Journal of Experimental Child Psychology 146 (2016) 50–65

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Here, but not there: Cross-national variability of gender effects in arithmetic

Chen Shen, Marina Vasilyeva *, Elida V. Laski

Boston College, Chestnut Hill, MA 02467, USA

ARTICLE INFO

Article history:
Received 7 October 2015
Revised 25 January 2016

Keywords:
Arithmetic
Strategy
Cross-cultural differences
Gender differences
Problem-solving
Math

ABSTRACT

The current study examined gender differences in arithmetic performance among first graders from the United States, Russia, and Taiwan. Children ($N = 250$, $M_{age} = 7$ years 2 months) solved simple (single-digit) and complex (mixed- and double-digit) addition problems and explained their strategies. On simple problems, there were gender differences in strategies that varied across countries but no differences in accuracy. On complex problems, there were gender differences among American and Russian students in strategy use that mediated differences in accuracy. In contrast, among Taiwanese students, there were no gender differences in strategies or accuracy. The pattern of results suggests that educational context may play a role in gender differences in mathematics.

Introduction

Understanding the nature, and in particular the extent of malleability, of gender differences in mathematics has important implications for educational practice, and thus it is not surprising that this issue has attracted the attention of psychologists and educators for decades (e.g., Carr, Steiner, Kyser, & Biddlecomb, 2008; Else-Quest, Hyde, & Linn, 2010; Fennema, Carpenter, Jacobs, Franke, & Levi, 1998; Hyde, 2005). A promising approach for exploring the extent to which gender differences can be modified is offered by cross-national investigations that include students from different educational contexts. A consistent pattern of findings across countries with diverse educational environments would

* Corresponding author.
E-mail address: vasilyev@bc.edu (M. Vasilyeva).

http://dx.doi.org/10.1016/j.jecp.2016.01.016
0022-0965/

suggest that gender differences are stable and insulated from experience. Alternatively, differences in gender patterns across countries would suggest that gender differences are malleable and implicate structural characteristics of the educational system and/or specific pedagogical approaches as facilitating or impeding gender parity in math performance.

In the current study, we took a cross-national approach to explore gender differences among young elementary school students in three countries with distinct math curricula. We examined both the accuracy demonstrated by boys and girls on arithmetic problems and the strategies they used. Looking at the strategies provided a window into children's reasoning underlying differences in accuracy. Below we review literature on the two issues central to the current investigation: cross-national research about gender differences in math performance and research about gender differences in strategy use in solving arithmetic problems.

Cross-national variability in mathematics gender differences

The majority of studies that have compared the size of gender differences across countries examined high school and upper middle school students' performance on large-scale international assessments such as Trends of International Mathematical and Science Study (TIMSS) and Program for International Student Assessment (PISA) (e.g., Ayalon & Livneh, 2013; Bedard & Cho, 2010; Else-Quest et al., 2010; Penner, 2003). Many of these investigations reveal variability among countries in the size of gender differences in math achievement. When the data are averaged across countries, a small boys' advantage typically emerges. When the data are examined by country, however, some countries (e.g., Norway, Japan) show no significant gender difference and others (e.g., Iceland) demonstrate an advantage for girls over boys. Smaller scale studies that have compared boys' and girls' math performance have also demonstrated variability in gender patterns, with differences between boys and girls being greater in the United States than in Asian countries. Byrnes, Hong, and Xing (1997) examined American and Chinese high school students' performance on Scholastic Aptitude Test (SAT)-type items and found a gender difference favoring boys among American, but not Chinese, students. Cai (1995) examined the performance of sixth graders on a battery of math tests and found gender differences among American, but not Japanese, students. Overall, both large- and small-scale studies suggest that Asian students might not demonstrate the same extent of gender differences observed among their American peers.

This pattern of findings leads to a critical question: Why do gender differences seem to exist in some contexts but not in others? One possibility is cross-national differences in mathematics instruction. Because participants in the cross-national studies of gender differences to date have typically included middle and high school students with many years of educational experience, it is possible that the observed variability in the patterns of gender findings in part reflect cross-national differences in structural educational factors such as the presence or absence of tracking and class placement. Consistent with this explanation, Bedard and Cho (2010) demonstrated that countries that practice tracking in upper grades are more likely to reveal gender differences in high school math achievement. This trend may reflect differences in boys' and girls' educational experiences resulting from a higher proportion of boys placed in advanced classes.

It is not entirely clear from the available research whether cross-national variability in gender effects also exists among younger elementary school students who have not yet received differential instruction. There are very few cross-national studies that have examined gender effects in young students' math performance, and these have not reported gender differences within or across countries. For example, no differences were found among Chinese and Finnish 4- to 7-year-old children on several numeric tasks such as number naming and counting objects (Aunio et al., 2006). Similarly, Geary, Bow-Thomas, Liu, and Siegler (1996) tested children from kindergarten through third grade from China and the United States using simple arithmetic tasks (single-digit addition) and found no gender effects on accuracy of performance in either country.

It is possible, however, that this null effect is partly due to the kinds of problems children were asked to solve. Recent work on the interaction between task difficulty and gender (e.g., Gibbs, 2010; Penner, 2003) suggests that using simple tasks or even aggregating results from both challenging and simple tasks may mask differences between boys and girls. Consistent with this view, a recent study in which students were given only challenging computational and geometric problems revealed strong gender differences on both types of problems (Ganley et al., 2013). Thus, gender differences in arithmetic accuracy may emerge if young students are asked to solve more challenging problems such as addition tasks involving multi-digit numbers.

Furthermore, gender differences may emerge among young children when examining aspects of problem solving other than accuracy. Similarities in accuracy might mask potential differences in underlying approaches to solving arithmetic problems. Investigators looking into individual differences in academic performance are increasingly aware of the need to examine strategies in conjunction with the overall level of performance (e.g., Carr & Davis, 2001; Vasilyeva, Casey, Dearing, & Ganley, 2009). If young children demonstrate gender differences in strategy use, even while showing no differences in accuracy on simple problems, these tendencies may eventually translate into differences in performance on more complex problems.

