Child Development, in press

Pathways to Third-Grade Calculation versus Word-Reading Competence: Are They More Alike or Different?

Supplemental File

Lynn S. Fuchs[1], David C. Geary[2], Douglas Fuchs[1], Donald L. Compton[1], and Carol L. Hamlett[1]

[1]Vanderbilt University and [2]University of Missouri

Method

Participants

This study was conducted in a U.S. southeastern school district of 81,000 students. The district encompasses a county of 668,000 residents living in inner-city and surrounding suburban and rural communities. The county's population is diverse socio-economically, with middle-class, working-class, and high-poverty populations. The county is home to large populations of Mexicans, Kurds, Vietnamese, Cambodians, Laotians, Arabs, and Bantus, with smaller communities of Pashtuns from Pakistan and Afghanistan, and the county has the largest community of Kurdish people in the U.S.

At the time data-collection occurred, the district's students were 33% white non-Hispanic, 55% black, 10% white Hispanic, 2.0% other; 71% of students in the schools received subsidized lunch. From a large number of schools in this district representing the full socio-cultural population, we randomly sampled participants, while stratifying on classroom and start-of-first-grade mathematics performance. The sample, recruited in four annual cohorts, comprised 973 children, with no more than eight children from the same class. Data for the present study came from three measurement occasions: the start of grade 1 and the end of grades 2 and 3. Of the 973 children, 226 moved beyond the study's reach before the end of third grade. The remaining 747 children were in 228 first-grade classrooms in 37 schools. They dispersed to 258 second-grade classrooms in 73 schools and then to 452 third-grade classrooms across 86 schools (creating too many unique classroom and school sequences to make clustering relevant). The 226 students who did not complete the study were not significantly different on any first-grade predictor from the 747 students who did complete the study. Data were complete for the 747 students.

At the start of first grade, the mean age of these children (49% girls) was 6.52 years (*SD* = 0.37 months); 76% received subsidized lunch; 8% received English as a Second Language services; 42% came from single-parent homes; 60% were African American, 30% non-Hispanic white, 8% Hispanic white, and 2% other. The mean IQ on the 2-subtest *Wechsler Abbreviated Scale of Intelligence* (WASI; Wechsler, 1999) was 93.65 (*SD* = 12.28). At the start of first grade, scores on the *Arithmetic* and *Reading* subtests of the *Wide Range Achievement Test* (WRAT-3; Wilkinson, 1993) averaged 97.93 (*SD* = 14.34) and 99.98 (*SD* = 14.96), respectively.

First-Grade Predictor Measures

Attentive behavior. The *Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior* (SWAN; Swanson et al., 2004) samples items from the *Diagnostic and Statistical Manual of Mental Disorders-IV* criteria for Attention Deficit Hyperactivity Disorder for inattention (9 items). Teachers rate students on a 1-7 scale for each of the 9 items. On this sample, α was .97.

Reasoning. Nonverbal reasoning was assessed with WASI-*Matrix Reasoning* (Wechsler, 1999), using all 4 problem types. With each item, children induce the missing section in a matrix by choosing among five response options. As reported in the test manual, reliability is .89 - .94.

Listening comprehension. With *Woodcock Diagnostic Reading Battery* (WDRB)-*Listening Comprehension* (Woodcock, 1997), children supply the word missing at the end of sentences or passages that progress from simple verbal analogies and associations to discerning implications. As reported in the test manual, reliability is .80.

Central executive working memory. From the *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001), *Counting Recall* from the central executive battery. Children count a set of 4, 5, 6, or 7 dots on a card; after counting a series of cards, they recall the number of counted dots of each card. The subtest has six items at each span level from 1-6 to 1-9. Passing four items at a level moves the child to the next level. At each level, the number of items to be remembered increases by one. Failing three items terminates the subtest. We used the trials correct score, as these are more reliable than the span scores (Geary et al., 2010. As per the test developer, test-retest *r* is .80.

Phonological short-term memory. From the WMTB-C (Pickering & Gathercole, 2001), we used *Nonword List Recall*. The format/structure is the same as Counting Recall, except the child simply repeats stimuli spoken by the tester in the same order. Note that use of nonwords forces reliance on temporary representation in the phonological store (whereas use of real words would permit recall of stored knowledge). As per the test developer, test-retest *r* is .83.

