UC Irvine
UC Irvine Previously Published Works

Title

Gender Differences in Children's Arithmetic Performance Are Accounted for by Gender Differences in Language Abilities

Permalink

https://escholarship.org/uc/item/703388gz

Journal

Psychological Science, 23(3)

ISSN

0956-7976

Authors

Wei, Wei
Lu, Hao
Zhao, Hui
et al.

Publication Date

2012-03-01

DOI

10.1177/0956797611427168

Copyright Information

Peer reviewed

Gender Differences in Children's Arithmetic Performance Are Accounted for by Gender Differences in Language Abilities

Wei Wei[1], Hao Lu[1], Hui Zhao[1], Chuansheng Chen[2], Qi Dong[1], and Xinlin Zhou[1]

[1]National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, and [2]Department of Psychology and Social Behavior, University of California, Irvine

Abstract

Studies have shown that female children, on average, consistently outperform male children in arithmetic. In the research reported here, 1,556 pupils (8 to 11 years of age) from urban and rural regions in the greater Beijing area completed 10 cognitive tasks. Results showed that girls outperformed boys in arithmetic tasks (i.e., simple subtraction, complex multiplication), as well as in numerosity-comparison, number-comparison, number-series-completion, choice reaction time, and word-rhyming tasks. Boys outperformed girls in a mental rotation task. Controlling for scores on the word-rhyming task eliminated gender differences in arithmetic, whereas controlling for scores on numerical-processing tasks (number comparison, numerosity estimation, numerosity comparison, and number-series completion) and general cognitive tasks (choice reaction time, Raven's Progressive Matrices, and mental rotation) did not. These results suggest that girls' advantage in arithmetic is likely due to their advantage in language processing.

There is an extensive literature on gender differences in mathematical performance. Boys have been shown to outperform girls in mathematical problem solving (e.g., Benbow & Stanley, 1983; Geary, 1996; Hyde, Fennema, & Lamon, 1990; for reviews, see Hyde, Lindberg, Linn, Ellis, & Williams, 2008; Zhu, 2007), but girls generally outperform boys in arithmetic (e.g., Linn & Hyde, 1989; Willingham & Cole, 1997; but see Fennema, Carpenter, Jacobs, Franke, & Levi, 1998, and Hyde et al., 1990, for some exceptions). Boys' advantage in mathematical problem solving has been attributed to their superior spatial abilities (e.g., Casey, Nuttall, & Pezaris, 1999; Geary, 1996). The reasons for girls' advantage in arithmetic, however, are not yet understood.

One plausible reason for this advantage is that girls have an early advantage in verbal abilities (e.g., Burman, Bitan, & Booth, 2008; Hyde & Linn, 1988; Maccoby & Jacklin, 1974). Maccoby and Jacklin (1974) remarked that "female superiority on verbal tasks has been one of the more solidly established generalizations in the field of sex differences" (p. 75). Compared with boys, girls experience an earlier onset of verbal ability and faster vocabulary acquisition, have better reading skills, use more word roots, and speak in longer utterances

(Bornstein, Haynes, Painter, & Genevro, 2000; Roulstone, Loader, Northstone, & Beveridge, 2002), and girls' advantage in reading skills appears to be consistent throughout the high-school years (Mann, Sasanuma, Sakuma, & Masaki, 1990). The risk of language disorders also varies by gender. Flannery, Liederman, Daly, and Schultz (2000) concluded that language and reading disorders are approximately twice as common in boys as they are in girls.

Arithmetic performance is highly dependent on language processing (e.g., Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). For example, exact calculation involves more language processing than does approximate calculation (Dehaene et al., 1999). Arithmetic calculation elicits greater activations in language-processing areas (including the left superior temporal gyrus, the left precentral gyrus, the left inferior frontal gyrus, and the motor area) than does mathematical problem solving (Lu et al., 2009). Because of the importance of

Corresponding Author:

Xinlin Zhou, National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
E-mail: zhou_xinlin@bnu.edu.cn

language processing in arithmetic, it is plausible that girls' advantage in language processing explains their advantage in arithmetic. We tested this hypothesis in the current study.

First, we tested for gender differences in simple calculation (i.e., subtraction using numbers smaller than 20) and complex calculation (i.e., multiplication involving double-digit factors; e.g., 73×2). We expected that girls would outperform boys in both tasks.

Second, we examined gender differences in arithmetic performance after statistically controlling for each of the following factors separately: basic numerical processing (including numerosity comparison, numerosity estimation, and number comparison), number-series completion, word rhyming, reaction time, mental rotation, and Raven's Progressive Matrices. We used a word-rhyming task to assess language processing because it measured a critical component of language ability: phonological awareness (Anthony et al., 2002; Bradley & Bryant, 1983). We expected that we would find no significant differences in arithmetic performance after controlling for performance on this task.

Finally, in order to investigate whether performance on all tasks (i.e., both the general cognitive tasks and the numerical-processing tasks), with the exception of the word-rhyming task, could account for gender differences in arithmetic, we conducted an analysis that included performance on all tasks except for word rhyming as covariates. We expected that gender differences would remain significant after we controlled for scores on the general-ability and numerical-processing tasks.