Gender differences in arithmetic strategies

Although there are no cross-national studies of gender differences in arithmetic strategy use, there is evidence of such differences in the United States. One line of work has focused on the distinction between manipulative strategies, which involve the use of physical representations of numbers (e.g., counters, child's fingers) and mental strategies such as verbal counting and retrieval (first graders: Carr & Davis, 2001; Carr & Jessup, 1997; second graders: Carr et al., 2008; second to fourth graders: Carr & Alexeev, 2011). Carr and colleagues have consistently found that girls tend to use manipulatives more than boys, who in turn prefer to use mental strategies. When forced to use a particular kind of strategy, boys are equally capable of using manipulatives, whereas, girls do not employ mental strategies to the same extent as boys (Carr & Davis, 2001). Similar findings have been reported by other researchers. For example, Jordan, Kaplan, Ramineni, and Locuniak (2008) tracked children from the beginning of kindergarten to the end of second grade and found that girls used finger-based strategies more often than boys on single-digit addition and subtraction problems.

Other researchers examining elementary students' approach to arithmetic problems distinguish between computational strategies that are based on counting (which may involve the use of manipulatives or mental operations) and memory-based strategies (decomposition and retrieval) (e.g., Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Geary et al., 1996). *Counting* involves enumerating either one or both of the addends (e.g., 6 + 3: "7, 8, 9"). *Decomposition* involves transforming the original problem into two or more simpler problems (e.g., 6 + 5: 6 + 4 = 10, 10 + 1 = 11). The *retrieval* strategy involves recalling a number fact stored in memory rather than active computation. Of these three strategy types, retrieval is the most efficient, but it is not always available because it requires memorizing number facts, which in turn requires sufficient experience in solving problems with specific numbers. When children cannot retrieve a response from memory, they can use the remaining two strategies as a backup. Counting works quite well with small numbers but becomes increasingly inefficient and error prone with larger numbers, in which case decomposition offers a better alternative.

Several studies examining the use of these computational strategies in solving simple arithmetic problems have found distinct preferences for boys and girls. For example, in a longitudinal study of American students from first grade to third grade, Fennema and colleagues (1998) reported gender differences in strategy use, with girls, but not boys, preferring to use counting strategies on addition and subtraction tasks. Another longitudinal study included a sample of American students who were tracked from first grade to sixth grade (Bailey, Littlefield, & Geary, 2012). Presented with single-digit addition problems, boys across grades demonstrated a preference for retrieval over other strategies. A similar result was reported in a study of older students (fourth to sixth graders) from Belgium who were presented with single-digit arithmetic problems (Imbo & Vandierendonck, 2007). Furthermore, boys demonstrate higher fluency on single-digit arithmetic problems (measured by response time on correctly solved problems), which is consistent with their use of more efficient problem solving approaches (Carr et al., 2008).

Studies investigating gender differences in strategy use provide insights into divergent ways in which boys and girls approach arithmetic problems. Yet, there are many remaining questions. In particular, it is unknown whether the gender-based strategy preferences found on simple arithmetic problems are also present on complex ones. A few studies that included both simple and complex problems examined strategies broadly without distinguishing between cognitive strategies such as counting versus decomposition (e.g., Carr & Davis, 2001; Carr et al., 2008). Other studies that did look specifically at the various computational strategies in young children included only simple problems (e.g., Bailey et al., 2012). Based on the work of Siegler and colleagues, one could expect that the kinds of strategies that are more prevalent in solving complex arithmetic problems would differ from those used with simple problems (Shrager & Siegler, 1998; Siegler, 1988). Consequently, the pattern of gender effects may vary depending on the type of problem. A further question is whether the gender differences in strategy choices reported in studies conducted with American children can be replicated across different educational contexts. There is a substantial body of cross-national research that has documented both country and gender effects on arithmetic accuracy, but no studies have examined patterns of gender differences in strategy use across countries.

Another important issue that remains to be investigated concerns the relation between gender differences in strategy use and in accuracy. This relation might not be apparent when examining performance on simple problems. For example, even when using a relatively inefficient strategy, such as counting as opposed to retrieval or decomposition, on single-digit addition (e.g., 4 + 2), children may easily arrive at an accurate response, whereas using the same strategy on a problem involving double-digit addends (e.g., 36 + 27) makes the solution process extremely inefficient. Thus, the choice of a more versus less advanced strategy on complex, as opposed to simple, problems is more likely to impact accuracy. Consequently, an investigation of gender differences on complex arithmetic problems may provide important information concerning the relation between strategy use and accuracy.

The current study

The current study investigated first graders' performance on arithmetic problems, extending the work on addition strategies to cross-national research on gender. The design integrated several approaches so as to extend the current understanding of gender effects. First, children were presented with both simple and complex problems, which enabled us to determine whether/how gender patterns vary as a function of problem difficulty. Second, accuracy was examined in conjunction with the strategies used by boys and girls, and the relation between these two aspects of performance was explored. Third, the cross-national nature of the study allowed us to look at gender differences both within and across educational contexts to examine the extent to which gender differences may be malleable, potentially reflecting cross-national differences in instructional practices.

Study participants were recruited in three countries: Taiwan, Russia, and the United States. Prior to conducting the study, we examined curricular guidelines and instructional materials used for teaching first-grade math in the participating schools. We found that in Taiwan, which has a centralized curriculum, students received more intense practice with basic number facts compared with the other two countries. Furthermore, the Taiwanese curriculum (*The Nine-Year School Curriculum;* Ministry of Education of Taiwan, 2011) placed a greater emphasis on the use of decomposition strategies as opposed to counting. In the United States, participating schools used several different curricula (primarily *Investigations in Number, Data, and Space* (TERC, 2008) and *Everyday Mathematics* (University of Chicago School Mathematics Project, 2012)), all of which included instruction in arithmetic strategies. Decomposition was introduced during instruction, but it was typically presented as one of many available strategies from which children could choose. The Russian first-grade curriculum (*Elementary School of the XXI Century;* Istomina, 2011) also introduced a variety of strategies to first graders, although these teaching materials contained less explicit instruction on strategies than either the Taiwanese or American materials. A large portion of instruction focused on teaching children how to solve contextual word problems that had simple computational requirements rather than on the mechanics of computation itself. Thus, the countries included in this investigation offered a range of approaches to teaching children arithmetic and, in particular, arithmetic strategies.

