Journal of Experimental Child Psychology 113 (2012) 78–92

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades

Drew H. Bailey, Andrew Littlefield, David C. Geary *

Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 7 September 2011
Revised 25 April 2012
Available online 15 June 2012

Keywords:
Arithmetic
Memory retrieval
Skill
Sex differences
Mathematics achievement
Practice

ABSTRACT

The ability to retrieve basic arithmetic facts from long-term memory contributes to individual and perhaps sex differences in mathematics achievement. The current study tracked the codevelopment of preference for using retrieval over other strategies to solve single-digit addition problems, independent of accuracy, and skilled use of retrieval (i.e., accuracy and reaction time [RT]) from first to sixth grades inclusive ($N = 311$). Accurate retrieval in first grade was related to working memory capacity and intelligence, and it predicted a preference for retrieval in second grade. In later grades, the relation between skill and preference changed such that preference in one grade predicted accuracy and RT in the next grade as RT and accuracy continued to predict future gains in preference. In comparison with girls, boys had a consistent preference for retrieval over other strategies and had faster retrieval speeds, but the sex difference in retrieval accuracy varied across grades. Results indicate that ability influences early skilled retrieval, but both practice and skill influence each other in a feedback loop later in development and provide insights into the source of the sex difference in problem-solving approaches.

Introduction

The strategies children use to solve addition problems have been well characterized and range from finger counting to retrieval of answers from long-term memory (Carpenter & Moser, 1984; Geary,

* Corresponding author.
E-mail address: gearyd@missouri.edu (D.C. Geary).

http://dx.doi.org/10.1016/j.jecp.2012.04.014

Brown, & Samaranayake, 1991; Siegler & Shrager, 1984). Fast and accurate retrieval has been identified as an important outcome for children by the National Mathematics Advisory Panel (2008) and the National Council of Teachers of Mathematics (2006), and indeed the more basic facts that can be correctly retrieved, the higher the performance on paper-and-pencil arithmetical ability and mathematics achievement tests for children and adults (Geary, Bow-Thomas, Liu, & Siegler, 1996; Geary & Widaman, 1987; Siegler, 1988).

Identification of the mechanisms underlying the development of children's ability to retrieve facts, thus, is an important educational and scientific goal. Children's use of counting strategies appears to contribute to the formation of the long-term memory representations needed to support fact retrieval and, thereby, is one such mechanism (Siegler, 1987; Siegler & Shrager, 1984), but this does not appear to be the only one. Brain imaging and cognitive studies indicate that retrieval of basic addition facts engages the prefrontal cortex, requires attentional control, and is a more active and effortful process for children compared with adults (Cho, Ryali, Geary, & Menon, 2011; Geary, Hoard, Nugent, & Bailey, 2012; Rivera, Reiss, Eckert, & Menon, 2005).

The implication is that individual differences in the capacity of the central executive component of working memory and perhaps other domain-general abilities contribute to individual differences in children's early ability to correctly retrieve basic facts. Siegler (1988) found, however, that some children who were capable of skilled retrieval did not always use retrieval but rather relied on counting or other procedures for problem solving, whereas other children used retrieval even when their answers were likely to be incorrect. In other words, preference for the use of one problem-solving approach over another can influence strategy choices independent of skill and domain-general abilities.

The relation between preference for retrieval and skilled use of retrieval is not well understood but can be framed with two separate but not mutually exclusive hypotheses. In general, the talent hypothesis for the codevelopment of preference and skill is that early talent at some activity influences one's preference for that activity in the future. In the context of the codevelopment of skill at and preference for retrieval as a strategy in single-digit addition, the talent hypothesis is that individual differences in working memory or other general cognitive abilities influence early skilled retrieval as well as grade-to-grade increases in skill. Skilled retrieval in turn contributes to a later preference for retrieval over other strategies.

The second hypothesis that might explain the codevelopment of preference and skill is the practice hypothesis. In general, an early preference for some activity causes individuals to practice that activity more, thereby increasing their future skill at that activity. In the context of the codevelopment of skill at and preference for retrieval as a strategy in single-digit addition, the practice hypothesis is that a preference for retrieval contributes to growth in skilled retrieval. In this case, the developmental emergence of retrieval skill is preceded by an earlier preference for use of retrieval over other strategies independent of working memory or other domain-general competencies. It is likely that the practice and talent hypotheses both are important for explaining the codevelopment of skill and preference in many domains. This is because if skill can influence practice, and practice can influence skill, a feedback loop may arise where individuals who prefer to participate in some activity will practice it more, which increases their skill, which in turn may further increase their preference, and so on.

The question of the developmental relation between preference for and skilled use of retrieval is also important because there are early emerging sex differences on these dimensions (Carr & Davis, 2001), and there is some evidence that skilled retrieval, which is higher in males, results in a sex difference on mathematics tests in which basic arithmetic is embedded (Royer, Tronsky, Chan, Jackson, & Marchant, 1999). By analyzing the cross-lagged relation between preference and skill from first to sixth grades inclusive, we were able to test these hypotheses. We demonstrate that the direction of the relation between skilled retrieval and preference for use of retrieval varies across grades and that the pattern of relations differs for boys and girls.

Addition development

By the time children start first grade, most of them have developed a suite of strategies that can be used to solve formal addition problems, for example, "How much is 3 + 2?" (Carpenter & Moser, 1984; Geary, 1990; Siegler, 1987). Most of these children will use counting to solve at least some of the problems, sometimes using their fingers (finger counting strategy) and sometimes not using them (verbal counting strategy). Whether or not children use their fingers, the min and sum procedures are the two most common ways in which children count (Groen & Parkman, 1972; Siegler & Shrager, 1984). The min procedure involves stating the larger valued addend and then counting a number of times equal to the value of the smaller addend. The sum procedure involves counting both addends starting from 1. The less common max procedure involves stating the smaller addend and counting the larger one.

The use of counting appears to result in the development of memory representations of basic facts (Siegler & Shrager, 1984), which eventually support the use of memory-based processes for problem solving. The most common are direct retrieval and decomposition. With direct retrieval, children state an answer that is associated with the presented problem in long-term memory, such as stating "eight" when asked to solve 5 + 3. Decomposition involves reconstructing the answer based on the retrieval of a partial sum; for example, 6 + 7 might be solved by retrieving the answer to 6 + 4 and then adding 3.

