Journal of Experimental Child Psychology **77,** 337–353 (2000)
doi:10.1006/jecp.2000.2594, available online at http://www.idealibrary.com on IDEAL®

Sex Differences in Spatial Cognition, Computational Fluency, and Arithmetical Reasoning

David C. Geary, Scott J. Saults, Fan Liu, and Mary K. Hoard

University of Missouri at Columbia

Alternative explanations for the male advantage in arithmetical reasoning, as measured by the ability to solve complex word problems, include a male advantage in spatial cognition and a male advantage in computational fluency. The current study was designed to test these competing hypotheses. To this end, 113 male and 123 female undergraduates were administered arithmetical computations and arithmetical reasoning tests, along with an IQ test and a test of spatial cognition. There was no sex difference on the IQ test, but males showed significantly higher mean scores on the arithmetical computations, arithmetical reasoning, and spatial cognition measures. A series of structural equation models indicated that individual differences in arithmetical reasoning were related to individual differences in IQ, spatial abilities, and computational fluency. Moreover, the results suggested that the male advantage in arithmetical reasoning is mediated by the male advantages in both computational fluency and spatial cognition.

Key Words: sex differences; arithmetical reasoning; mathematical word problems; spatial cognition; computational fluency.

Developmental research has focused on sex differences in childhood and in adulthood for many decades (e.g., Brown, 1958; Maccoby, 1990; Maccoby & Jacklin, 1974). Among the areas of investigation are studies of the nature and extent of sex differences in mathematical competencies (Benbow, 1988; Geary, 1996; Hyde, Fennema, & Lamon, 1990). Within the broad domain of mathematics, sex differences, favoring males, are often found for the speed and accuracy with which word problems can be solved (Benbow, 1988; Casey, Nuttall, & Pezaris, 1997; Casey, Nuttall, Pezaris, & Benbow, 1995; Geary, 1996; Johnson, 1984; Marshall & Smith, 1987). The male advantage in the solving of word problems is found as early as first grade and for children, adolescents, and young

We thank Nicole Harris, Aaron Stratman, and Natalija Popovic for their assistance with various aspects of the study and Jeff Bisanz, James Royer, and an anonymous reviewer for comments on an earlier draft. The research was supported by a grant from the Research Board of the University of Missouri and a Summer Research Fellowship award to the first author by the Research Council of the University of Missouri at Columbia.

Address correspondence and reprint requests to David C. Geary, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211-2500. Fax: (573) 882-7710. E-mail: GearyD@Missouri.edu.

0022-0965/00 $35.00

adults in the United States, China, Japan, and a host of European nations (Harnisch, Steinkamp, Tsai, & Walberg, 1986; Lummis & Stevenson, 1990; Stevenson et al., 1990). The magnitude of the male advantage in this area is small to moderate, ranging between 0.2 and about 0.6 standard deviations, depending on the age of the sample and the complexity of the test. Mean differences of this magnitude indicate that roughly 60 to 70% of males score higher than the average female, although this gap widens considerably at the upper end of the ability distribution (Benbow, 1988).

The source of the male advantage in solving mathematical word problems is currently debated. Alternative explanations include sex differences, favoring males, in spatial cognition, computational fluency, and attitudes toward mathematics (Casey et al., 1997; Geary, 1996; Johnson, 1984; Royer, Tronsky, Chan, Jackson, & Marchant, 1999). Casey and her colleagues found that the male advantage on the mathematics section of the Scholastic Achievement Test (SAT-M) was mediated by a male advantage in three-dimensional spatial abilities and by more positive attitudes toward mathematics. In a series of experiments, Johnson found that male college students consistently outperformed their female peers on tests that involved the solving of algebraic word problems, but not on tests of nonmathematical problem solving. Johnson also found that individual differences on the tests of algebraic problem solving were correlated with spatial abilities, IQ, and attitudes toward mathematics (e.g., rated importance of mathematics). In Johnson's studies and those of Casey and her colleagues, the sex difference in mathematical problem solving was more strongly related to spatial abilities than to attitudes toward mathematics (see also Benbow, 1988). In an analysis of the academic achievement of children in Japan, Taiwan, and the United States, Lummis and Stevenson (1990) found no overall sex differences in attitudes toward mathematics or on a general test of mathematics achievement. However, in both first and fifth grade and in all three nations, boys outscored girls on three mathematics subtests, word problems, visualization, and estimation.