Method
Participants

Third- to sixth-grade pupils from 12 primary schools in the greater Beijing area of China ($N = 1,556$; 803 males, 753 females; 8–11 years old) participated in this study. Six schools were urban schools, and 6 were rural schools. Four classes of pupils (one class per grade; approximately 30 to 40 children per class) were randomly selected from each school. All participants were native Chinese speakers and had normal or corrected-to-normal eyesight. The study was approved by the Institute of Cognitive Neuroscience and Learning at Beijing Normal University, the administrative departments of education in the relevant counties, and the principals of the schools.

Procedure

Tests were administered to each class of students in a computer classroom. Each class was monitored by either six or seven experimenters (4 to 6 children per experimenter) as well as the teacher of that class. Instructions were given and a practice session was completed before each formal test. The tasks were administered in the same order for all students. For 9 of the 10 tasks, the children indicated their responses by pressing

one of two keys on a computer keyboard; for the numerosity-estimation task, they entered a numerical value. Students' responses were automatically recorded and sent over the Internet to a server located in our laboratory at Beijing Normal University. All data were collected between November 12, 2009, and December 24, 2009.

The practice session for each task consisted of either four or six trials, which were similar to those used in the formal test. The computer provided the child with feedback after each practice trial. For 9 of the 10 tasks, feedback for correct responses was "Correct! Can you go faster?" and feedback for incorrect responses was "It is wrong. Try again." For the numerosity-estimation task, in which the children had to estimate the number of dots in a dot array, the feedback was the correct number of dots. The children could ask experimenters any questions they had during the practice session. After all the children in a class had finished the practice session and had no more questions for the experimenters, the main experimenter said, "Start," and the children pressed any key to begin the formal test.

Tasks

All the tasks were programmed using Web-based applications available at www.dweipsy.com. For each of the time-limited tasks (i.e., simple subtraction, complex multiplication, number-series completion, mental rotation, and Raven's Progressive Matrices), we calculated scores by subtracting the number of incorrect responses from the number of correct responses. For each of the timed tasks (i.e., choice reaction time, numerosity comparison, number comparison, and word rhyming), we calculated each participant's median reaction time and error rate. For the numerosity-estimation task, we calculated the absolute value of the deviation of the estimated quantity from the correct quantity for each trial and then calculated the mean of the absolute values. All tests appeared to have acceptable reliability (see Table 1).

Simple subtraction. For all 92 simple-subtraction problems (e.g., $6 - 2$, $17 - 8$), the minuends were 18 or smaller, and the differences were single-digit numbers. Two candidate answers were presented beneath each problem. Participants were asked to press the "Q" key to choose the answer on the left and the "P" key to choose the answer on the right. For this task, each incorrect candidate answer was within the range of the correct answer plus or minus 3 (i.e., ± 1, ± 2, or ± 3). The children were allotted 2 min to complete this task.

Complex multiplication. All 76 problems in the multiplication task involved one double-digit number multiplied by one single-digit number (e.g., 67×9). Every problem required carrying. Four candidate answers were presented beneath each problem: the correct answer and three incorrect candidate answers (i.e., the correct answer plus or minus 1, the correct answer plus or minus 10, and the correct answer plus or minus

Table 1. Mean Results and Gender Differences for All Tasks

| | Younger children (ages 8–9) | | | | Older children (ages 10–11) | | | | Tests of gender differences (multilevel model) | | Split-half reliability |
| | Urban | | Rural | | Urban | | Rural | | | | |
Task	Male	Female	Male	Female	Male	Female	Male	Female	b	t	
Simple subtraction	37.7 (9.5)	38.77 (8.5)	35.6 (8.3)	35.8 (8.1)	42.0 (8.3)	44.0 (7.9)	39.7 (9.0)	40.9 (7.2)	1.19	$t(1510) = 2.97$**	.96
Complex multiplication	17.4 (7.1)	18.3 (6.5)	17.0 (5.4)	18.0 (5.8)	23.7 (6.1)	25.6 (5.4)	22.3 (6.2)	23.0 (5.7)	0.93	$t(1502) = 3.30$***	.92
Word rhyming (RT)	2,435 (941)	2,350 (691)	2,384 (971)	2,731 (819)	2,330 (745)	2,201 (598)	2,499 (777)	2,465 (648)	16.02	$t(1533) = 0.41$	.95
Word rhyming (error rate)	.29 (.16)	.23 (.16)	.37 (.15)	.34 (.16)	.21 (.15)	.13 (.12)	.27 (.15)	.22 (.16)	−0.06	$t(1516) = -7.59$***	.71
Numerosity comparison (RT)	926 (229)	864 (164)	905 (207)	885 (183)	810 (158)	778 (153)	799 (154)	780 (159)	−33.62	$t(1518) = -3.85$***	.80
Numerosity comparison (error rate)	.13 (.10)	.13 (.09)	.12 (.08)	.13 (.10)	.10 (.08)	.09 (.07)	.10 (.09)	.10 (.08)	−0.002	$t(1530) = -0.40$	.76
Numerosity estimation	20.3 (9.9)	21.0 (9.9)	23.4 (10.8)	23.5 (10.4)	18.2 (8.2)	17.5 (6.6)	19.1 (7.7)	18.6 (7.1)	0.07	$t(1529) = 0.30$	.74
Number comparison (RT)	754 (157)	768 (140)	753 (135)	782 (128)	679 (122)	672 (119)	673 (127)	668 (110)	5.75	$t(1510) = -0.91$	.93
Number comparison (error rate)	.10 (.07)	.09 (.06)	.10 (.05)	.10 (.05)	.08 (.05)	.07 (.04)	.08 (.05)	.08 (.05)	−0.01	$t(1538) = -2.87$**	.80
Number-series completion	12.1 (4.8)	12.6 (4.2)	11.0 (3.9)	14.8 (4.0)	13.4 (4.6)	11.4 (3.9)	12.1 (4.1)	12.6 (3.8)	0.68	$t(1522) = 3.27$**	.71
Choice reaction (RT)	467 (117)	462 (124)	493 (125)	496 (121)	412 (97)	405 (92)	434 (131)	446 (177)	1.16	$t(1519) = 0.19$	.73
Choice reaction (error rate)	.06 (.10)	.05 (.08)	.05 (.09)	.03 (.04)	.03 (.05)	.03 (.05)	.05 (.09)	.04 (.08)	−0.01	$t(1542) = -3.08$**	.91
Mental rotation	18.4 (8.6)	17.2 (6.9)	16.7 (7.5)	16.2 (6.2)	21.8 (9.0)	20.3 (8.0)	20.3 (8.2)	18.6 (7.6)	−1.28	$t(1537) = -3.21$**	.87
Raven's Progressive Matrices (Raven, 1998)	19.6 (6.4)	18.9 (6.7)	17.0 (6.5)	17.0 (6.4)	21.8 (6.0)	21.3 (6.1)	17.8 (6.1)	18.5 (6.4)	−0.13	$t(1526) = -0.41$	.83