Our study addressed several research questions. The first set of questions concerned gender differences in accuracy and strategy use on single-digit addition problems. Based on prior cross-national findings, we hypothesized that there would be no significant gender differences in accuracy on this type of problems. With respect to strategies, we expected that boys and girls may prefer different approaches to solving arithmetic problems, as suggested by research conducted in the United States. At the same time, our examination of math curriculum suggested that the pattern of strategy use may vary across countries. For example, Taiwanese students, whose curriculum includes a large amount of practice in memorizing simple number facts, may be more likely to use memory-based strategies, such as retrieval and decomposition, compared with their counterparts in the other two countries. Thus, we aimed to determine whether students' use of particular strategies varied by gender and country and whether there was an interaction between these two factors. Given the differences in instructional emphases on arithmetic strategies across the countries, a gender by country interaction would indicate that gender differences in strategy use, and consequently arithmetic accuracy, are malleable and can be attenuated with targeted instruction.

The second set of questions concerned gender differences in children's performance on more complex addition problems that included at least one double-digit addend. Few studies have examined young children's performance on complex problems, which made it difficult to make clear predictions with respect to gender. Recent studies in the United States, however, suggest that gender effects are more likely to emerge on challenging mathematical tasks (Ganley et al., 2013; Gibbs, 2010). Thus, it seemed possible that gender differences would be present among first graders on complex arithmetic problems—in terms of both accuracy and strategy use.

With respect to strategies, we were particularly interested in the use of decomposition, which becomes critically important in solving complex arithmetic problems. For this type of problems, the retrieval strategy might not be available (because it is hard to memorize a large set of number facts beyond single-digit numbers) and the counting strategy is extremely inefficient and error prone. Given the differences among participating countries in the emphasis on this strategy, and given prior empirical findings indicating greater use of decomposition by Asian students (Carr et al., 2008; Geary et al., 1996), we expected that Taiwanese participants would use this strategy more often. It is further possible that intensive instruction and practice with decomposition may reduce gender differences in using this strategy, leading to cross-national variability in gender effects.

The third set of questions concerned the relation between accuracy and strategy use. We anticipated differences between boys and girls in both accuracy and strategies on complex arithmetic problems. We aimed to determine whether gender differences in strategies could account for gender differences in the accuracy of students' performance. We focused specifically on differences in the use of decomposition strategy because that strategy has been shown to be associated with higher levels of performance on arithmetic problems (Geary et al., 2004). Because the use of decomposition on a particular problem leads to a greater probability of solving that problem correctly, students who use decomposition on a higher percentage of trials should have greater overall accuracy than students who use decomposition on a lower percentage of trials. To explore this relation, we conducted a mediation analysis with frequency of decomposition use as a potential mediator of the relation between gender and accuracy.

Method

Participants

Participants were 250 first graders in three countries: United States ($n = 92$), Russia ($n = 98$), and Taiwan ($n = 60$). The samples were comparable in age and gender distribution (United States: $M_{age} = 7;2$ [years;months], 52% girls; Russia: $M_{age} = 7;8$, 51% girls; Taiwan: $M_{age} = 6;7$, 53% girls). Participating students were from middle- and upper middle-class families. In the United States, children were from suburbs largely populated by well-educated professionals (75% of adults held a bachelor's or higher degree). In Russia and Taiwan, where demographic data were not publicly available, we relied on information about schools' reputations provided by local researchers. Based on this information, students were from schools primarily serving families with high educational and professional status.

All of the students were from mainstream elementary classrooms in mixed-gender schools. After a careful examination of math curricula and interviews with teachers, we determined that by the time of the study participants in all three countries had a substantial amount of practice with single-digit addition but less practice with addition tasks involve double-digit numbers. It is possible that some children were exposed to such problems outside of school, but this exposure would be relatively minimal compared with single-digit addition.

Materials

Materials were 24 addition problems. Of these, 12 were single-digit problems in which the addends varied between 1 and 9 (excluding problems with identical addends); these are referred to as simple problems. Half of the simple problems had sums up to 10 (e.g., 6 + 2), and the other half had sums over 10 (e.g., 3 + 8). The other 12 addition problems included one or two double-digit addends; these are referred to as complex problems. Half of the complex problems did not involve carryover (e.g., 5 + 22, 11 + 17), and the other half involved carryover (e.g., 18 + 3, 25 + 37). Doubles (e.g., 44, 88) were excluded as potential addends, and problems with identical addends were also excluded. The criteria for the two problem types—simple and complex—were consistent with previous research (e.g., Carr & Alexeev, 2011; Geary et al., 1997).

Each item was printed on a separate sheet of paper. Two sets of items (A and B) were compiled; they included the same problems but in different semi-random orders. Both sets began with a block of 6 simple addition problems, followed by a block of 12 complex problems, followed by another block of 6 simple problems. This order was intended to encourage children by presenting easier problems at the beginning of the task and to provide relief toward the end of the task. The two sets of items varied in the order of items within each block.

Procedure

The procedure involved presenting children with addition problems and asking them to solve each problem and then to explain how they arrived at the solution. Children were randomly assigned to receive Problem Set A or B. They were presented with one problem at a time and allowed as much time as needed to solve the problem. They were not provided with any supplies, such as pencil and paper, but were allowed to use fingers or count out loud if they wished. The experimenter observed the children and recorded any overt signs of strategy use (e.g., counting out loud, silently moving lips). When there were no overt behaviors, the tester asked the children how they "figured it out." Children's responses were audio-recorded and later transcribed.

Coding arithmetic strategies

Children's responses were coded following the strategy assessment scheme presented by Geary and colleagues (Geary et al., 1996, 2004). The data used in the coding process included children's explanations transcribed from audiotapes as well as the tester's notes made while the children were solving the problems and explaining the solutions. When the tester had not observed any overt behaviors, the coding relied primarily on the children's explanations. This kind of combination of behavioral observations during problem solving and retrospective self-reports has been found to lead to valid strategy classifications (Rittle-Johnson & Siegler, 1999; Siegler, 1987).

