.201	0.255	1	.614
WP change → CALC change	1.634	1	.201	0.254	1	.614

SB = Satorra-Bentler; WP = word-problem solving skill, KeyMath-R Problem Solving (Connolly, 1998); CALC = calculation skill, Wide Range Achievement Test-3 Arithmetic (Wilkinson, 1993).

Table 6

Results of Covariate Effects on Intercepts and Slopes

Covariate	CALC Intercept		CALC Slope		WP Intercept		WP Slope	
	path	p	path	p	path	p	path	p
Word reading	0.268	<.001	0.143	<.001	0.285	<.001	0.112	.004
Listening comprehension	0.059	.287	-.0006	.846	0.300	<.001	0.261	<.001
Working memory	0.176	<.001	0.041	.243	0.151	.001	0.086	.028
Nonverbal reasoning	0.166	.001	0.146	<.001	0.205	<.001	0.155	.001
R ²	.340		.318		.516		.459	

CALC = calculation skill, Wide Range Achievement Test-3 Arithmetic (Wilkinson, 1993); WP = word-problem skill, KeyMath-R Problem Solving (Connolly, 1998); Word reading = Wide Range Achievement Test-3 Reading (Wilkinson, 1993); Listening comprehension = Woodcock Diagnostic Reading Battery (Woodcock, 1997); Working memory = Working Memory Test Battery for Children Backward Digit Recall (Pickering & Gathercole, 2001); Nonverbal reasoning = Wechsler Abbreviated Scale of Intelligence Matrix Reasoning (Psychological Corporation, 1999).