Across the elementary school years, basic change reflects a decreased use of counting procedures and increased use of retrieval-based processes (Ashcraft, 1982), but this is not simply a switch from the use of less sophisticated to more sophisticated strategies. Rather, at any time, children can use one of many strategies to solve a given problem. They may retrieve the answer to 3 + 1 but count to solve 5 + 8. What changes is the mix of strategies, with sophisticated ones used more often and less sophisticated ones used less often (Siegler, 1996).

Domain-general abilities

General cognitive mechanisms have been found to influence rate and ease of learning across academic domains. The current study focused specifically on the central executive component of working memory and intelligence (Carroll, 1993; Gottfredson, 1997). Intelligence is the best individual predictor of achievement across academic domains, including mathematics (e.g., Deary, Strand, Smith, & Fernandes, 2007; Stevenson, Parker, Wilkinson, Hegion, & Fish, 1976; Walberg, 1984). Working memory may contribute to performance on intelligence tests, but at the core is the ability to think logically and systematically and to learn novel information and concepts (Cattell, 1963; Embretson, 1995).

At a very basic level, the core component of working memory is the central executive, which is expressed as attention-driven control of information represented in two passive memory systems: the phonological loop and the visuospatial sketchpad (Baddeley, 1986; Baddeley & Hitch, 1974; Cowan, 1995). It has been well established that the higher the capacity of the central executive, the better the performance on measures of mathematics achievement and cognition (Bull, Espy, & Wiebe, 2008; DeStefano & LeFevre, 2004; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Klein & Bisanz, 2000; Mazzocco & Kover, 2007; Noël, 2009; Passolunghi, Vercelloni, & Schadee, 2007; Swanson & Sachse-Lee, 2001). Individual differences in the phonological loop and visuospatial sketchpad are related to some aspects of mathematical performance and learning (De Smedt et al., 2009; Krajewski & Schneider, 2009; Swanson, Jerman, & Zheng, 2008) but do not appear to be as critical to developmental change in the use of retrieval-based processes as the central executive (Geary, Hoard, & Nugent, in press).

Sex differences

Sex differences in arithmetic tend to be small, but one aspect in which boys and girls consistently differ is the mix of strategies used to solve addition problems. First-grade boys, for instance, use retrieval-based strategies with greater frequency and accuracy than girls; that is, as a group, boys show greater preference for and more skilled use of retrieval than girls (Carr & Alexeev, 2011; Carr & Davis, 2001; Carr & Jessup, 1997; Geary et al., 1996). Royer and colleagues (1999) proposed that sex differences in the speed of arithmetic fact retrieval underlie sex differences in mathematics achievement in difficult subject areas both by allowing boys and men to finish tests faster (difficult mathematics problems require individuals to solve a large number of embedded arithmetic problems) and by automatizing much of the problem-solving process, which frees working memory for other aspects of problem solving (Geary & Widaman, 1992).

However, the mechanisms underlying these sex differences are unclear and are unlikely to be related to the central executive. If anything, young girls' advantage over same-age boys in effortful (Allan & Lonigan, 2011) and executive control (Wiebe, Espy, & Charak, 2008) should result in more skilled retrieval in girls than in boys, but this is not the case (Carr & Davis, 2001; Carr & Jessup, 1997; Geary et al., 1996; Royer et al., 1999). The issue of whether or not there are sex differences in intelligence is vigorously debated (Colom, Juan-Espinosa, Abad, & García, 2000; Lynn & Irwing, 2004), but even theorists arguing for such a difference do not believe it exists before 15 years of age (Lynn & Irwing, 2004). This does not mean that intelligence can be ruled out as a contributor to individual differences in skilled retrieval, but it does suggest that intelligence is an unlikely mediator of the early sex differences. Another possibility is that boys' preference for retrieval is attributable to sex differences in intellectual risk taking (Byrnes, Miller, & Schafer, 1999), with boys willing to risk the use of an underdeveloped strategy in order to quickly solve the problem as a form of social display (Geary, 2010; Peterson & Fennema, 1985).

In other words, if boys are more likely to want to quickly answer an arithmetic problem as a way of competing with their classmates, they will be biased toward a preference for retrieval because it can be executed more quickly than counting strategies (Siegler, 1987). To the extent that the practice hypothesis is correct, boys' early preference for retrieval will contribute to a later advantage over same-grade girls in skilled retrieval (Royer et al., 1999).

The current study

Our primary goal was to test the practice and talent hypotheses using cross-lagged patterns across six measurements (first to sixth grades inclusive) of children's single-digit addition strategy choices. The focus was on retrieval because this is a desired educational goal (National Council of Teachers of Mathematics, 2006; National Mathematics Advisory Panel, 2008) and because of the just-noted sex differences. If the practice hypothesis is correct, high levels of preference for retrieval (i.e., frequent use independent of accuracy) in earlier grades should predict increases in retrieval skill in subsequent grades. If the talent hypothesis is correct, retrieval skill in early grades should be related to working memory or intelligence and early skilled retrieval should predict later skilled retrieval as well as a preference for retrieval over other strategies.

Method

Participants

All kindergarteners from 12 elementary schools that serve children from a wide range of socioeconomic backgrounds were invited to participate. Parental consent and child assent were received for 37% ($N = 311$) of these children, and 288 of them completed the first year of testing, 249 remained in the study through the first 5 years (i.e., end of fourth grade), and 229 continued through fifth and sixth grades. All available data were used, with missing data estimated in Mplus Version 6.0 (Muthén & Muthén, 1998–2010) using full information maximum likelihood.

At the end of first grade, the mean intelligence and variability of the sample were average ($M = 99$, $SD = 15$) based on the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). The mean age was 75 months ($SD = 4$) at the time of the first assessment (spring, kindergarten) and 141 months ($SD = 4$) at the time of the last assessment (fall, sixth grade). Girls comprised 53% of the total sample, and 70% of the sample was White; the remaining children were Black (12%), Asian (4%), or more than one race (7%), with the parents of the remaining children identifying them as Native American, Pacific Islander, or "unknown." Across racial categories, 4% of the sample identified as ethnically Hispanic. The mathematics curriculum when the children began the study was *Investigations in Number, Data, and Space* (Scott Foresman, 1999), and they continued with this curriculum through fifth grade. We did not have information on parental socioeconomic status, but we did have information on the percentage of children eligible for free or reduced price lunches at the 12 schools from which the initial sample was drawn, where 34% of children attending these schools were eligible.