Strategies were coded into several categories: counting, decomposition, retrieval, refusal, and other. A strategy was coded as *counting* if the child verbally enumerated each unit in one or both of the addends or exhibited counting through behavioral cues (e.g., counting on fingers) while solving the problem or reported enumerating addend(s) when describing the solution. A strategy was coded as *decomposition* if the child reported breaking the problem apart into a series of smaller problems. A strategy was coded as *retrieval* if the child provided a quick response (within 3 sec) with no overt evidence of counting or decomposition and reported knowing the answer. If the child claimed to "just know" the answer but had not responded within approximately 3 sec, the experimenter prompted the child to explain some more. This usually resulted in a modified report of the strategy used (e.g., "Well, I counted part of it") or a description of breaking the problem apart and retrieving intermediate solutions from memory (i.e., decomposition). Similarly, if the child's overt behavior indicated use of a counting strategy, the strategy was coded as counting even if the child reported retrieval. There were, of course, instances where the child's reported strategy was unclear and could not be clarified by further prompting. These instances, along with cases where the child reported guessing, were coded as *other.* Any instance in which the child did not attempt the problem (e.g., saying that he or she did not know how to solve it) was coded as *refusal.* Two raters independently coded strategies on 25% of trials, and the level of agreement was 96%. The raters discussed the few cases where they initially disagreed and together arrived at the final code.

Results

First, we examined results for simple problems, analyzing gender effects in both strategies and accuracy within each country. Next, we conducted a parallel analysis of strategies and accuracy for complex arithmetic problems. Finally, based on the pattern of results obtained with complex problems, we conducted a mediation analysis to test whether gender differences in strategy choice accounted for gender differences in accuracy on this type of problems.

Simple problems

Strategies

Table 1 shows the frequency of using different strategies (retrieval, counting, decomposition, and other) for solving simple arithmetic problems. There were no "refusals" on any simple problems. As shown in the table, retrieval was the most frequent strategy in all three countries. There were gender differences in the use of each of the three main strategies, but the pattern varied by country. To examine the effects of gender and country statistically, we conducted analyses of variance (ANOVAs) with the outcome variable as the percentage of simple problems (out of 12) on which children used retrieval, counting, and decomposition strategies, respectively. Two independent variables—child's gender and country—were used as fixed factors.

With respect to retrieval, we found main effects of gender, $F(1, 459) = 31.33$, $p < .001$, $\eta^2 = .06$, and country, $F(2, 459) = 35.89$, $p < .001$, $\eta^2 = .14$, but no interaction ($p > .05$) between these two factors. The gender effect indicated that across countries boys used retrieval more frequently than girls. Follow-up pairwise comparisons exploring the main effect of country indicated that Taiwanese students used retrieval more frequently than Russian students ($p < .05$), whereas the difference between U.S. participants and students in the other two countries did not reach statistical significance (both *ps* > .05).

With respect to counting, we found a main effect of gender, $F(1, 459) = 17.62$, $p < .001$, $\eta^2 = .04$, moderated by country, $F(2, 459) = 19.49$, $p < .001$, $\eta^2 = .008$, as seen in the gender by country interaction, $F(2, 459) = 3.20$, $p < .05$, $\eta^2 = .01$. Simple effects tests indicated that in both the United States and

Table 1
Simple problems: percentage of trials on which each strategy was used.

	United States ($n = 92$, 42 boys)	Russia ($n = 98$, 48 boys)	Taiwan ($n = 60$, 28 boys)	Total ($n = 250$, 118 boys)
Retrieval				
Boys	49 (4)	50 (2)	73 (5)	55 (2)
Girls	43 (4)	37 (3)	54 (4)	44 (2)
Decomposition				
Boys	21 (3)	34 (2)	15 (4)	25 (2)
Girls	18 (3)	30 (4)	30 (4)	25 (2)
Counting				
Boys	27 (5)	11 (4)	11 (5)	17 (2)
Girls	37 (5)	26 (4)	14 (5)	27 (2)
Other				
Boys	1 (1)	5 (1)	1 (1)	3 (1)
Girls	2 (1)	6 (1)	2 (1)	3 (1)

Note: Standard errors are indicated in parentheses.

Russia girls used counting more frequently than boys ($p < .05$ and $p < .001$, respectively), whereas in Taiwan the use of this strategy did not vary by gender ($p > .05$).

With respect to decomposition, there was a main effect of country, $F(2, 459) = 14.08$, $p < .001$, $\eta^2 = .06$, which was qualified by a gender by country interaction, $F(2, 459) = 11.87$, $p < .001$, $\eta^2 = .05$. Simple effects tests indicated that Taiwanese girls used decomposition more frequently than boys ($p < .001$), whereas in the other two countries there were no gender differences in the use of this strategy ($p > .05$).

Accuracy

Next, we examined the accuracy of children's performance on simple problems. Fig. 1 shows the percentage of problems solved correctly by country and gender. As shown in the figure, all children performed at a very high level (>90%). Thus, the distribution of scores in all countries was highly skewed and the range of variability was extremely limited, especially in Taiwan, where girls and boys performed near or at ceiling (99% and 100% correct, respectively). Based on these data properties, we used non-parametric statistics (Mann–Whitney *U*) to examine gender effects. The results showed that there were no gender differences in accuracy on single-digit addition problems in any country (all *p*s > .05).

Complex problems

Strategies

Table 2 shows the frequency with which children used different strategies when solving complex arithmetic problems. Unlike in the case of simple problems, when presented with complex problems, some children refused to solve them. Thus, in analyzing performance on complex problems, we calculated the frequency of each strategy in two ways: as a proportion out of all problems presented and as a proportion of those problems attempted. The pattern of results obtained with both types of analyses was very similar. In Table 2, we report the frequency of each strategy out of all trials so as to show the frequency of refusals along with other strategies. In subsequent statistical analyses, we analyzed the frequencies of the main strategies computed as a proportion of attempted problems.

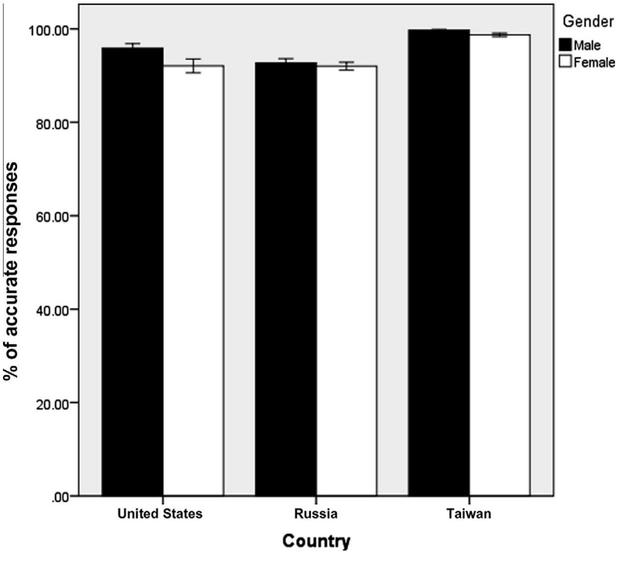

Fig. 1. Simple problem accuracy by gender and country.