Note: Reaction times (RTs) are given in milliseconds. Gender was coded as 0 for male and 1 for female. Standard deviations are shown in parentheses.
**$p < .01$. ***$p < .001$.

100). The other aspects of the procedure (stimuli presentation, method of responding) were the same as those for the simple-subtraction task. Children were allotted 2 min to complete this task.

Numerosity comparison. The numerosity-comparison task was adapted from the second edition of the Test of Early Mathematics Ability (Ginsburg & Baroody, 1990). Two sets of dots of varying sizes were presented simultaneously on the screen, and participants were asked to judge which dot array contained more dots while ignoring the sizes of individual dots. Participants pressed "Q" if they thought the array on the left contained more dots and "P" if they thought the array on the right contained more dots. The number of dots in each set varied from 5 to 12. The total combined area of all dots in each set was controlled to be the same. The test consisted of 36 trials.

Numerosity estimation. The numerosity-estimation task was adapted from Krueger (1982). Participants were asked to quantify the number of dots in an array of 11 to 99 dots. Each dot array was presented in the middle of the screen for 1,000 ms. Participants entered a value into a box at the bottom of the screen. After each trial, participants received feedback showing the correct number of dots and were asked to use the feedback to improve their estimation on subsequent trials. The test consisted of 27 trials.

Number comparison. The number-comparison task was adapted from a Stroop-like number-comparison task used in previous research (Girelli, Lucangeli, & Butterworth, 2000; Zhou, Chen, Chen, et al., 2007). Eighty-four pairs of single-digit Arabic numbers of varying sizes were presented in a random order. For each pair, participants were asked to decide which number was larger in numerical magnitude, while ignoring the differences in the physical size of the numbers. Participants pressed the "Q" key to choose the answer on the left and the "P" key to choose the answer on the right. The magnitude of the numbers could be congruent, incongruent, or neutral with respect to the physical sizes of the numbers (e.g., if the pair of numbers was "3-8," the 3 could be physically smaller than the 8, physically larger than the 8, or the same size as the 8, respectively). In pairs of differently sized numbers, the ratio of the physical size of the two numbers was 1:2. The number-comparison test consisted of three sessions (28 trials per session), separated by two 30-s rest periods.

Number-series completion. The number-series-completion test was adapted from the Cognitive Abilities Test 3 (Smith, Fernandes, & Strand, 2001). A series of numbers was presented in the middle of the screen. Participants were asked to detect the pattern of these numbers and deduce the next number in the series. For example, the series of numbers "2, 4, 6, 8" would be followed by 10. Two candidate answers were presented beneath each number sequence. Participants pressed the "Q" key to choose the answer on the left and the "P" key

to choose the answer on the right. The children were allotted 4 min to complete this test.

Choice reaction time. On each trial of the choice reaction time test, a white dot was presented on a black screen, either to the left or to the right of a fixation cross. The position of the dot was within 15° of visual angle from the cross. Participants were asked to press the "Q" key if the dot appeared on the left and the "P" key if it appeared on the right. There were 30 trials in total (15 trials with the dot on the left and 15 trials with the dot on the right); the size of the screen on which the dot appeared varied randomly across trials. Interstimulus intervals varied randomly between 1,500 ms and 3,000 ms.

Word rhyming. The word-rhyming task consisted of 40 trials and was similar to the task used by Tan et al. (2001). Two Chinese characters were presented simultaneously on the screen. Participants were asked to indicate whether or not the two characters rhymed by pressing the "Q" key for rhyming pairs and the "P" key for nonrhyming pairs. The stimuli remained on the screen until participants responded or 4 s had lapsed.

Mental rotation. The mental rotation task was based on the mental rotation task used by Shepard and Metzler (1971). On each trial, one three-dimensional image was presented on the upper part of the screen, and two more were presented on the lower part of the screen. Participants were asked to choose which image from the bottom of the screen matched the image at the top; the matching image could be identified only by mental rotation. Participants pressed the "Q" key to choose the image on the left and the "P" key to choose the image on the right. The mental rotation test consisted of 180 trials; participants were allotted 3 min to complete it. The rotation angles of the matching images ranged from 15° to 345°, in intervals of 15°. On each trial, the stimuli remained on the screen until the participant responded by pressing the "P" or the "Q" key.