Visuospatial short-term memory. From the WMTB-C (Pickering & Gathercole, 2001), we used *Mazes Memory*. The format/structure is the same as Counting Recall, but the nature of the task differs. The tester presents a maze with more than one solution and a picture of an identical maze with a path showing one solution. The picture is removed, and the child duplicates the path in the response booklet. At each level, the maze get larger by one wall. As per the test developer, test-retest *r* is .80.

Processing speed. With *WJ-III Visual Matching* (Woodcock, McGrew, & Mather, 2001), children locate and circle two identical numbers in rows of six numbers. They have 3 min to complete 60 rows. As per the test developer, reliability is .91.

RAN. Following Denkla and Rudel (1976), the tester presents 5 letters or numerals with which the child completes three practice trials. Then the tester presents a 5 X 10 matrix of the same letters or numerals, which the child names as quickly as possible without making mistakes. RT is measured with a stopwatch. We combined RTs across letters and numerals (r = .69); the reliability of the composite, estimated with the Spearman-Brown prediction formula was high (ρ = .82). We multiplied scores by -1 so higher values indicate faster processing.

Counting knowledge. Following Geary (2011), the tester introduces the child to a puppet that is learning to count and needs feedback on counting accuracy. During each trial, an array of 7, 9, or 11 chips of alternating color is placed behind a screen. For each trial, the screen is removed; the puppet counts the chips; and the child states whether the puppet counted correctly. There are four types of trials: correct (chips counted sequentially and correctly from the child's left to right), right-left (chips counted sequentially and correctly from right to left), pseudo-error (chips counted correctly from left to right; first one color is counted and then the other), and error (chips counted sequentially from left to right; the first chip is counted twice). Each trial type occurs once for each array size, with one additional pseudo-error as the last trial. The score is the number of trials correctly identified as counted correctly or not (15% of the sample achieved the maximum score). On this sample, α was .91.

Early numerical competencies. WRAT-Arithmetic (Wilkinson, 1993) comprises an oral and a written component. The 15-item oral portion includes counting objects (3 items), identifying Arabic numerals (5 items), holding up a specified number of fingers (2 items), naming the larger of two numbers (2 items), and answering simple word problems (3 items). The written component provides students with 10 min to answer on calculation problems of increasing difficulty (grades K-12). All students finished working in < 10 min. In this sample, the mean score at start of first grade was 15.25 (SD = 3.12), which most students achieved by scoring 13 on the oral portion and correctly answering 2 problems on the written portion. The first four items on the written portion are simple adding or subtracting problems (1 + 1, 5 – 1, 2 + 7, 8 – 4), also a form of early numerical competencies. With a *SD* of 3, 68% of participants were within a score of 12 and 18. On this sample, α was .89.

Early reading-related competencies. WRAT-Reading (Wilkinson, 1993) is administered orally, without a time limit. With the first 15 items, students name capital letters presented in random order. Letter naming is a reading-related competency that strongly predicts later reading achievement (Lambrecht et al., 2008). Then students read words of increasing difficulty. In this sample, the mean score at start of first grade was 19.98 (SD = 4.81), which most students achieved by correctly naming the 15 letters and reading the first five simple, high frequency words (*in*, *cat*, *book*, *tree*, *how*). With a *SD* of 4, 68% of participants were within a score of 16 and 24. On this sample, α was .91.

End-of-Second-Grade Measure Involved in Indirect Effects

Addition retrieval. With *Addition Strategy Assessment* (Geary et al., 2007), 14 simple addition problems are presented horizontally, one at a time at the center of a computer screen. Addends are 2 - 9; doubles problems are not used. Half the problems sum to 10 or less; half have the smaller addend in the first position. The child solves each problem without paper as “quickly as possible without making too many mistakes” using whatever strategy is easiest to get the answer (no time limit was imposed). After solving each problem, the tester asks how the child derived the answer. This combined with the tester's observations is used to classify each trial as finger counting, verbal counting, decomposition, or retrieval (i.e., correct problem for which the child speaks answer quickly without any indication of counting or decomposition; includes child report of guessing or "I just knew it"). These methods have proven useful for measuring strategy choice (Geary, 1990; Siegler, 1987). Validity is supported at second grade where the mean RT for retrieval responses is 2,789 ms (*SD* = 1,892; some quicker; some longer); 4,152 ms (*SD* = 2784) for decomposition; 4,980 ms (*SD* = 3,928) for verbal counting; and 6,662 ms (*SD* = 4,153) for finger counting (all pairwise comparisons significant). In the present study, we found the same relation between strategy and RT. We used was the percentage of correct problems for which the child retrieved answers.