Addition strategy choices

Fourteen simple addition problems were presented horizontally, one at a time, at the center of a computer monitor; the problems were presented on cards in first grade and on the monitor thereafter. The problems consisted of the integers 2 through 9, with the constraint that the same two integers (e.g., 2 + 2) were never used in the same problem. Half of the problems summed to 10 or less, and the smaller valued addend appeared in the first position for half of the problems.

Each child was asked to solve each problem (without pencil and paper) as quickly as possible without making too many mistakes to balance speed and accuracy. It was emphasized that the child could use whatever strategy was easiest to get the answer, and the child was instructed to speak the answer into a microphone that was interfaced with the computer, which in turn recorded reaction time (RT) from onset of problem presentation to microphone activation. After solving each problem, the child was asked to describe how he or she got the answer. Based on the child's description and the experimenter's observations, the trial was classified based on problem-solving strategy, with the four most common being counting fingers, verbal counting, retrieval, and decomposition. The combination of experimenter observation and child reports immediately after each problem is solved has proven to be a useful measure of children's strategy choices (Geary, 1990; Siegler, 1987). The validity of this information is supported by findings showing that finger-counting trials have the longest RTs, followed by verbal counting, decomposition, and direct retrieval, respectively (e.g., Geary, Hoard, & Bailey, in press; Siegler, 1987).

Previous studies have operationalized skilled retrieval as the number of problems correctly retrieved (i.e., *p*[correct & retrieved]) (e.g., Geary et al., 2007). Although useful, this operationalization conflates children's tendency to attempt to retrieve answers to addition problems (i.e., *p*[retrieved]) and their tendency to answer these problems correctly at a high rate (i.e., *p*[correct | retrieved]. This conflation is expressed schematically in Fig. 1, where the area of the circle represents the total number of problems solved through retrieval, *C* is the number of problems solved correctly (i.e., the common operationalization of skilled retrieval), and *I* is the number of problems solved incorrectly. The number of problems a child retrieved correctly (i.e., *C*) can be increased by increasing skill (i.e., $C/[C + I]$; see Fig. 1B) or by increasing preference for retrieving answers (i.e., $C + I$; see Fig. 1C). We disentangle these components by defining preference as *p*(retrieved) and skill as *p*(correct | retrieved).

Because of very high skill scores in fourth to sixth grades inclusive, and to improve model fit, the skill variable was dichotomized across all grades, such that participants who correctly recalled all of the problems attempted using retrieval were coded as a 1, whereas participants who attempted at least one problem using retrieval but made one or more retrieval errors were coded as a 0. Participants who did not use retrieval for any problems were coded as missing for skill, which occurred in 35%, 13%, 10%, 4%, 3%, and 1% of cases in first to sixth grades, respectively. We also used RT for correctly retrieved problems as a second measure of skilled retrieval. We excluded RTs less than 500 ms and greater than 3000 ms as outliers. The former were uncommon (<1% correct retrieval trials) and likely represented

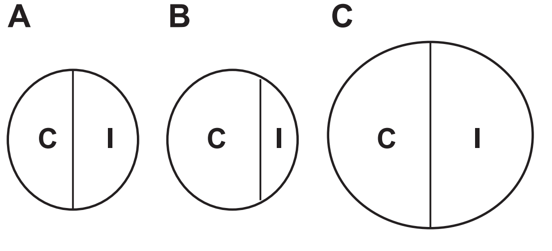

Fig. 1. Schematic representation of the relation between preference and skill. The area of the circle represents overall use of retrieval, and A represents an individual for whom 50% of the retrieved answers were correct and 50% were incorrect. This individual could retrieve more correct answers by making fewer errors (i.e., increasing skill), with no overall change in the number of retrieved answers (B; i.e., no change in preference) or could retrieve more correct answers by retrieving more overall (C). In the latter case, the individual's preference for retrieval increased but skill (i.e., % correct) did not.

premature triggering of the voice-activated relay by extraneous noise, and the latter (11% of trials) is often used as an upper limit for retrieval (e.g., Jordan, Hanich, & Kaplan, 2003; Jordan & Montani, 1997). The RTs were skewed and, thus, submitted to a natural log transformation. Unlike the other variables, which were collected in first to sixth grades, RT was collected only in second to sixth grades because the task was not computerized in first grade. The means and standard errors of the preference and skill variables are shown in Fig. 2.

Working memory and intelligence

The Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) consists of nine subtests that assess the central executive, phonological loop, and visuospatial sketchpad. All of the subtests have six items at each span level. Across subtests, the span levels range from one to six to one to nine. Passing four items at one level moves the child to the next level. At each span level, the number of items (e.g., words) to be remembered is increased by one. Failing three items at one span level terminates the subtest. The first ($\alpha = .69$) and fifth ($\alpha = .62$) grade central executive span scores were used in these analyses because central executive measures are more consistently related to arithmetic development than are measures of the phonological loop or visuospatial sketchpad, as noted earlier (Geary et al., 2007, in press; Swanson et al., 2008).

The central executive is assessed using three dual-task subtests. Listening Recall requires the child to determine whether a sentence is true or false and then to recall the last word in a series of sentences. Counting Recall requires the child to count a set of four, five, six, or seven dots on a card and then, at the end of the series, to recall the number of counted dots on each card. Backward Digit Recall is a standard format backward digit span. The score is the mean across these three subtests.

The children were administered the Vocabulary and Matrix Reasoning subtests of the WASI (Wechsler, 1999) and the Raven's Coloured Progressive Matrices (CPM; Raven, Court, & Raven, 1993), a nontimed test that is considered to be an excellent measure of fluid intelligence, The CPM scores were standardized based on the 287 children who were administered the test ($M = 100$, $SD = 15$), and the WASI standard scores were converted to the same metric. For preliminary analyses, the Vocabulary subtest was used as a measure of verbal IQ, and the mean of the CPM and Matrix Reasoning tests was used as a measure of nonverbal IQ. A full IQ score (i.e., mean of the three tests, $\alpha = .65$), however, typically provided better and more parsimonious fits to the data than separate verbal and nonverbal scores and, thus, was used in the analyses.