Raven's Progressive Matrices. The Raven's Progressive Matrices test (Raven, 1998) was used to assess general intelligence. For this task, participants had to identify the missing segment that would complete a figure's inherently regular pattern. On each trial, two segments were presented side by side beneath an incomplete figure; participants were instructed to press "Q" if the figure's missing segment was on the left and "P" if it was on the right. The test consisted of 80 trials; participants were allotted 3 min to complete it.

Data analysis

Because we used a large sample of students from 48 classes, it was necessary to first determine whether the nested data needed to be analyzed with multilevel models. We used the unconditional-means model to compute the intraclass correlation coefficients (*ICC*s; Peugh & Enders, 2005). The *ICC*s were .22 for subtraction and .37 for multiplication, suggesting

significant between-class variability. We therefore conducted multilevel modeling by using the MIXED procedure in SPSS for all analyses. The following equation was used for Level 1:

$$\text{score}_{ij} = \beta_{0j} + \beta_{1j}(\text{age}_{ij}) + \beta_{2j}(\text{gender}_{ij}) + \beta_{3j}(\text{covariates}_{ij}) + \gamma_{ij}.$$

In this equation, score_{ij} is the simple-subtraction or complex-multiplication score for participant i in class j, and β_{0j} is the mean score for class j. β_{1j}, β_{2j}, and β_{3j} are the slopes of age, gender, and covariates (i.e., scores for various tests) predicting the score within class j; γ_{ij} is the random component of the score for participant i in class j.

The following equations were used for Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{residence}_{ij}) + \mu_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$

In these equations, β_{0j} is the mean score for class j, γ_{00} is the grand mean score across all classes, γ_{01} is the slope of the Level 2 residence variable (urban vs. rural) predicting the mean score for class j, and μ_{0j} is the random component of the mean score for class j; γ_{1j}, γ_{2j}, and γ_{3j} are the slopes of age, gender, and covariates predicting the mean score for class j.

The combined equation was as follows:

$$\text{score}_{ij} = \gamma_{00} + \gamma_{01}(\text{residence}) + \gamma_{10}(\text{age}) + \gamma_{20}(\text{gender}) + \gamma_{30}(\text{covariates}) + \mu_{0j} + \gamma_{ij}.$$

Results

Table 1 shows the mean scores and standard deviations for all tasks and the coefficients for gender from the multilevel models. The left panels of Figures 1 and 2 show the mean scores for simple subtraction and complex multiplication, respectively, by gender, age, and area of residence (without controlling for other covariates).

Older children significantly outperformed younger children on all tasks. None of the interactions involving gender and arithmetic performance (i.e., performance on the simple-subtraction and complex-multiplication tasks) were significant. Girls outperformed boys in arithmetic (i.e., simple subtraction, complex multiplication) as well as in numerosity comparison, number comparison, number-series completion, choice reaction time, and word rhyming. Boys outperformed girls in mental rotation.

Urban children's performance was superior to that of rural children on several tasks—simple subtraction: $b = -2.55$, $t(37.17) = -2.40$, $p = .02$; numerosity estimation: $b = -0.89$, $t(41.36) = -2.43$, $p = .02$; number-series completion: $b = -1.52$, $t(37.25) = -3.95$, $p < .0001$; mental rotation: $b = -1.53$, $t(38.09) = -2.84$, $p = .007$; Raven's Progressive Matrices: $b = -2.88$, $t(42.27) = -5.12$, $p < .0001$, and word rhyming: $b = 185.89$, $t(44.14) = 3.02$, $p = .004$, for reaction times and $b = 0.09$, $t(40.92) = 4.91$, $p < .0001$, for error rates. Urban children's reaction times on the choice reaction time task were also superior to those of rural children, $b = 32.09$, $t(41.72) = 2.43$, $p = .02$.
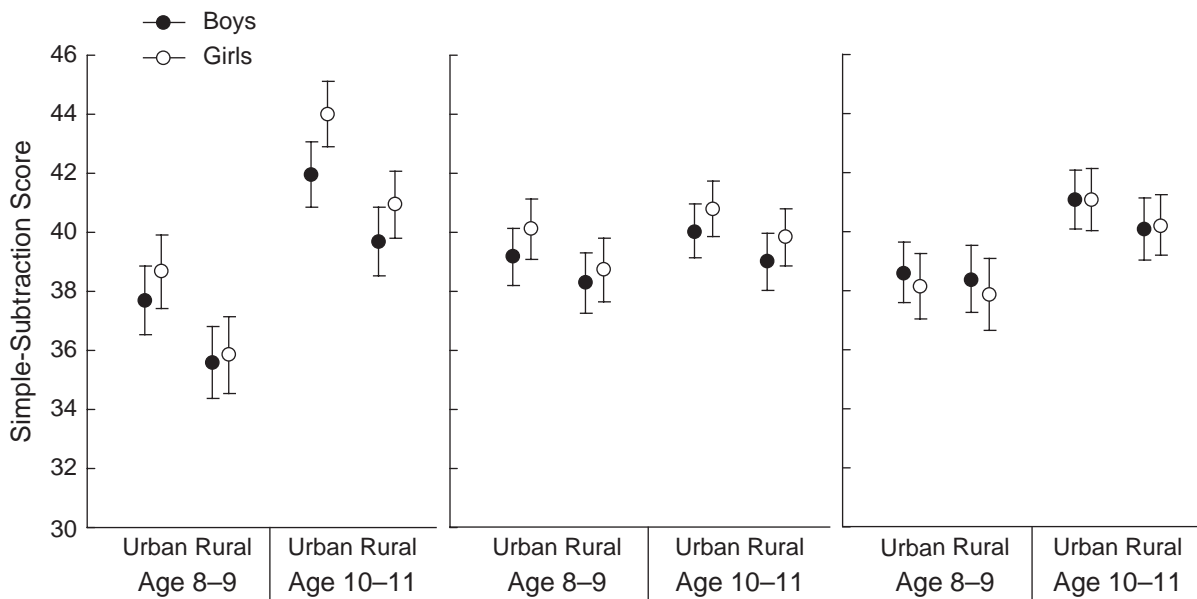