Third-Grade Outcomes

We used the written portion of WRAT-Arithmetic (Wilkinson, 1993) to index third-grade calculation skill (no child required administration of the individual portion to meet the basal). All students finished work in < 10 min. On this sample, α was .94. We used the word-reading portion of WRAT-Reading to index third-grade word-reading skill (no child required administration of the letter naming segment to meet the basal). On this sample, α was .93.

Procedure

Testers were trained to criterion at each testing occasion. They were unfamiliar to the children they tested on all measurement occasions and did not otherwise interact with children they assessed. In fall of first grade, children were assessed on the predictors (WASI-Matrix Reasoning, WDRB-Listening Comprehension, WMTB-C subtests, WJ-III Visual Matching, WRAT-Arithmetic and Reading), and classroom teachers completed ratings of inattentive behavior. In spring of second grade, children were assessed on the Addition Strategy

Assessment; in the spring of third grade, on WRAT–Arithmetic and Reading. All sessions were audiotaped; 15% of tapes were selected randomly, stratifying by tester, for accuracy checks by an independent scorer. Agreement exceeded 99%.

Results

See Table S1 for means and *SD*s (raw scores and standard scores for nationally normed tests) as well as *r*s among measures. Raw scores, transformed to sample-based *z*-scores, were simultaneously entered into analyses that examined total effect, the direct effect, and the indirect effect of each start-of-first-grade predictor. The predictors were attentive behavior, reasoning, central executive working memory, language comprehension, phonological memory, visuospatial memory, processing speed, RAN, counting knowledge, early math skill, and early word-reading skill. With these 11 predictors, we ran two models: one estimating the total, direct, and indirect effect of each predictor on third-grade calculations; the other estimating these effects on third-grade word reading. To examine indirect effects, we assessed the effect of the predictors on second-grade retrieval (*a* paths), and the effect of retrieval on each of the third-grade outcomes (*b* paths).

We used the Preacher and Hayes (2008) SPSS mediate macro to obtain estimates, using path analytic mediation routine with bootstrapping (5000 draws to estimate standard errors) applied to construct 95% confidence intervals for indirect effects. In line with current thinking (Hayes, 2009), we interpret significant indirect effects (the product of the *a* and *b* paths [$a*b$]) even when the total, direct, *a* path, or *b* path is not significant. Table S2 shows path coefficients for the model predicting each outcome.

In Figure S1, shaded boxes signify significant total effects, solid arrows show significant direct effects, and dotted lines indicate significant indirect effects. The top panel shows significant paths for effects on third-grade calculations. The total effect (direct plus indirect) was significant for attentive behavior, reasoning, central executive, early math, and early reading. The omnibus test of the total effect on third-grade calculations was significant, $R^2 = .44$, $F(11,735) = 53.39$, $p < .001$. Direct effects were significant for attentive behavior, reasoning, working memory, early math, and early reading. The path *b* effect for retrieval on the calculation outcome was significant. The path *a* effect on retrieval was significant for attentive behavior, reasoning, central executive, visuospatial short-term memory, RAN, early math, and early reading, and the indirect effect via retrieval was significant for each of these predictors.

The bottom panel of Figure S1 shows significant paths for effects on third-grade word reading. The total effect was significant for language comprehension, phonological memory, RAN, and word reading. The omnibus test of the total effect on third-grade word reading was significant, $R^2 = .55$, $F(11,735) = 85.28$, $p < .001$. Direct effects were significant for these same predictors (language comprehension, phonological memory, RAN, word reading). The path *b* effect for retrieval on the word-reading outcome was significant. The path *a* effect on retrieval was significant for attentive behavior, reasoning, central executive, visuospatial short-term memory, RAN, early math, and early reading, with significant indirect effects via retrieval for each of these predictors.

To help readers with interpretation of coefficients and the magnitude of their effects, we operationalize the direct and indirect effect for the word-reading on the calculation outcome. The significant direct effect coefficient indicates that for every increase in early reading competency of one unit, later calculation skill improves by .14. The significant indirect effect for early reading on later calculation skill via retrieval is the product of paths *a* and *b*, in which every unit improvement in early reading is associated with an increase of .10 in second-grade retrieval (path *a*) and every unit increase in retrieval raises third-grade calculations by .20 (path *b*). This produces an indirect effect of .02 (.10*.20). So the total (direct + indirect) effect for early reading competency on the calculation outcome is .16. Together, this suggests that (a) improving early reading competencies is directly associated with development of retrieval and later calculation skill and (b) early reading's effect on third-grade calculations is further enhanced when second-grade addition retrieval is strengthened.