Assessments

The CPM and WASI were administered during the spring of kindergarten and first grade, respectively. The addition tasks were administered as part of a broader mathematical cognition assessment

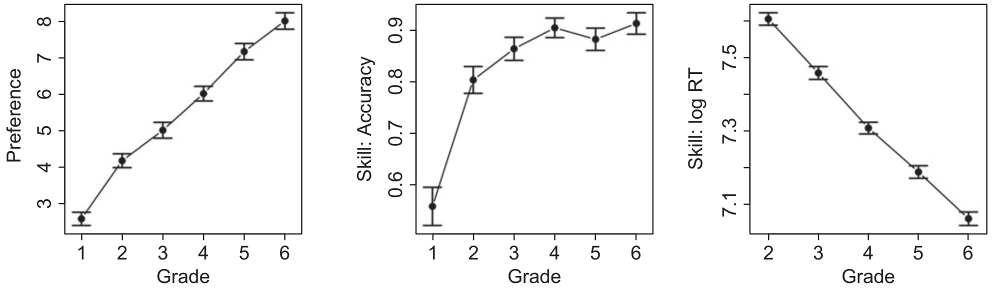

Fig. 2. Means (and standard errors) for preference for retrieval and skilled retrieval (i.e., accuracy and reaction time [RT]) across grades. Preference is the number of problems (out of 14) on which a child attempted to retrieve an answer. Accuracy is the proportion of problems for which the child attempted to retrieve the answer that were answered correctly. Log RT is the natural log of the time a child took to answer retrieved problems correctly.

Table 1
Correlations among accuracy, RT, and preference by grade

Grade	Accuracy with RT	Accuracy with preference	RT with preference
1	NA	–.31***	NA
2	.10	–.20**	.19**
3	.22**	–.14*	.04
4	.17**	–.03	.23***
5	.10	–.07	.27***
6	.12	.01	.31***

Note. RT = –1 ∗ log RT. NA, not applicable.
* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

during the fall of first to sixth grades inclusive. The majority of children were tested in a quiet location at their school site and occasionally on the university campus or in a mobile testing van. Testing in the van occurred for children who had moved out of the school district and for administration of the WMTB-C (e.g., on the weekend or after school). The mean ages at the times of the first- and fifth-grade WMTB-C assessments were 84 months (SD = 6) and 128 months (SD = 5), respectively. The mathematical cognition assessments required between 20 and 40 min, and the WMTB-C required approximately 60 min.

Results

We first report the correlations among the skill and preference variables and the codevelopment of skill and preference across grades. Then, we report on the relations between domain-general cognitive abilities and these variables, specifically testing the talent and practice hypotheses. Lastly, we report on sex differences in skill and preference.

Correlations among skill and preference

Raw correlations among the skill (i.e., accuracy and RT) and preference variables are shown in Table 1.[1] The two skill variables, accuracy and RT (for all of the following analyses, RT was negatively scored, so that higher values indicate lower RTs and higher speeds), were slightly positively correlated across years, with the correlations reaching significance in third and fourth grades. The significant negative correlations between accuracy and preference from first to third grades inclusive likely reflect the attendant trade-off of higher preference resulting in an increased risk of a retrieval error. Similarly, children who use retrieval only when they are absolutely sure that it will result in the correct answer will attempt to retrieve a lower number of answers. However, these correlations become decreasingly negative over time, such that in fourth to sixth grades they are close to 0. During these same years, preference and RT have their highest correlations, ranging from .23 to .31. Taken together, findings suggest that skill and preference become increasingly positively (or decreasingly negatively) related across grades. However, this observation is consistent with both the skill and practice hypotheses. To attempt to differentiate between these, it is necessary to determine how these relations change in cross-lagged models.

Cross-lagged models

Preference and accuracy

The results are shown in the upper section of Fig. 3. The gist is that accurate retrieval in first grade is associated with a preference for retrieval over other strategies in second grade. The magnitude of

[1] These differ slightly from those estimated in the path models below because missing data were not estimated for these correlations as they were in the path models.

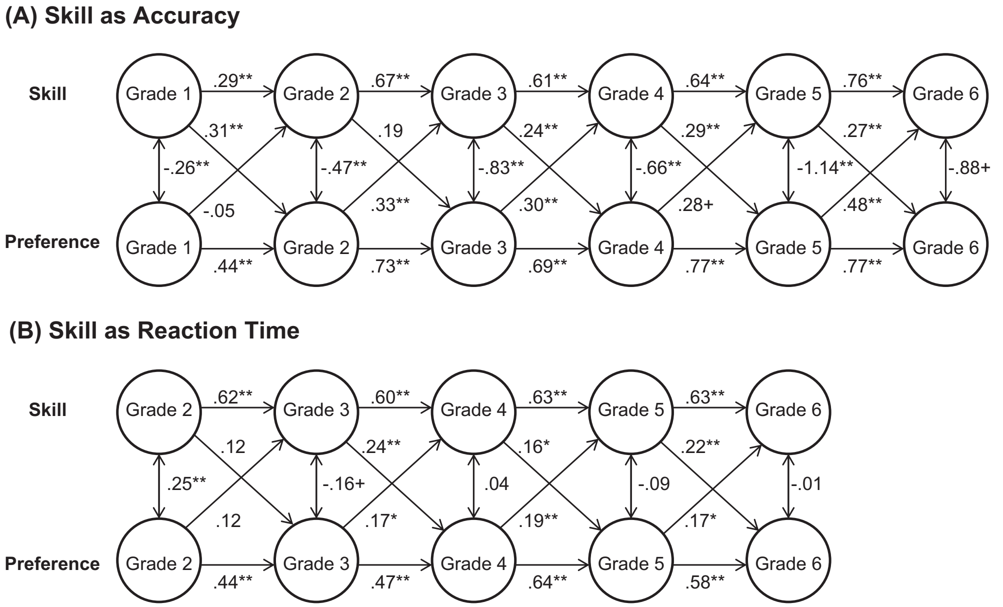

Fig. 3. Cross-lagged models of the relation between skilled retrieval and preference for retrieval: (A) skill as accuracy; (B) skill as reaction time. $^{+}p \leqslant .05$; $^{*}p \leqslant .01$; $^{**}p \leqslant .001$. The −.88 skill, preference covariance in Grade 6, $p = .054$.

this cross-lagged pattern remains similar across grades. Preference for retrieval in first grade was not related to retrieval accuracy in second grade, but higher preference in second grade predicted higher accuracy in third grade. This cross-lagged relation remained significant thereafter and grew in magnitude through fifth grade.