Fig. 1. Average number of correct trials (out of 92 trials) in the simple-subtraction task. The graph on the left shows the raw data; the graph in the middle shows results after controlling for performance on all other tasks except the word-rhyming task; and the graph on the right shows results after controlling for performance on the word-rhyming task only. Error bars indicate 95% confidence intervals.
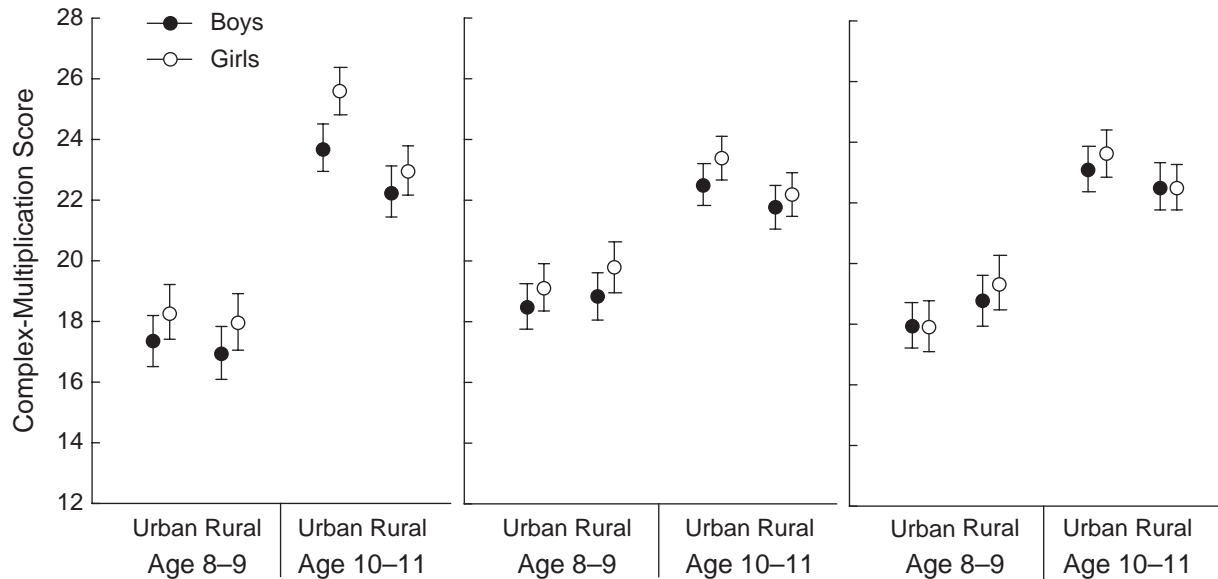
Fig. 2. Average number of correct trials (out of 76) for the complex-multiplication task. The graph on the left shows the raw data; the graph in the middle shows results after controlling for performance on all other tasks except the word-rhyming task; and the graph on the right shows results after controlling for performance on the word-rhyming task only. Error bars indicate 95% confidence intervals.

Table 2 shows the intertask correlations. With few exceptions (e.g., word-rhyming reaction times), correlations were significant and in the expected direction. Most notably, among all tasks other than the arithmetic tasks, error rates for word rhyming had the highest correlations with simple subtraction and complex multiplication. Figures 3 and 4 show the distribution of language ability, the distribution of arithmetic performance, and their relationship by gender. The moderate correlations (approximately −.50) between arithmetic performance (both subtraction and multiplication) and language ability were similar for boys and girls. Forward stepwise regression showed that word rhyming was the most powerful predictor of arithmetic performance (for more details about and results from the stepwise regression analysis, see the Supplemental Material available online).

Multilevel model analysis showed that after controlling for performance (both error rates and reaction times) on the word-rhyming task, there were no gender differences in performance on the simple-subtraction and complex-multiplication tasks (see Table 3 and the right panels of Figs. 1 and 2). Controlling for performance on the number-series-completion task eliminated gender differences for simple subtraction but not for complex multiplication. No other covariates accounted for gender differences in arithmetic performance. The results of this multilevel model analysis are displayed in Table 3. We further examined whether gender differences in word rhyming could explain gender differences in other tasks. Results showed that after word-rhyming performance was controlled for, boys' and girls' performance was similar on the symbolic number-related tasks, including number comparison, $b = -0.003$, $t(1541) = -1.08$, $p = .28$, and number-series completion, $b = 0.11$, $t(1533) = 0.58$,

$p = .57$. However, gender differences remained for tasks that did not involve symbolic numbers, including choice reaction time, $b = -0.01$, $t(1540) = -2.58$, $p = .01$; numerosity comparison, $b = -28.85$, $t(1519) = -3.30$, $p = .001$; and mental rotation, $b = -1.94$, $t(1548) = -4.94$, $p < .0001$.