References

Denkla, M.B., & Rudell, R. (1976). Rapid automatized naming (RAN: Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*, 471-479.

Geary, D.C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, *49*, 363-383. doi:10.1016/0022-0965(90)90065-G

Geary, D.C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539-1552. doi:10.137/a0025510

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1459.

Geary, D.C., Hoard, M. K., & Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *Journal of Experimental Child Psychology, 113*, 49–65.

Hayes, A.F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408-420.

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children.* London: The Psychological Corporation.

Preacher, K.J., & Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891. [Supplementary material].

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250-264.

Swanson, J. et al. (2004). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales. Downloaded from www.adhd.net on 12/20/2004.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wilkinson, G.S. (1993). *Wide Range Achievement Test 3*. Wilmington: Wide Range.

Woodcock, R.W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside.

Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Table S1

Means, Standard Deviations, and Correlations (n=747)

	Raw Scores			Standard Scores		Correlations												
						Predictors											Mediator	Outcome
	Mean	(*SD*)	Max[a]	Mean	(*SD*)	A	R	W	L	PL	V	PS	RN	CT	EM	ER	RT	C3
Predictors																		
Attentive Behavior (A)	37.35	(12.98)	63	NA														
Reasoning (R)	7.88	(5.10)	32	47.73	(8.89)	.38												
Central Executive (W)	10.61	(4.53)	42	NA		.42	.38											
Language (L)	13.54	(4.98)	38	86.40	(15.58)	.44	.40	.36										
Phonological Mem (P)	15.24	(9.76)	36	NA		.37	.38	.38	.43									
Visuospatial Mem (V)	4.39	(3.68)	42	NA		.26	.30	.33	.23	.26								
Processing Speed (PS)	8.20	(2.94)	60	92.66	(19.07)	.45	.44	.42	.39	.39	.39							
RAN (RN)	71.96	(30.89)	NA	NA		.42	.23	.37	.30	.30	.25	.33						
Counting (CT)	9.30	(2.27)	13	NA		.24	.27	.27	.28	.21	.15	.24	.17					
Early Math (EM)	15.25	(3.12)	55	97.93	(14.34)	.55	.52	.49	.49	.48	.36	.46	.46	.28				
Early Reading (ER)	19.23	(4.81)	57	99.98	(14.96)	.54	.50	.37	.47	.42	.27	.44	.50	.22	.61			
Mediator																		
Retrieval (RT)	32.10	(26.43)	100	NA		.22	.24	.24	.17	.17	.22	.21	.20	.15	.34	.24		
Outcomes																		
Calculations	25.85	(4.25)	55	97.93	(15.52)	.52	.43	.43	.40	.37	.29	.41	.39	.23	.58	.54	.36	
Word Reading	31.15	(4.85)	57	99.97	(13.41)	.48	.42	.37	.45	.42	.24	.38	.50	.16	.52	.71	.25	.54

For all *r*s, *p*<.001. Attentive behavior is Attentive Behavior from Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior (Swanson et al., 2004). Reasoning is Matrix Reasoning from Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Working memory is Counting Recall from Working Memory Test Battery–Children (WMTB; Pickering & Gathercole, 2001). Language is Listening Comprehension from Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1999). Phonological memory is Nonword List Recall from WMTB. Visuospatial short-term memory is Mazes Memory from WMTB. Processing speed is Visual Matching from Woodcock-Johnson III (Woodcock et al., 2001). RAN is rapid automatized naming- digits and letters, following Denkla and Rudel (1976); we multiplied scores by -1. Counting is Counting Knowledge, following Geary (2011). Early math and calculations outcome are the Wide Range Achievement Test-Arithmetic (WRAT; Wilkinson, 1993). Early reading and word-reading outcome are WRAT-Reading. Standard scores are mean=100, *SD* = 15, except Reasoning (mean=50, *SD* = 10).

[a] Max is maximum possible score.