These patterns suggest that the male advantage for solving mathematical word problems might be mediated, in part, by their advantage in spatial cognition. More specifically, it appears that males are better at generating spatial representations, or diagrams, of the relational information conveyed in word problems (Geary, 1996; Johnson, 1984), which has been shown to reduce the frequency of problem-solving errors (Lewis, 1989). However, this is by no means a consensus view. In a meta-analysis, Friedman (1995) found that performance on mathematics tests tended to be more strongly correlated with verbal abilities than with spatial abilities and concluded that these patterns "are not convincing evidence that spatial skill is well related to mathematical ability" (p. 40; see also Lubinski & Humphreys, 1990). Linn and Petersen (1985) suggested that the relation between spatial and mathematical abilities is actually due to their covariation with intelligence (IQ) and not a causal spatial–mathematics relation. More recently, Royer and his colleagues argued that the male advantage in mathematics is not related to spatial cognition at all, but rather to a male advantage in computational arithmetic, specifically, in the speed of retrieving arithmetic facts from long-term memory (Royer et al., 1999).

Across a series of nine experiments, Royer and his colleagues showed that speed of solving simple (e.g., $4 + 6$) and complex (e.g., $4 \times 7 \times 3$) arithmetic problems was correlated with performance on paper-and-pencil arithmetical computational and arithmetical reasoning (i.e., word problem) tests (Royer et al., 1999; see also Geary & Widaman, 1987, 1992). In fourth through eighth grades there were no sex differences in speed of computational problem solving, but as a group boys were more variable in their performance. At these grade levels, the fastest boys were faster than the fastest girls at basic computations, whereas the slowest boys were slower than the slowest girls. For college students, males were faster, on average, at solving arithmetic problems than were their same-age female peers, and this sex difference appeared to contribute to a male advantage in mathematical problem solving, that is, on tests that required the solving of word problems.

Other studies of elementary-school children suggest no sex difference in overall arithmetical performance, but sex differences in problem-solving approaches are often found (Carr & Jessup, 1997; Carr, Jessup, & Fuller, 1999; Fennema, Carpenter, Jacobs, Franke, & Levi, 1998; but see Siegler, 1988). In these studies, boys used direct retrieval and covert strategies (e.g., mental counting) more often than girls did and girls used finger counting and overt strategies (e.g., blocks) more often than boys did. To assess these differences further, Chinese and American kindergarten through third-grade children's use of finger counting and direct retrieval to solve simple addition problems (e.g., $4 + 3$) was analyzed using data from several cross-national studies (Geary, Bow-Thomas, Liu, & Siegler, 1996; Geary, Fan, & Bow-Thomas, 1992; see also Geary, 1994). Although the differences were not always statistically significant, in all five samples American girls used finger counting more frequently than did American boys. In third grade, U.S. boys correctly retrieved significantly more addition facts than did U.S. girls ($p < .05$). In kindergarten, Chinese boys used finger counting more frequently than Chinese girls did ($p < .10$), and in first grade boys correctly retrieved more addition facts than did girls ($p < .05$), but there were no other sex differences in the Chinese samples.

To determine if similar sex differences might exist in U.S. adults, data on paper-and-pencil test performance, strategy choices, and speed of processing for solving simple and complex subtraction problems were reanalyzed (Geary, Frensch, & Wiley, 1993); the sample included 36 college students (26 female and 10 male) and 36 older (mean age = 72 years; 22 female and 14 male) adults. There were no sex differences in either generation for performance on the paper-and-pencil tests. For the college students, males and females used direct retrieval to solve 86 and 66% of the simple subtraction problems, respectively ($p < .07$); reconstructive strategies, as counting, were used to solve the remaining problems. There was no sex difference in the speed of retrieving subtraction facts from long-term memory, but during the solving of complex problems (e.g., as in $43 - 8$) males executed the borrowing procedure more than twice as quickly (1028 vs 2175 ms, $p < .09$) as females did. However, there were no sex differences in strategy choices or in the speed of processing for the older generation.