Finally, to investigate whether performance on all tasks (i.e., both general cognitive tasks and numerical-processing tasks) except for word rhyming could account for gender differences in arithmetic performance, we conducted analyses that included performance on all tasks except for word rhyming as covariates. Results from these analyses of performance on the simple-subtraction task and the complex-multiplication task, respectively, are shown in the middle panels of Figures 1 and 2. Gender differences remained significant for both simple subtraction, $b = 0.80$, $t(1522) = 2.29$, $p = .022$, and complex multiplication, $b = 0.55$, $t(1496) = 2.20$, $p = .028$.

Discussion

Our goal in this study was to examine whether gender differences in children's arithmetic performance could be accounted for by gender differences in language abilities. Our results showed that gender differences in arithmetic were significant and favored girls. Controlling for scores on the word-rhyming task, however, eliminated gender differences in arithmetic performance. In contrast, scores on the basic numerical-processing tasks (number comparison, numerosity estimation, and numerosity comparison) and general cognitive tasks (choice reaction time, Raven's Progressive Matrices, and mental rotation) did not account for gender differences in arithmetic performance. Finally, controlling for performance on the

Table 2. Correlations Among All Performance Variables

Task	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Simple subtraction	—												
2. Complex multiplication	.56**	—											
3. Word rhyming (RT)	.00	-.01	—										
4. Word rhyming (error rate)	-.50**	-.49**	-.01	—									
5. Numerosity comparison (RT)	-.08*	-.16**	.18**	.18**	—								
6. Numerosity comparison (error rate)	-.39**	-.30**	-.15**	.25**	-.24**	—							
7. Numerosity estimation	-.32**	-.29**	.12**	-.32**	-.11**	-.18**	—						
8. Number comparison (RT)	-.22**	-.26**	.22**	.20**	.56**	-.05	-.14**	—					
9. Number comparison (error rate)	-.35**	-.38**	-.25**	.30**	-.19**	.41**	-.18**	-.19**	—				
10. Number-series completion	.47**	.46**	.06*	-.44**	-.07**	-.24**	.28**	-.06*	-.29**	—			
11. Choice reaction (RT)	-.20**	-.18**	.12**	.23**	.45**	-.09**	-.14**	.41**	-.10**	-.15**	—		
12. Choice reaction (error rate)	-.15**	-.12**	-.06*	.10**	-.02	.14**	-.07**	-.02	.16**	-.11**	.33**	—	
13. Mental rotation	.26**	.28**	.08**	-.28**	-.06*	-.18**	.18**	-.07**	-.25**	.25**	-.12**	-.09**	—
14. Raven's Progressive Matrices (Raven, 1998)	.32**	.29**	.05	-.35**	-.11**	-.21**	.24**	-.08**	-.20**	.38**	-.17**	-.08**	.31**

Note: RT = reaction time.
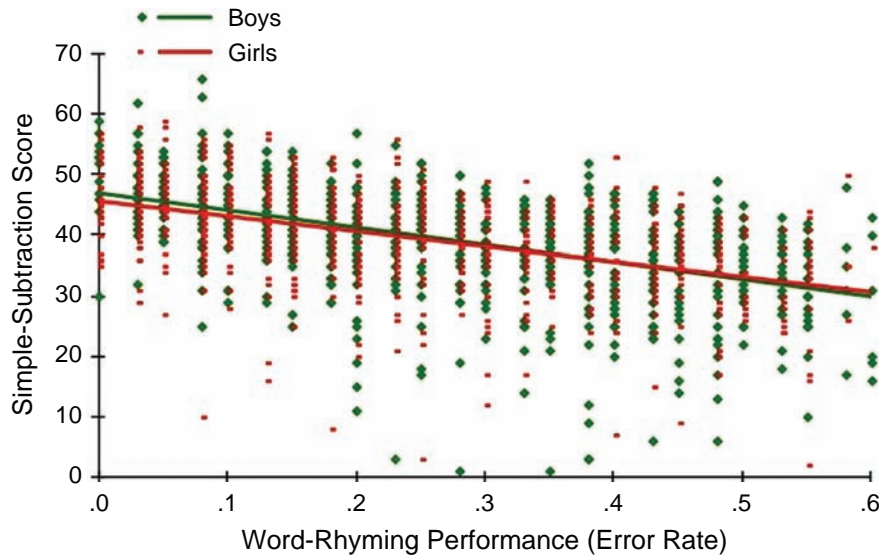*$p < .05$. **$p < .01$.

Fig. 3. Scatter plot (with best-fitting regression lines) showing scores on the simple-subtraction task as a function of the proportion of errors on the word-rhyming task. Results are shown separately for boys ($r = -.51, p < .0001$) and girls ($r = -.50, p < .0001$).

number-series-completion task eliminated the gender gap for performance on the simple-subtraction task, but not the complex-multiplication task. Gender differences were generally consistent across age groups and areas of residence (urban vs. rural; i.e., there were few interactions between gender and age group or gender and area of residence). These results suggest that girls' advantage in arithmetic was likely due to their advantage in language processing, rather than an advantage in basic numerical processing or particular cognitive abilities.