The autoregressive model (AR1) estimates in the upper panel of Fig. 3 provided an acceptable fit to the data, $\chi^2(40) = 82.70$, $p = .0001$, root mean square error of approximation (*RMSEA*) = .06, comparative fit index (*CFI*) = .90.[2] Higher accuracy in first grade was associated with gains in preference in second grade ($\beta = .31$, $p < .001$), but the corresponding cross-lagged path from first grade preference to second grade accuracy was not significantly different from 0. Whereas the pattern of cross-lagged paths from accuracy to preference remained similar across grades, the magnitude of the cross-lagged paths from preference to accuracy increased across grades, such that preference for retrieval in fifth grade was associated with an improvement in accurate retrieval in sixth grade ($\beta = .48$, $p < .001$). The residual covariance between accuracy and preference across time were consistently negative, indicating that, controlling for previous accuracy and preference, there is a trade-off in all grades such that a strong preference for retrieval was associated with less accurate retrieval.

Preference and RT

Beginning in third grade, cross-lagged relations were found between faster retrieval in one grade and stronger preference for retrieval in the next grade and between preference for retrieval in one grade and faster retrieval in the next grade.

The autoregressive model estimates for the lower panel of Fig. 3 reveal a marginal fit as indicated by the *RMSEA* but good fit as indicated by the *CFI*, $\chi^2(24) = 88.98$, $p < .0001$, *RMSEA* = .10, *CFI* = .93. Lower RT (i.e., higher speed) in second grade was not significantly associated with preference in third grade, and the corresponding cross-lagged path from second grade retrieval preference to third grade

[2] Kline (2005) recommended a comparative fit index (*CFI*), which measures the model against a null model in which measures are constrained to be uncorrelated with each other, of greater than .90 and a root mean square error of approximation (*RMSEA*), which measures fit adjusting for sample size, of less than .08.

retrieval RT was also not significant. However, both sets of cross-lagged paths were significant in all subsequent grades, with higher preference associated with future gains in retrieval speed and faster retrieval speed associated with future increases preference for retrieval. The residual covariances between speed and preference were significant in second and third grades but not thereafter. Faster retrieval was associated with stronger preference for retrieval in first grade but with weaker preference for retrieval in second grade.

Domain-general abilities

Correlations

Skill variables and preference relate differently to central executive and IQ scores across grades (Table 2). The pattern in first grade was consistent with the talent hypothesis; this is when higher capacities of the central executive and IQ were most strongly related to frequency of accurate retrieval and were unrelated to preference for retrieval. However, in later grades, as the magnitude of the correlations between domain-general abilities and accurate retrieval declined and the correlations between preference and capacity increased in magnitude, preference began to predict next-year gains in accurate retrieval (above), consistent with the practice hypothesis. Unlike accuracy, moderate correlations between RT and capacity were maintained across grades, and with the exceptions of fourth and fifth grades, IQ was unrelated to RT.

Mediation

Based on the finding that first-grade retrieval accuracy was correlated with central executive capacity and IQ, first-grade accuracy was assessed as a potential mediator of the effects of first-grade capacity and IQ on second-grade preference. First, we assessed the accuracy variable's potential utility as a mediator following Baron and Kenny (1986). The mediator, first-grade retrieval accuracy, is significantly related to the initial variable, first-grade central executive (Table 2). In the following models, we used only observations for which first- and second-grade preference, first-grade accuracy, and first-grade central executive were not missing ($n = 161$), so that each model is based on the identical observations.

Model 1 (Table 3) shows that first-grade central executive predicts second-grade preference ($\beta = .23$, $p < .01$), with simultaneous control of first-grade preference. Model 2 shows that second-grade preference is related to both first-grade central executive scores ($\beta = .18$, $p < .05$) and first-grade accuracy ($\beta = .17$, $p < .05$), but the magnitude of the former effect is reduced relative to Model 1. Mediation tests were performed using the distribution of the product method (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) using the R package RMediation (Tofighi & MacKinnon, 2011). Although the effect of first-grade central executive is not fully mediated by first-grade accuracy, mediation was significant (95% confidence interval for the mediated effect = [.0003, .037]); thus, the effect

Table 2
Correlations among central executive, skill, and preference

Grade	Central executive, Grade 1			Central executive, Grade 5			IQ, Grade 1		
	Accuracy	RT	Preference	Accuracy	RT	Preference	Accuracy	RT	Preference
1	.38***	NA	−.10	.31***	NA	−.02	.36***	NA	−.03
2	.32***	.24***	.17**	.14	.23**	.28***	.30***	.14	.13*
3	.30***	.27***	.18**	.26***	.27***	.28***	.19**	.12	.22***
4	.19**	.37***	.14*	.18+	.38***	.26***	.02	.18*	.35***
5	.19**	.41***	.25***	.21**	.38***	.33***	.18**	.26**	.24***
6	.15*	.32***	.27***	.23**	.34***	.24**	.22**	.13	.12

RT = $-1 * ln$ RT. NA, not applicable.
* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

Table 3
First- and second-grade mediation models

First-grade variable	Predicting second-grade preference for retrieval				
	β (Model 1)	β (Model 2)	β (Model 3)	β (Model 4)	β (Model 5)
Accuracy		.17*		.22**	.16
Preference	.20**	.24***	.16*	.21**	.24**
Central executive	.23**	.18*			.16
IQ			.19*	.13	.04
R^2	.08**	.10***	.05**	.09***	.10**

* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

of first-grade central executive on second-grade preference is partially mediated by first-grade accuracy.