The pattern across studies suggests that sex differences might exist in strategic approaches to solving arithmetic problems and in the speed of executing some component processes, although these patterns are most evident for young adults, adolescents, and children in the United States and might not be found in other nations (e.g., China) or for older Americans. Either way, the findings for younger Americans suggests that Royer and his colleagues' results might not simply be due to a male advantage in speed of arithmetic fact retrieval. Rather, their findings may be due to a combination of sex differences in the pattern of strategy choices and in the speed of executing component processes other than retrieval, such as borrowing (Geary, 1999). Whatever the underlying cause, the Royer et al. finding of a male advantage in computational fluency is provocative, as is their suggestion that this sex difference mediates the male advantage on mathematics achievement tests. In fact, the Royer et al. computational fluency hypothesis represents a plausible alternative to the proposal that the male advantage in the ability to solve mathematical word problems is mediated by the sex difference in spatial cognition.

The current study is the first to simultaneously assess the sex difference in the ability to solve arithmetical word problems as potentially related to sex differences in spatial cognition and computational fluency and thus the first to test these competing hypotheses. At the same time, an IQ test was included in the study so that the relation between the sex differences in spatial cognition, computational fluency, and arithmetical reasoning could be assessed after statistically controlling for the covariation between performance in these domains and IQ. In this way, the hypotheses that the male advantage in spatial cognition or computational fluency contributes to their advantage in solving mathematical word problems could be assessed, while statistically controlling for a potential confounding variable, IQ (Linn & Petersen, 1985).

Because Royer et al. (1999) found that the male advantage in computational fluency was most consistently found in young adults, the current sample also consisted of young adults. If the Royer et al. findings are replicated, then follow-up studies with children and adolescents would be warranted, although, in any such studies, the assessment of sex differences in computational fluency should focus on strategy choice and speed of processing differences. This is because sex differences on paper-and-pencil computational tests are not always evident for adolescents and children (Carr & Jessup, 1997; Carr et al., 1999;

Fennema et al., 1998). In any case, before such studies are conducted, an attempt to replicate Royer et al.'s most robust finding seems wise.

METHOD

Participants

The participants were 236 (113 male and 123 female) general psychology students from the University of Missouri at Columbia. The mean age of the males and females was 19.0 ($SD = 2.0$) and 18.5 ($SD = 2.4$) years, respectively, and did not differ significantly, $F(1, 229) = 2.66$, $p > .10$.

Ability Measures

Arithmetical computations. Three tests, each including two forms, were used to assess computational abilities. One of these, the Addition Test, was from the Educational Testing Service (ETS) kit of factor-referenced tests (Ekstrom, French, & Harman, 1976) and other two, Simple Subtraction and Complex Subtraction, were used based on previous research (Geary et al., 1993). Each form of the Addition Test, hereafter Complex Addition (to emphasize that the items were complex, not simple, e.g., 4 + 3, problems), allows 2 min to solve as many complex addition problems as possible (e.g., 19 + 8 + 27). For each form of the Simple Subtraction (e.g., 9 − 5) and Complex Subtraction (e.g., 78 − 9) tests, participants were allowed 1 min to solve as many problems as possible. For all tests, the score was the number of items solved correctly across both forms. The reliability of each measure was estimated by means of Cronbach's alpha and were uniformly high; Simple Subtraction (.95), Complex Subtraction (.89), and Complex Addition (.91).

Arithmetical reasoning. Both arithmetical reasoning tests—Necessary Arithmetic Operations (NAO) and Arithmetic Aptitude (AA)—were from the ETS kit of factor-referenced tests (Ekstrom et al., 1976). Again, both forms of each test were administered, with 5 min allowed for each form of NAO and 10 min for each form of AA. Both tests include multistep arithmetic word problems, but NAO only requires participants to indicate the order in which arithmetic operations (e.g., multiplication then division) need to be used to solve the problem—no computations are needed. Items for AA require participants to solve multistep word problems. The score was the number of items solved correctly minus a fraction of the number of items solved incorrectly (to control for guessing) across both forms. The reliability estimates, again based on Cronbach's alpha, were .69 for NAO and .77 for AA.