Table 2
Complex problems: Percentage of trials on which each strategy was used.

	United States (n = 92, 42 boys)	Russia (n = 98, 48 boys)	Taiwan (n = 60, 28 boys)	Total (n = 250, 118 boys)
Retrieval				
Boys	0 (0)	0 (0)	21 (4)	5 (0)
Girls	0 (1)	0 (0)	6 (1)	1 (0)
Decomposition				
Boys	58 (9)	62 (9)	62 (12)	61 (6)
Girls	31 (4)	40 (6)	71 (13)	44 (4)
Counting				
Boys	33 (5)	17 (2)	15 (3)	22 (2)
Girls	56 (8)	40 (6)	20 (4)	41 (4)
Other				
Boys	9 (1)	17 (2)	2 (0)	11 (1)
Girls	0 (1)	12 (2)	3 (1)	9 (1)
Refusal				
Boys	0 (0)	4 (1)	0 (0)	2 (0)
Girls	3 (0)	9 (1)	1 (0)	5 (0)

Note: Standard errors are indicated in parentheses.

As shown in Table 2, very few responses on complex problems were coded as retrieval, which is not surprising given children's limited experience with computations involving double-digit numbers. One group of students who produced a higher percentage of such responses was Taiwanese boys. It is not clear, however, whether these students indeed retrieved a stored number fact from memory when solving complex problems or instead carried out mental computations (e.g., using a decomposition strategy) very quickly—an issue we return to in the Discussion.

With respect to the other two strategies—counting and decomposition—the data presented in Table 2 suggest that there were gender differences in strategy choice and, furthermore, that the pattern of gender differences varied across countries. To examine the effects of gender and country statistically, we conducted two ANOVAs with two dependent variables: the percentage of attempted problems on which children used a counting strategy and the percentage of attempted problems on which children used decomposition, respectively. The independent variables were gender (boys or girls) and country (United States, Russia, or Taiwan).

With respect to counting, there was a significant effect of gender, $F(1, 459) = 38.26$, $p < .001$, $\eta^2 = .07$, and a main effect of country, $F(2, 459) = 24.05$, $p < .001$, $\eta^2 = .09$. The effect of gender was moderated by country, as seen in the gender by country interaction, $F(2, 459) = 6.00$, $p < .01$, $\eta^2 = .02$. To better understand the nature of the interaction, tests of simple effects were conducted. They revealed significant gender differences in the use of a counting strategy in two of the countries: United States and Russia (both $ps < .001$). In contrast, there was no gender difference in the use of counting among the Taiwanese students ($p > .05$). With respect to the decomposition strategy, a parallel pattern of findings was observed. Specifically, there was a main effect of gender, $F(1, 459) = 16.13$, $p < .001$, $\eta^2 = .071$, and a main effect of country, $F(2, 459) = 13.65$, $p < .001$, $\eta^2 = .09$. Furthermore, the effect of gender differed by country, as seen in the significant gender by country interaction, $F(2, 459) = 12.45$, $p < .001$, $\eta^2 = .022$. Tests of simple effects revealed significant gender differences in the use of a decomposition strategy in two of the countries: United States and Russia (both $ps < .001$). In contrast, there was no gender difference in the use of decomposition among the Taiwanese students ($p > .05$).

Accuracy

Next, we examined the accuracy of children's performance on complex problems. Fig. 2 presents accuracy results by country and gender. As can be seen in the figure, the accuracy of Taiwanese students—both boys and girls—is very high, which makes the distribution of scores skewed and restricts the range of variability, leading to violations of the ANOVA assumptions and decreasing the power of detecting interaction effects (McClelland & Judd, 1993). Thus, similar to the analysis of accuracy on

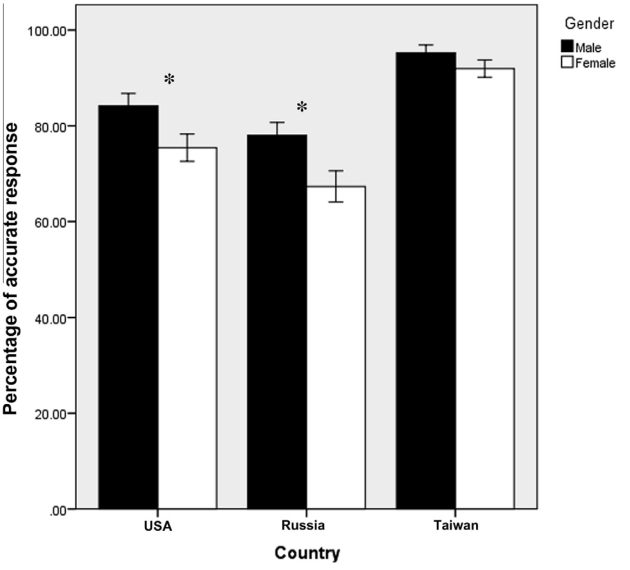

Fig. 2. Complex problem accuracy by country and gender. *$p < .01$.

simple problems, we used non-parametric analyses (Mann–Whitney *U*) to examine gender differences in each of the countries. These analyses showed significant gender differences in the United States and Russia (both *p*s < .05), with boys outperforming girls, and no difference in accuracy between the Taiwanese boys and girls ($p = .16$).

Mediation analysis

The analysis of performance on complex addition problems showed gender differences in strategy use among the students from the United States and Russia and corresponding differences in accuracy; on the other hand, boys and girls in Taiwan did not differ from each other either in the frequency of using particular strategies or in accuracy. We conducted a bias-corrected bootstrapping mediation analysis (Dearing & Hamilton, 2006; Preacher & Hayes, 2008) to examine whether the use of a decomposition strategy mediated gender differences in accuracy on all complex problems among American and Russian children. We reasoned that if the use of decomposition on a particular problem leads to a greater probability of solving that problem correctly, a higher frequency of using decomposition across all problems would lead to higher overall accuracy. The mediation analysis allowed us to determine whether gender differences in the use of decomposition accounted for the observed differences in accuracy.