Models 3 and 4 show the same pattern but replacing the central executive variable with IQ (these models are based on observations for which first- and second-grade preference, first-grade skill, and IQ were not missing, $n = 173$). The mediator, first-grade retrieval accuracy, is significantly related to the initial variable, first-grade IQ (Table 2). Model 3 (Table 3) shows that first-grade IQ predicts second-grade preference ($\beta = .19$, $p < .05$), with simultaneous control of first-grade preference. Model 2 shows that second-grade preference is related to first-grade accuracy ($\beta = .22$, $p < .01$), but the effect of first-grade IQ is no longer statistically significant. Mediation was significant (95% confidence interval for the mediated effect = [.0003, .027]); thus, the effect of first-grade IQ on second-grade preference is fully mediated by first-grade accuracy.

To test whether first-grade central executive and IQ both contribute unique variance to second-grade preference, or whether the covariance between one of these variables and second-grade preference can be attributed to the relation between IQ and central executive, both variables were entered into the same model predicting second-grade preference along with first-grade accuracy and preference (Model 5, Table 3). The standardized regression weights for central executive and IQ were no longer statistically significant, suggesting that shared variance between central executive and IQ contributed to second-grade preference.

The skill variables (i.e., accuracy and RT) could not be assessed as mediators of the relationship between fifth-grade central executive and sixth-grade preference because after controlling for fifth-grade preference, fifth-grade central executive was not significantly related to sixth-grade preference ($\beta = -.03$, $p = .74$). Similarly, preference could not be assessed as a mediator of the relation between fifth-grade central executive and sixth-grade RT because after controlling for fifth-grade RT, fifth-grade central executive was not significantly related to sixth-grade RT ($\beta = .10$, $p = .10$). Finally, preference could not be assessed as a mediator of the relation between fifth-grade central executive and sixth-grade accuracy because after controlling for fifth-grade central executive and fifth-grade accuracy, fifth-grade preference was no longer significantly related to sixth-grade accuracy (log odds = .07, $p = .50$).

Thus, early cross-lagged paths between skill (specifically accuracy) and preference suggest that the influence of domain-general abilities on later preference for retrieval is partially mediated by accurate retrieval. However, cross-lagged paths in later grades seem to reflect the emergence of domain-specific influences of skill and preference on each other.

Sex differences

Both boys' and girls' means and standard errors for across-grade skill and preference are shown in Fig. 4. Sex differences in preference were assessed across grades using random intercept models in R (nlme package; Pinheiro, Bates, DebRoy, Sarkar, & R Development Core Team., 2010). A model with preference as the dependent variable and sex, grade (coded such that the intercept is assessed at first grade), and a sex by grade interaction yielded main effects for grade, $t(1205) = 20.23$, $p < .00005$

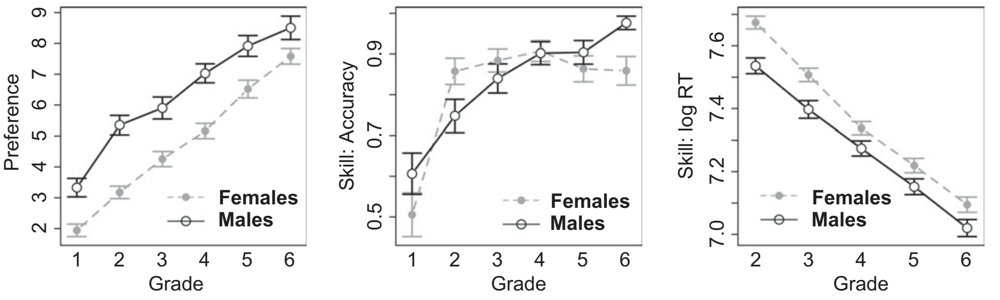

Fig. 4. Means (and standard errors) for boys' and girls' preference for retrieval and skilled retrieval (i.e., accuracy and reaction time [RT]) across grades.

(individuals increased over time on preference for retrieval), and sex, $t(289) = 5.75$, $p < .00005$ (boys had a higher preference), but the interaction of grade and sex was nonsignificant, $t(1205) = -0.98$, $p = .33$, indicating that boys' higher preference for retrieval remained consistent across grades.

Sex differences in accuracy were assessed across grades using generalized linear mixed models in R (lme4; Bates & Maechler, 2010). Intercept was a random variable, and because the skill variable was binary, the logit link function was used. Because of the large increase in accuracy for both sexes between first and second grades, which causes an irregular pattern of growth across time, only time points from second to sixth grades inclusive were analyzed. For first-grade scores, a logistic regression model indicated that boys' higher accuracy was not significant ($z = 1.36$, $p = .17$). A model with accuracy as the dependent variable and sex, grade (coded so that the intercept is assessed at second grade), and a sex by grade interaction as predictors yielded no main effect for grade (881 observations, $z = -0.74$, $p = .46$; girls' skilled retrieval did not increase significantly across time), but the effect for sex was significant ($z = -2.02$, $p = .04$; girls were significantly higher on accurate retrieval in second grade), qualified by a sex by grade interaction ($z = 2.68$, $p = .007$; boys' accurate retrieval increased significantly faster than girls' across grade). A follow-up logistic regression model indicated that boys were significantly higher on accurate retrieval than girls in sixth grade ($z = 2.49$, $p = .01$).

Sex differences in RT were assessed across grades using random intercept models. A model with RT (negatively scored, so that higher values indicate faster speeds) as the dependent variable and sex, grade (coded such that the intercept is assessed at second grade), and a sex by grade interaction yielded main effects for grade, $t(768) = 28.91$, $p < .00005$ (individuals became faster at retrieval across grades), and sex, $t(263) = 4.62$, $p < .00005$ (boys had faster retrieval times across grades), and a smaller but statistically significant interaction of grade and sex, $t(768) = -2.87$, $p = .004$, indicating that girls' retrieval speed increased slightly more across grades than did boys'.

To test whether the sexes differed on working memory, t tests were run on central executive scores and revealed no significant sex difference for either first grade, $t(266) = 0.44$, $p = .66$, or fifth grade, $t(266) = -0.30$, $p = .77$. A sex difference was found in IQ, favoring boys, $t(286) = 2.20$, $p = .03$, Cohen's $d = 0.26$, indicating a small effect. However, adjusting for IQ in the multilevel, generalized linear mixed, and logistic regression models described above did not affect any of the sex differences presented in this section.