Table S2

Effects of First-Grade Predictors on Third-Grade Calculation and Word-Reading Outcomes, with Effects of Second-Grade Retrieval in the Model (n=747)

Outcome/Predictor	Total Effect on Outcome Path *c* (*SE*)	Direct Effect on Outcome Path *c'* (*SE*)	Effect on Retrieval Path *a* (*SE*)	Effect of Retrieval on Outcome Path *b* (*SE*)	Indirect Effect on Outcome Path (CI)
Outcome: Calculations				**.20 (.03)[c]**	
Attentive Behavior	**.20 (.04)[c]**	**.18 (.04)[c]**	**.11 (.04)[c]**		**.02 (.043 .005)**
Reasoning	**.08 (.04)[a]**	**.06 (.03)[a]**	**.12 (.04)[a]**		**.02 (.007 .043)**
Central Executive	**.11 (.03)[b]**	**.10 (.04)[b]**	.06 (.04)		.01 (-.030 .004)
Language	.03 (.03)	.03 (.03)	-.01 (.04)		.00 (-.019 .015)
Phono Mem	.02 (.03)	.01 (.03)	.01 (.04)		.00 (-.014 .019)
Visuospa Mem	.03 (.03)	.03 (.03)	**.09 (.04)[a]**		**.02 (.030 .035)**
Processing Speed	.03 (.03)	.02 (.03)	.02 (.04)		.00 (-.012 .021)
RAN	.05 (.03)	.04 (.03)	**.10 (.04)[a]**		**.02 (.037 .003)**
Counting	.01 (.03)	.02 (.03)	-.02 (.04)		-.00 (-.019 .010)
Early Math	**.23 (.04)[c]**	**.18 (.04)[c]**	**.23 (.05)[c]**		**.05 (.004 .073)**
Early Reading	**.16 (.04)[c]**	**.14 (.04)[c]**	**.12 (.05)[a]**		**.02 (.004 .045)**
Outcome: Word Reading				**.06 (.03)[a]**	
Attentive Behavior	.05 (.03)	.04 (.03)	**.11 (.04)[c]**		**.01 (.020 .0004)**
Reasoning	.04 (.03)	.04 (.03)	**.12 (.04)[a]**		**.01 (.001 .017)**
Working Memory	.04 (.03)	.04 (.03)	.06 (.04)		-.00 (-.012 .001)
Language	**.10 (.03)[b]**	**.10 (.03)[b]**	-.01 (.04)		-.00 (-.007 .005)
Phono Mem	**.08 (.03)[b]**	**.08 (.03)[b]**	.01 (.04)		.00 (-.005 .007)
Visuospa Mem	-.02 (.03)	-.02 (.03)	**.09 (.04)[a]**		**.01 (.0002 .010)**
Processing Speed	-.02 (.03)	-.02 (.03)	.02 (.04)		.00 (-.004 .008)
RAN	**.17 (.03)[c]**	**.16 (.03)[c]**	**.10 (.04)[a]**		**.01 (.010 .0002)**
Counting	.05 (.03)	.05 (.03)	-.02 (.04)		-.00 (-.007 .004)
Early Math	.02 (.04)	.00 (.04)	**.23 (.05)[c]**		**.01 (.0003 .018)**
Early Reading	**.50 (.04)[c]**	**.49 (.04)[c]**	**.12 (.05)[a]**		**.01 (.002 .030)**

[a]p<.05; [b]p<.01; [c]p<.001. Bolded coefficients are statistically significant.

Attentive behavior is Attentive Behavior from Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior (Swanson et al., 2004). Reasoning is Matrix Reasoning from Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Central executive is Counting Recall from the Working Memory Test Battery–Children (WMTB; Pickering & Gathercole, 2001). Language is Listening Comprehension from Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1999). Phonological memory is Nonword List Recall from WMTB. Visuospatial short-term memory is Mazes Memory from WMTB. Processing speed is Visual Matching from Woodcock-Johnson III (Woodcock et al., 2001). RAN is rapid automatized naming- digits and letters, following Denkla and Rudel (1976); we multiplied scores by -1. Counting is Knowledge of Counting, following Geary (2011). Early math and calculations outcome are the Wide Range Achievement Test-Arithmetic (WRAT; Wilkinson, 1993). Early reading and word-reading outcome are WRAT-Reading.

Figure S1. Direct effects (solid arrows) and indirect effects (dotted arrows) on the calculation outcome (top panel) and on the word-reading outcome (bottom panel). Shading refers to total effects (direct + indirect effects).