To conduct the analysis, the SPSS macro "PROCESS," created by Preacher and Hayes (2004), was used to randomly select 10,000 samples with replacement from the complete data file. Regression coefficients were estimated for each of the bootstrap samples and averaged across all samples. This method allowed for detection of direct and indirect effects (i.e., mediation) of gender on arithmetic accuracy via decomposition strategies. The indirect effect is considered to be statistically significant if the 95% confidence interval (CI) does not include zero (Preacher & Hayes, 2008). The significance of mediation effects is based on the estimates of confidence intervals and is the recommended analytic method for smaller samples. The results of the mediation analysis are presented in Fig. 3.

Prior to entering the mediation term, the regression analysis showed that gender was a significant predictor of accuracy on complex problems, $c = 0.097$, $s_c = 0.03$, $p < .01$. When both gender and decomposition strategy were entered into the regression model, the results showed that decomposition predicted accuracy, $b = 0.26$, $s_b = 0.04$, $p < .001$, whereas gender was no longer a significant predictor,

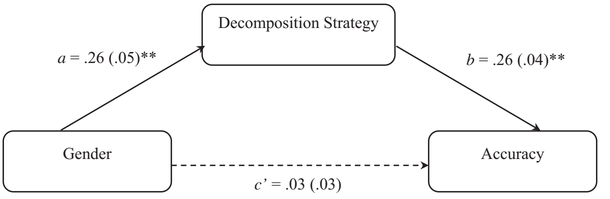

Fig. 3. Decomposition strategy as a mediator of the relation between gender and accuracy on complex arithmetic problems. Beta coefficients are reported (with standard errors in parentheses). $^{**}p < .001$.

$c' = 0.03$, $s_{c'} = 0.03$, $p = .37$. Thus, the findings indicate that there is a significant indirect effect of gender on arithmetic accuracy via decomposition, 95% CI [0.037, 0.104].

Discussion

The current study integrated several existing lines of research, including cross-national comparisons of math performance (Geary et al., 1996; Mullis, Martin, Foy, & Arora, 2012), investigations of gender differences in strategy use (Carr & Davis, 2001; Carr et al., 2008), and research on the relation between gender and problem complexity (Gibbs, 2010; Penner, 2003). The findings highlight the intricate nature of the phenomenon of gender differences in mathematics. Rather than observing an overall difference between boys and girls, we found that the pattern of their performance varied as a function of both problem characteristics and educational/cultural context. The results suggest that gender differences in arithmetic accuracy can be attenuated with instructional experience.

Differences in the pattern of gender findings across countries

Accuracy

Whereas previous studies that examined accuracy on math tasks with older students reported cross-national variability in gender findings (Bedard & Cho, 2010; Byrnes et al., 1997), the few studies that involved elementary school students did not (e.g., Geary et al., 1996). In fact, these studies with younger students, which typically included simple math tasks, did not even find within-country gender differences. Similarly, when examining students' accuracy on simple arithmetic problems in the current investigation, we found no gender differences in any of the participating countries. However, unlike other cross-national studies of younger students, our investigation also included complex arithmetic problems. This approach produced a more nuanced picture of performance, revealing cross-national variability in gender patterns even among elementary school students.

The key difference was observed between Taiwanese participants and children in the other two countries. In particular, the findings obtained with Russian and American students showed that boys outperformed girls in their respective countries when solving complex arithmetic problems. On the other hand, among Taiwanese participants, no gender differences were found in the level of accuracy on such problems. These results parallel some of the previous findings with older children, indicating a smaller gender gap in mathematics achievement among Asian students compared with American students (Byrnes et al., 1997; Cai, 1995). The observed pattern of gender differences in accuracy calls for a closer analysis of underlying mathematical reasoning in boys and girls in different countries that may account for the recurring finding. The focus on arithmetic strategies in the current study allows us to address this issue.

Strategies

Unlike with accuracy, gender differences in strategy use were found on both simple and complex problems. On simple problems, boys tended to use retrieval more often than girls. Furthermore, there was no interaction between gender and country in the use of this strategy. For example, whereas Taiwanese students used retrieval more frequently than Russian students, the pattern of gender differences was the same in both groups. Thus, the current cross-national findings were consistent with previous research examining American children that demonstrated boys' preference for retrieval in solving single-digit arithmetic problems (Bailey et al., 2012).

Although girls generally used retrieval less often than boys, the strategy use among girls on simple problems varied across countries. Taiwanese girls were more likely than American and Russian girls to use retrieval. Furthermore, when they did not use retrieval, they tended to use decomposition and did so more often compared with Taiwanese boys. On the other hand, American and Russian girls used counting strategies most frequently when they used a strategy other than retrieval and did so more often compared with boys in their respective countries.

On complex problems, girls' tendencies persisted. The use of decomposition strategies increased between simple and complex problems among Taiwanese girls, whereas the use of counting strategies increased among American and Russian girls. In contrast, boys in all three countries shifted from using predominantly retrieval on simple problems to using decomposition most frequently on complex problems. As a result, there was a gender difference in strategy use in the United States and Russia, with boys using decomposition more often compared with girls in their respective countries, but no gender difference in Taiwan, where boys and girls were comparable in their use of decomposition strategies.

The shift from retrieval on simple arithmetic problems to alternative strategies on complex problems is not surprising given that retrieval of the sums of complex problems is seldom possible. Even adults most likely do not memorize all of the number facts involving double digits in the same manner as with single digits. Rather, adults typically solve complex arithmetic problems by breaking them into components involving basic number facts that can be retrieved from memory—a kind of decomposition. Yet, it should be noted that when examining Taiwanese students' strategies on complex problems, 20% of their responses were coded as retrieval because the answer was produced quickly (within 3 s), there was no behavioral evidence of using counting, and when asked to explain their solutions the children replied "I just knew it." Most such responses were produced by Taiwanese boys. It is possible that extensive computational practice allowed these children to commit some of the more complex number facts to memory. However, it is also possible that some of the instances that were categorized as retrieval actually presented very quickly executed calculations. Our data do not allow us to differentiate between these possibilities. What is clear based on the data, however, is that some Taiwanese students—predominantly boys—demonstrated a high degree of fluency not only with basic computations but also with more complex computations.