Discussion

It has been well documented that children use a mix of strategies to solve sets of addition and other academic problems (Siegler, 1996) and that their choices are systematic; children's use of one strategy or another reflects a trade-off between the time needed to execute the strategy and the probability that the generated answer will be correct (Siegler, 1987; Siegler & Shrager, 1984). The mechanisms underlying these trade-offs are not fully understood, but they include in part some combination of working memory capacity and intelligence (DeStefano & LeFevre, 2004; Geary & Burlingham-Dubree, 1989; Geary, Hoard, Byrd-Craven, & DeSoto, 2004). Not all choices are driven by these trade-offs, however, as some appear to involve a preference for one strategy over others independent of strategy accuracy (Siegler, 1988), and boys and girls may differ in these preferences, especially with respect to the use of fact retrieval (Carr & Davis, 2001). Our 6-year longitudinal study contributes to these literatures by documenting the cross-lagged relations between preference or skill in one grade and preference and skill in the next grade and testing whether these relations are best explained by talent, practice, or both and whether they vary across grades or sex.

Codevelopment of skill and preference

Skilled retrieval and preference for use of retrieval over other strategies, independent of retrieval accuracy, codeveloped across first to sixth grades inclusive. The practice hypothesis is that preference for using retrieval in earlier grades leads to improved skilled retrieval in later grades independent of previous skill and general cognitive abilities. The talent hypothesis is that higher working memory capacity and intelligence facilitates early skilled retrieval that in turn results in a preference for retrieval in later grades.

Consistent with the talent hypothesis, first graders with higher working memory capacity and intelligence showed more accurate retrieval than their peers and accurate retrieval in first grade predicted preference in second grade, but preference in first grade did not predict accurate retrieval in second grade. The importance of working memory and intelligence for early accurate retrieval, combined with the relation between retrieval and achievement, may contribute to the often-found correlation between these cognitive abilities and performance on mathematics achievement tests (Bull et al., 2008; Geary et al., 1996; Mazzocco & Kover, 2007; Noël, 2009; Passolunghi et al., 2007; Royer et al., 1999; Swanson & Sachse-Lee, 2001).

These findings are also consistent with brain imaging studies showing more engagement of the prefrontal cortex during children's arithmetic fact retrieval than during adults' (Cho et al., 2011; Rivera et al., 2005). Rivera and colleagues' (2005) results further suggest that as retrieval becomes more automatized, the foci of brain activity shift from prefrontal to parietal regions. Our results are in keeping with this pattern; that is, the importance of working memory and intelligence for accurate retrieval declined across grades and the contributions of an earlier preference for retrieval increased across grades. The pattern suggests that the relative contribution of talent versus practice shifts across grades, such that preference for retrieval predicts future increases in accuracy and speed of retrieval just as much as these skill variables predict future increases in preference. The pattern of early preference predicting later skill was repeated from second to sixth grades for accuracy and from third to sixth grades for RT. The strength of the relation between early skill and later preference remained similar in magnitude across grades in both models.

Our results suggest that the talent and practice hypotheses are simultaneously true for the majority of the developmental period we observed. At least by third grade, there seems to be a feedback loop, where increases in preference are associated with later increases in skill, which are associated with later increases in preference, and so on. Furthermore, the influence of central executive on later skill and preference is lower in fifth grade than in first grade. Talent in the form of high working memory and IQ in early grades influences future preference via skill increases, but skill increases in later grades may be influenced more by the skill preference feedback loop than by domain-general abilities per se.

Sex differences

Previous studies have shown that boys and girls differ in the mix of strategies they use for solving addition problems, with boys showing a bias toward the use of retrieval and girls showing a bias for the use of counting procedures (Carr & Davis, 2001; Carr & Jessup, 1997; Geary et al., 1996). The same was found in the current study for all six grades, with boys always preferring retrieval to alternative strategies more frequently than girls regardless of retrieval accuracy. The developmental pattern for skilled retrieval, however, was more complex. In first grade, boys had a higher preference for retrieval than girls, but the sexes did not differ significantly on accuracy. By second grade, girls used retrieval less than boys but were more accurate than boys when they did use it, reminiscent of Siegler's (1988) findings for "perfectionists" and of Carr and Alexeev's (2011) finding that children in a group who were less likely to adopt retrieval as a preferred strategy across grades were more likely to be girls.

The former are students who are more capable of accurate retrieval than suggested by their strategy choices. Rather than risk a retrieval error, they resort to more time-consuming but typically accurate counting strategies. Girls' retrieval accuracy stayed constant from second to sixth grades as their use of retrieval steadily increased, suggesting that their criterion for avoiding errors did not change. Boys' retrieval accuracy, in contrast, increased steadily from first to sixth grades, and at this point they outperformed girls.

Boys' early advantage for accurate retrieval was not due to an advantage on the capacity of the central executive, given that there were no sex differences on these measures, nor was it related to boys' slight advantage in IQ in this sample. Although we do not have direct measures of risk taking, this should be considered as a strong possibility. This is because males generally take more risks than females, especially in situations where their performance will be socially evaluated (Byrnes et al., 1999; Geary, 2010), and tend to express themselves more freely in classroom settings that encourage competition. One way they do so in these classrooms is through getting the answer before anyone else (Peterson & Fennema, 1985). In the current study, boys had shorter RTs than girls across grades, although girls caught up slightly over this time frame. Interestingly, boys' advantage in RT did not increase across grades, as it did for accuracy, despite their consistently higher preference for retrieval. One potential reason is that competitiveness may affect RT more directly (as it affects preference) than it affects accuracy. That is, boys across grades seek to answer single-digit addition problems faster and more often than girls, perhaps because they are more competitive or less perfectionist, which may harm boys' addition accuracy in early grades but put them at an advantage later via the practice hypothesis.