Language processing has previously been shown to be involved in arithmetic processing. In an earlier review of behavioral studies, Aiken (1971) concluded that verbal processing is associated with arithmetic performance. More recent imaging studies have further shown that exact calculation (e.g., addition and subtraction) involves more verbal processing than does approximate calculation (e.g., Dehaene et al., 1999). Similarly, Fedorenko, Gibson, and Rohde (2007) showed that language processing was closely related to
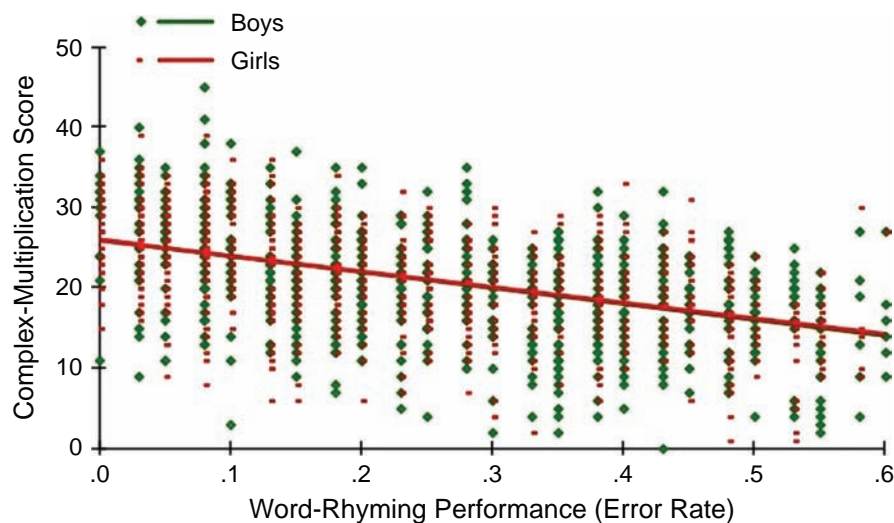

Fig. 4. Scatter plot (with best-fitting regression lines) showing scores on the complex-multiplication task as a function of the proportion of errors on the word-rhyming task. Results are shown separately for boys ($r = -.47, p < .0001$) and girls ($r = -.49, p < .0001$).

Table 3. Results From Multilevel Modeling Showing Gender Differences in Simple-Subtraction and Complex-Multiplication Performance

Covariate	Simple subtraction			Complex multiplication		
	b	*SE b*	*t*	*b*	*SE b*	*t*
None	1.19	0.40	$t(1510) = 2.97$**	0.93	0.28	$t(1502) = 3.30$***
Word-rhyming performance	−0.03	0.38	$t(1514) = -0.08$	0.09	0.26	$t(1500) = 0.35$
Numerosity-comparison performance	1.03	0.38	$t(1511) = 2.71$**	0.84	0.27	$t(1499) = 3.06$**
Numerosity-estimation performance	1.17	0.39	$t(1511) = 3.00$**	0.92	0.28	$t(1501) = 3.33$***
Number-comparison performance	0.85	0.38	$t(1515) = 2.25$*	0.66	0.26	$t(1501) = 2.53$*
Number-series-completion performance	0.69	0.37	$t(1513) = 1.85$[†]	0.57	0.26	$t(1500) = 2.22$*
Choice reaction time	1.07	0.40	$t(1508) = 2.68$**	0.85	0.28	$t(1499) = 3.03$**
Mental rotation performance	1.45	0.40	$t(1512) = 3.67$***	1.12	0.28	$t(1502) = 4.04$***
Raven's Progressive Matrices (Raven, 1998)	1.24	0.39	$t(1511) = 3.17$**	0.96	0.28	$t(1500) = 3.48$***

Note: Gender was coded as 0 for male and 1 for female.
[†]$p = .055$. *$p < .05$. **$p < .01$. ***$p < .001$.

arithmetic processing because the two types of processing share working memory resources. Moreover, in children, dyscalculia is often accompanied by reading difficulties (Jordan, Hanich, & Kaplan, 2003; Landerl, Bevan, & Butterworth, 2004). There is also evidence that verbal skills are more important for arithmetic (especially mental arithmetic) performance than are other cognitive skills, such as spatial skills (e.g., Solan, 1987).

Consistent with this literature is our finding that performance on the language-processing (word-rhyming) task, among all tasks, was most closely linked to arithmetic performance. More important, our results suggest that gender differences in verbal processing are responsible for gender differences in arithmetic performance. This link might be due to differences in the ways in which boys and girls acquire arithmetic facts. Fennema et al. (1998) showed that beginning in the first grade, girls prefer to use concrete strategies, such as counting or modeling, whereas boys tend to use abstract strategies. The use of concrete strategies, in addition to conceptual knowledge, can improve performance in arithmetic (Siegler & Shrager, 1984).

We found that after controlling for scores on the number-series-completion task, gender differences in simple-subtraction performance became marginal, but gender differences in complex-multiplication performance remained significant. One possible explanation for this finding is that the verbal processing required for simple subtraction is similar to that required for number-series completion, but that complex multiplication involves more verbal processing than either simple subtraction or number-series completion. Previous research has shown that simple multiplication (i.e., multiplication using only single-digit factors) involves more verbal processing than does addition or subtraction (e.g., Zhou, Chen, Zang, et al., 2007; Zhou et al., 2006). On the basis of these findings, we inferred that complex multiplication should involve even more verbal processing than simple multiplication because it involves more verbal representation of multiplication facts.