Relation between strategy use and accuracy

Can boys' and girls' tendencies to use different strategies explain gender differences in accuracy on arithmetic problems? The current data suggest that these tendencies may matter only in the context of complex problems. Although boys and girls differed in their preferred strategy on simple problems, this was not reflected in gender differences in accuracy in any country. This finding is consistent with previous research that has found that backup strategies, such as counting, may be less efficient but are likely to lead to accurate responses on simple arithmetic problems (Shrager & Siegler, 1998). In fact, Siegler's (1988) strategy choice model posits that using backup strategies is actually an adaptive means for children to maximize accuracy when most efficient strategies are not available.

In contrast, strategy preferences seem to have a more profound impact on accuracy when solving complex problems. In the context of these problems, where addends are much larger and cannot be easily represented on fingers, the use of counting is not only inefficient but also likely to lead to errors. Decomposition, on the other hand, may present a more efficient approach, enabling students to reduce a complex problem to simple problems that can be solved without the use of counting. Thus, a tendency to use decomposition, rather than counting, on complex problems is likely to be related to accuracy on those problems and could account for gender differences in accuracy.

Indeed, the current data are consistent with this hypothesis. Taiwanese boys and girls demonstrated a comparably high frequency of using decomposition on complex problems, and no gender differences in accuracy were found between them. In contrast, in the two countries where boys relied more on decomposition and girls relied more on counting when solving complex problems (United States and Russia), boys demonstrated higher accuracy than girls. Furthermore, our analysis showed that the use of decomposition on complex problems mediated the gender differences observed on this type of problems. These results extend the recent findings on the critical role of decomposition in solving arithmetic problems by providing, for the first time, statistical evidence that gender differences in the use of decomposition fully account for gender differences in arithmetic accuracy (Carr et al., 2008; Fennema et al., 1998; Geary, Hoard, Nugent, & Bailey, 2013; Geary et al., 2004).

Implications for understanding the nature of gender differences

A key finding to emerge from the current study is the importance of gender differences in strategy choice for understanding gender differences in accuracy on mathematics tasks. This finding, in turn, leads to questions concerning the sources of gender variability in strategy preferences. For example, when looking at the performance of American students, why were boys more likely to choose decomposition, whereas girls from the same classrooms tended to rely more on counting when solving complex arithmetic problems?

One possibility is that predispositions to use particular strategies reflect general cognitive strengths that vary as a function of gender. Although there are many inconsistencies in findings on gender differences in cognitive skills, there appears to be a consensus that boys tend to have an advantage on spatial tasks, especially those involving visualization and mental transformation, whereas girls perform as well or better on verbal/analytic reasoning tasks (Halpern et al., 2007; Hyde, 2005; Voyer, Voyer, & Bryden, 1995). Perhaps differential preferences for particular computational strategies are in part related to these well-documented tendencies in cognitive skills. Interestingly, recent findings on the relation between spatial reasoning and mathematics performance parallel our findings on the relation between strategies and math performance. For example, when presented with measurement problems, boys tended to rely on spatial visualization more than girls, who relied more on analytic reasoning (Vasilyeva et al., 2009). This difference was not related to accuracy of performance on standard tasks that could be solved using known algorithms, but it was associated with boys' advantage on more challenging novel tasks where analytic approaches either were unavailable or became extremely inefficient.

The parallel findings in work on math strategies and spatial reasoning raises the possibility that the differences between boys and girls in the preferred strategy may be in part related to differences in spatial skill or at least in the extent to which it is used in problem solving. Consistent with this explanation, recent studies with elementary school students have demonstrated that a decomposition strategy is chosen more frequently by children with higher visuospatial memory capacity (Dulaney, Vasilyeva, & Laski, 2016; Laski et al., 2013). The authors suggested that children with greater spatial skills may be better at visualizing the relative magnitudes of numbers and manipulating them along a mental number line, which in turn helps them to implement a decomposition strategy (i.e., breaking larger numbers into smaller numbers).

If differences in spatial abilities lead to differential preferences for particular strategies, and if we assume, based on existing literature, that boys have greater spatial skills, it may provide a reasonable explanation for the findings obtained in the American and Russian samples. However, this explanation cannot account for the observed cross-national differences in gender patterns and, in particular, for the performance of the Taiwanese students. The existence of cross-national variability in gender differences points to a potential role of educational experiences in explaining strategy preferences observed in the current study. Especially informative in this context was the performance of Taiwanese girls, who used retrieval more frequently on simple problems and decomposition more frequently on complex problems than American and Russian girls. Furthermore, they used decomposition on complex problems about as frequently as Taiwanese boys, such that there were no gender differences in the use of this strategy in Taiwan.

As noted earlier, the curriculum used by our Taiwanese participants provided more intensive practice with basic number facts and a stronger emphasis on the use of decomposition compared with American and Russian early math instruction. This observation is consistent with other reports indicating that Asian students (e.g., Chinese, Korean) receive a large amount of computational practice, facilitating their memorization of number facts, and that instruction in arithmetic focuses explicitly on the use of most efficient strategies (Perry, 2000; Wang & Lin, 2005). Both aspects of instruction are likely to play a role in children's performance on complex arithmetic tasks. It should be noted that a decomposition strategy requires a high degree of fluency with basic number facts (DeStefano & LeFevre, 2004; Imbo & Vandierendonck, 2007). Consider the problem 36 + 27; if the child adds the tens and then begins to add the ones using either counting or further decomposition, the sum of the tens may disappear from memory. However, if the child can quickly and accurately retrieve a relevant single-digit addition fact, there is a better chance that final addition steps will be completed successfully. Thus, the explicit focus on decomposition may facilitate girls' use of this strategy in computations, and a substantial amount of computational practice may further increase their confidence in using this strategy.