Limitations

Like other authors who have examined the relation among constructs across time (e.g., Littlefield, Sher, & Wood, 2010; Scollon & Diener, 2006), we want to highlight the limitations of drawing strong conclusions about causality from the cross-lagged model or from any observational data (Freedman, 1987). Autoregressive techniques such as those used here model multiwave data as a series of two-wave "snapshots" (Rogosa, 1980, p. 255) that most likely represent an oversimplification of the development process. Furthermore, the length of time covered by the lags in these models needs to correspond to the time course of the proposed underlying causal process; lags that are too long or too short may lead to misleading results. Although the autoregressive models fit the data well, it should also be stressed that these analyses by no means exhaust the plausible theoretical and analytic models that could be applied to these data. The causal structure assumed by these proposed models may also be fit equivalently by models that make no such attributions of causal influence between skill and preference (see Tomarken & Waller, 2003). The accuracy variable's ceiling effect (and the relatively simple problems that the children solved) may have affected the loadings in the cross-lagged models and our resulting interpretations. However, accuracy and preference show significant cross-lagged loadings in both directions at the last four grades, and grade-to-grade consistency of the accuracy variable remains high at these grades. Finally, our instructions attempted to balance speed–accuracy trade-offs, and changing the emphasis on speed or accuracy of problem solving is likely to change the pattern of relations among skill and preference. Regardless, future research on skill and strategy preference in different types of mathematical tasks will be needed to demonstrate the replicability and generalizability of the talent and practice hypotheses.

Despite these limitations, the current study suggests a complex process by which talent and practice contribute to the development of the skilled use of retrieval-based addition strategies, especially given the pattern with which these variables relate to measures of domain-general cognitive abilities. Furthermore, this model parsimoniously explains observed sex differences in the development of skill at and preference for these strategies.

Acknowledgments

The authors acknowledge support from Grants R01 HD38283 and R37 HD045914 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and Grant F31AA019596 from the National Institute on Alcohol Abuse and Alcoholism to Andrew K. Littlefield. We thank Mary Hoard, Lara Nugent, Kendra Anderson Cerveny, Jennifer Byrd-Craven, Rachel Christensen, Caitlin Cole, Mike Coutts, Sara Ensenberger, Nick Geary, Larissa Haggard, Rebecca Hale, Patricia Hoard, Stacey Jones, Mary Lemp, Patrick Maloney, Rehab Mojid, Cy Nadler, Chatty Numtee, Amanda Shocklee, Jennifer Smith, Ashley Stickney, Jonathan Thacker, Leah Thomas, Jasmine Tilghman, Erin Twellman, Katie Waller, and Alex Wilkerson for help on various aspects of the project as well as two anonymous reviewers for helpful comments on an earlier draft.

References

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology, 47*, 905–915.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2*, 213–236.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–90). New York: Academic Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using S4 classes* (R package, Version 0.999375-33). Available at: <http://CRAN.R-project.org/package=lme4>.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functions in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205–228.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*, 367–383.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*, 179–202.

Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology, 103*, 617–631.

Carr, M., & Davis, H. (2001). Gender differences in arithmetic strategy use: A function of skill and preference. *Contemporary Educational Psychology, 26*, 330–347.

Carr, M., & Jessup, D. L. (1997). Gender differences in first-grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology, 89*, 318–328.

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

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence. A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How do you solve 7 + 8? Decoding children's brain activity patterns during counting versus retrieval. *Developmental Science, 14*, 989–1001.

Colom, R., Juan-Espinosa, M., Abad, F., & García, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence, 28*, 57–68.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*, 186–201.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386.

Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence, 20*, 169–189.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12*, 101–128.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363–383.

Geary, D. C. (2010). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.

Geary, D. C., Bow-Thomas, C. C., Liu, F., & Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044.

Geary, D. C., Brown, S. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787–797.

Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology, 47*, 175–192.

Geary, D. C., Hoard, M. K., & Bailey, D. H. (in press). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities*.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, C. M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*, 121–151.

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*, 1343–1359.

Geary, D. C., Hoard, M. K., & Nugent, L. (in press). Different cognitive mechanisms contribute to the development of retrieval and procedural competencies in addition. *Journal of Experimental Child Psychology*.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five year prospective study. *Journal of Educational Psychology, 104*, 206–223.
Geary, D. C., & Widaman, K. F. (1987). Individual differences in cognitive arithmetic. *Journal of Experimental Psychology: General, 116*, 154–171.
Geary, D. C., & Widaman, K. F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence, 16*, 47–80.
Gottfredson, L. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.
Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.
Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*, 103–119.
Jordan, N. C., & Montani, T. O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities, 30*, 624–634.
Klein, J. S., & Bisanz, J. (2000). Preschoolers doing arithmetic: The concepts are willing but the working memory is weak. *Canadian Journal of Experimental Psychology, 54*, 105–115.
Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.
Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual–spatial working memory, and preschool quantity–number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516–531.
Littlefield, A. K., Sher, K. J., & Wood, P. K. (2010). Do changes in drinking motives mediate the relation between personality change and "maturing out" of problem drinking? *Journal of Abnormal Psychology, 119*, 93–105.
Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481–498.
MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.
Mazzocco, M. M. M., & Kover, S. T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology, 13*, 18–45.
Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.
National Council of Teachers of Mathematics (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: Author.
National Mathematics Advisory Panel (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.
Noël, M.-P. (2009). Counting on working memory when learning to count and to add: A preschool study. *Developmental Psychology, 45*, 1630–1643.
Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability, and numerical competence. *Cognitive Development, 22*, 165–184.
Peterson, P. L., & Fennema, P. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal, 22*, 309–335.
Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children (WMTB-C) manual*. London: Psychological Corporation.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Development Core Team. (2010). *nlme: Linear and nonlinear mixed effects models* [documentation for nlme package]. Available at: <http://cran.r-project.org/web/packages/nlme/nlme.pdf>.
Raven, J. C., Court, J. H., & Raven, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H.K. Lewis.
Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex, 15*, 1779–1790.
Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88*, 245–258.
Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H. III, (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology, 24*, 181–266.
Scollon, C. N., & Diener, E. (2006). Love, work, and changes in extraversion and neuroticism over time. *Journal of Personality and Social Psychology, 91*, 1152–1165.
Foresman, Scott (1999). *Investigations in number, data, and space*. Glenview, IL: Author.
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.
Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59*, 833–851.
Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229–293). Hillside, NJ: Lawrence Erlbaum.
Stevenson, H. W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. (1976). Longitudinal study of individual differences in cognitive development and scholastic achievement. *Journal of Educational Psychology, 68*, 377–400.
Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*, 343–379.
Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294–321.
Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692–700.
Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, 112*, 578–598.
Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*, 19–27.
Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: PsychCorp, Harcourt Assessment.
Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*, 575–587.