Several limitations of the current study should be noted. First, we included only one verbal-processing task in our study. Although performance on this task accounted for gender differences in arithmetic performance, a broader array of verbal tasks would have allowed for finer analyses of the links between verbal and arithmetic processing. Second, our sample consisted of primary-school children, so our results may not generalize to children in other age groups. It is not clear whether the continued advantage in verbal processing among older female students (e.g., in high school and college) is linked to their performance in mathematics beyond arithmetic. Empirical evidence has shown that girls lag behind boys in high-school mathematics, even though their language skills are superior (Fleischman, Hopstock, Pelczar, & Shelley, 2010). Finally, it should be emphasized that, although the gender differences we found were consistent across age groups and between students from urban and rural areas, their magnitude was modest. Our results should therefore be interpreted cautiously.

In summary, numerous studies have demonstrated girls' advantages in verbal processing (e.g. Bornstein et al., 2000; Burman et al., 2008; Hyde & Linn, 1988; Maccoby & Jacklin, 1974). The current study shows that such advantages lead to

girls' superior performance in arithmetic, because of the important role verbal processing plays in arithmetic performance.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

Aiken, L., Jr. (1971). Verbal factors and mathematics learning: A review of research. *Journal for Research in Mathematics Education*, *2*, 304–313.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology*, *82*, 65–92.

Benbow, C., & Stanley, J. (1983). Sex differences in mathematical reasoning ability: More facts. *Science*, *222*, 1029.

Bornstein, M., Haynes, O., Painter, K., & Genevro, J. (2000). Child language with mother and with stranger at home and in the laboratory: A methodological study. *Journal of Child Language*, *27*, 407–420.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, *301*, 419–421.

Burman, D., Bitan, T., & Booth, J. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*, 1349–1362.

Casey, M., Nuttall, R., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology*, *35*, 1237–1247.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, *56*, 246–269.

Fennema, E., Carpenter, T., Jacobs, V., Franke, M., & Levi, L. (1998). A longitudinal study of gender differences in young children's mathematical thinking. *Educational Researcher*, *27*, 6–11.

Flannery, K., Liederman, J., Daly, L., & Schultz, J. (2000). Male prevalence for reading disability is found in a large sample of Black and White children free from ascertainment bias. *Journal of the International Neuropsychological Society*, *6*, 433–442.

Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context*. Alexandria, VA: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Geary, D. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, *19*, 229–247.

Ginsburg, H., & Baroody, A. (1990). *Examiner's manual of the Test of Early Mathematics Ability* (2nd ed.). Austin, TX: Pro-Ed.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104–122.

Hyde, J., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.

Hyde, J., Lindberg, S., Linn, M., Ellis, A., & Williams, C. (2008). Gender similarities characterize math performance. *Science*, *321*, 494–495.

Hyde, J., & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, *74*, 834–850.

Krueger, L. E. (1982). Single judgments of numerosity. *Attention, Perception, & Psychophysics*, *31*, 175–182.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*, *93*, 99–125.

Linn, M., & Hyde, J. (1989). Gender, mathematics, and science. *Educational Researcher*, *18*, 17–27.

Lu, J., Zhou, X., Zhao, H., Xu, J., Yuan, H., & Dong, Q. (2009, May). *Neural correlates of operation reasoning in mathematical cognition*. Paper presented at the Workshop on Cognitive Science: From Cellular Mechanisms to Computational Theories, Beijing, China.

Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Mann, V., Sasanuma, S., Sakuma, N., & Masaki, S. (1990). Sex differences in cognitive abilities: A cross-cultural perspective. *Neuropsychologia*, *28*, 1063–1077.

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. *Educational and Psychological Measurement*, *65*, 717–741.

Raven, J. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford, England: Oxford Psychologists Press.

Roulstone, S., Loader, S., Northstone, K., & Beveridge, M. (2002). The speech and language of children aged 25 months: Descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Child Development and Care*, *172*, 259–268.

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Smith, P., Fernandes, C., & Strand, S. (2001). *Cognitive Abilities Test 3: Technical manual and pupil book*. Windsor, England: NFER-Nelson.

Solan, H. (1987). The effects of visual-spatial and verbal skills on written and mental arithmetic. *Journal of the American Optometric Association*, *58*, 88–94.

Tan, L., Liu, H., Perfetti, C., Spinks, J., Fox, P., & Gao, J. (2001). The neural system underlying Chinese logograph reading. *NeuroImage*, *13*, 836–846.

Willingham, W., & Cole, N. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Zhou, X., Chen, C., Dong, Q., Zhang, H., Zhou, R., Zhao, H., . . . Guo, Y. (2006). Event-related potentials of single-digit addition, subtraction, and multiplication. *Neuropsychologia*, *44*, 2500–2507.

Zhou, X., Chen, C., Zang, Y., Dong, Q., Qiao, S., & Gong, Q. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, *35*, 871–880.

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., & Dong, Q. (2007). Chinese kindergartners' automatic processing of numerical magnitude in Stroop-like tasks. *Memory & Cognition*, *35*, 464–470.

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: A review of literature. *International Education Journal*, *8*, 187–203.