The overall pattern of within- and between-country gender differences found in the current study suggests that strategy preferences are due to both gender-related predispositions and educational experience. In other words, even if boys and girls differ in certain aspects of mathematical skills or cognitive processes, instruction can minimize or eliminate the subsequent differences in performance. This idea is consistent with the finding of no gender differences in the use of decomposition on complex problems in Taiwan, where this strategy was highly emphasized in instruction. Furthermore, the finding of gender differences in the use of decomposition in the United States and Russia, where there was not emphasis on this strategy, suggests that without intense instruction and when given a free choice, children are more likely to fall back on their original preferences.

These initial predispositions could reflect either biologically based differences or variability in earlier experience. Research in the area of spatial development has implicated both biological factors (e.g., differences in sex hormones; Williams, Barnett, & Meck, 1990) and sociocultural factors (e.g., differential exposure to toys and activities stimulating spatial thinking; Newcombe, Bandura, & Taylor, 1983) in gender differences. Levine and colleagues conducted a study of spatial skills in elementary school students and established that the pattern of gender differences varied as a function of socioeconomic status (SES) (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). The authors pointed out that although the observed variability across SES groups was not consistent with biological explanations of the male advantage, it was consistent with an environmental explanation. That is, differential engagement in activities that promote spatial skills may lead to a male advantage, but only to the extent that such activities are available in children's environment.

Whether the origin of the initial gender differences in cognitive predispositions is biological or environmental, or both, the critical issue is the malleability of the cognitive skills in question—the extent to which they can be improved through training. In the domain of spatial reasoning, there is accumulating evidence that skills in general and visuospatial memory in particular can be increased through specially designed interventions (Uttal et al., 2013; Wright, Thomson, Ganis, Newcombe, & Kosslyn, 2008). Furthermore, it has been shown that spatial training might not only improve the performance of all participants but also provide greater benefits to those participants who had lower performance in the beginning and, as a result, decrease the gender gap (Tzuriel & Egozi, 2010). Similarly, it is possible that the use of efficient strategies can be facilitated directly through instruction and that instruction focusing on such strategies may reduce differences in performance between boys and girls. Of course, currently this is only speculation that is consistent with the cross-national findings of the current study. The non-experimental nature of this study does not allow us to make definitive causal statements. Yet, the current findings raise hypotheses about a potential role of instruction that could be explored in future intervention studies. Interventions that start at the elementary school level may have a cumulative effect by increasing girls' confidence and their participation in more challenging math classes later in school.

References

Aunio, P., Niemivirta, M., Hautamäki, J., Van Luit, J. E., Shi, J., & Zhang, M. (2006). Young children's number sense in China and Finland. *Scandinavian Journal of Educational Research, 50*, 483–502.

Ayalon, H., & Livneh, I. (2013). Educational standardization and gender differences in mathematics achievement: A comparative study. *Social Science Research, 42*, 432–445.

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). 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. *Journal of Experimental Child Psychology, 113*, 78–92.

Bedard, K., & Cho, I. (2010). Early gender test score gaps across OECD countries. *Economics of Education Review, 29*, 348–363.
Byrnes, J. P., Hong, L., & Xing, S. (1997). Gender differences on the math subtest of the Scholastic Aptitude Test may be culture-specific. *Educational Studies in Mathematics, 34*(1), 49–66.
Cai, J. (1995). A cognitive analysis of U.S. and Chinese students' mathematical performance on tasks involving computation, simple problem solving, and complex problem solving. *Journal for Research in Mathematics Education* (Monograph No. 7, National Council of Teachers of Mathematics, Reston, VA).
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.
Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*, 61–75.
Dearing, E., & Hamilton, L. C. (2006). Contemporary advances and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development, 71*(3), 88–104.
DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386.
Dulaney, A., Vasilyeva, M., & Laski, E. (2016). Mediators of the working memory and arithmetic relation in children: Comparing visuospatial and verbal memory. Manuscript submitted for publication.
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 103–127.
Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher, 27*(5), 6–11.
Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology, 49*, 1886–1899.
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., Hamson, C. O., Chen, G. P., Liu, F., Hoard, M. K., & Salthouse, T. A. (1997). Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States. *Psychonomic Bulletin & Review, 4*, 425–430.
Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, 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., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS One, 8*(1), e54651.
Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research, 39*, 540–569.
Halpern, D. F., Benbow, C., Geary, D., Gur, D., Hyde, J., & Gernsbacher, M. A. (2007). The science of sex-differences in science and mathematics. *Psychological Science in the Public Interest, 8*(1), 1–52.
Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.
Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*, 284–309.
Istomina, N. B. (2011). *Elementary school of the XXI century: Mathematics/Grade 1*. Moscow, Russia: Ventana-Graf.
Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science, 11*, 662–668.
Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences, 23*, 123–130.
Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16*, 841–845.
McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376–390.
Ministry of Education of Taiwan. (2011). *The nine-year school curriculum: First-grade elementary school* (textbook Vol. 1). Taiwan: Taipei. Retrieved from: <http://mathtext.project.edu.tw>.
Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center.
Newcombe, N., Bandura, M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles, 9*, 377–386.
Penner, A. M. (2003). International gender × item difficulty interactions in mathematics and science achievement tests. *Journal of Educational Psychology, 95*, 650–655.
Perry, M. (2000). Explanations of mathematical concepts in Japanese, Chinese, and U.S. first- and fifth-grade classrooms. *Cognition and Instruction, 18*, 181–207.
Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*, 717–731.
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.
Rittle-Johnson, B. R., & Siegler, R. S. (1999). Learning to spell: Variability, choice, and change in children's strategy use. *Child Development, 70*, 332–349.
Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405–410.
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.

TERC. (2008). *Investigations in number, data, and space* (2nd ed.). Glenview, IL: Scott Foresman.

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81*, 1417–1430.

University of Chicago School Mathematics Project. (2012). *Everyday mathematics: Grade 1*. Chicago: Wright Group/McGraw-Hill.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*, 352–402.

Vasilyeva, M., Casey, B. M., Dearing, E., & Ganley, C. M. (2009). Measurement skills in low-income elementary school students: Exploring the nature of gender differences. *Cognition and Instruction, 27*, 401–428.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270.

Wang, J., & Lin, E. (2005). Comparative studies on U.S. and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher, 34*(5), 3–13.

Williams, C. L., Barnett, A. M., & Meck, W. H. (1990). Organization effects of early gonadal secretions on sexual differentiation in spatial memory. *Behavioral Neuroscience, 104*, 84–97.

Wright, R., Thomson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15*, 763–771.