Mental rotation test. The MRT requires the mental rotation of geometric figures in three-dimensional space and includes two forms (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978). For each of 10 items on each form, four test figures are compared to a standard figure. Across forms, 40 test figures (2/item) are rotations of the standard figure and 40 are rotated mirror images or rotated images that differ on one or several features from the standard figure. The task

TABLE 1
Mean Test Scores and Effect Sizes

Test	Males *M*	Males *SD*	Females *M*	Females *SD*	Effect size (*d*)
Computational tests					
Simple subtraction	121.8	22.1	114.9	22.6	.31
Complex subtraction	41.6	13.9	33.2	11.4	.66
Complex addition	37.0	10.6	36.6	7.3	.04
Reasoning tests					
Necessary Arthrimetic Operations	16.0	5.4	14.9	5.6	.20
Arithmetic Aptitude	14.9	6.6	12.4	6.1	.39
Mental Rotation Test	18.5	9.5	12.5	7.7	.70
IQ	114.6	11.6	113.4	12.0	.10

Note. The effect size is given by $d = M$ (males) $-$ M (females)/.5[SD (males) + SD (females)], where M (males) and M (females) are the mean scores of the males and females, respectively, and SD (males) and SD (females) are the respective standard deviations.

is to identify the test figures that are rotations of the standard figure. Participants were allowed 3 min to compete each form. The score was the number of test figures correctly identified minus the number of test figures incorrectly identified (to correct for guessing); reliability estimate, .70.

Intelligence. The Raven's Progressive Matrices served as the measure of general intelligence (Raven, Court, & Raven, 1993). Due to time constraints, the test was started at the item representing an IQ of 80 and participants were given credit for all items up to this level. Participants were then allowed 30 min to complete the remaining items and nearly all did so in the allotted time.

Procedure

All participants were tested in small groups, and all tests were administered in accordance with standardized instructions and in the following order: Simple Subtraction, Complex Subtraction, Complex Addition, NAO, AA, MRT, and the IQ test. Testing sessions lasted about 105 min.

RESULTS

The results are presented in two sections. The first describes mean sex differences on each of the measures, whereas the second describes structural equation models that assessed potential mediators of the male advantage on the arithmetical reasoning measures.

Mean Sex Differences

As shown in Table 1, males had higher means scores than did females on all of the measures, but these differences were significant for only four tests; Simple

TABLE 2
Correlations between Observed Variables

	1	2	3	4	5	6	7	8	9
1. Simple subtraction	—								
2. Complex subtraction	.63	—							
3. Complex addition	.63	.58	—						
4. NAO	.31	.44	.33	—					
5. AA	.45	.57	.47	.62	—				
6. MRT—Form 1	.11	.26	.07	.30	.34	—			
7. MRT—Form 2	.11	.16	.07	.22	.20	.54	—		
8. IQ	.21	.25	.14	.34	.35	.29	.21	—	
9. Sex	.15	.32	.02	.10	.20	.34	.21	.05	—

Note. The structural equation models were estimated using a covariance, not correlational, matrix; $r \geq .14$, $p < .05$; $r \geq .20$, $p < .01$; $r \geq .25$, $p < .0001$.

Subtraction, $F(1, 233) = 5.64$, $p < .05$; Complex Subtraction, $F(1, 233) = 25.97$, $p < .0001$; AA, $F(1, 234) = 9.49$, $p < .005$; and MRT, $F(1, 234) = 28.14$, $p < .0001$. The fifth column in Table 1 shows the mean differences in terms of effect size (d); that is, the difference between the male and female means divided by the mean of the two respective standard deviations (Cohen, 1988). The effect sizes ranged from .04 for Complex Addition to .70 for the MRT.

Structural Models

Structural equation models were used to assess the relations among the variables listed in Table 1 and to test substantive hypotheses. All models were based on covariance matrixes and were estimated by means of LISREL 8 (Jöreskog & Sörbom, 1996). The first section describes structural models that assessed the pattern of sex differences, whereas the second describes within-sex relations for the final structural model described in the first section.

Sex differences analyses. First, a null model hypothesizing no common factors or latent variables was estimated for the covariance matrix; the corresponding correlation matrix is shown in Table 2. The goodness-of-fit of the null and subsequent models was evaluated by the goodness-of-fit index (GFI) and the normed-fit index (NFI, not available for the null model; Bentler & Bonett, 1980; Jöreskog & Sörbom, 1996). Values greater than .90 are typically considered acceptable; that is, they indicate that the model provides an acceptable representation of the covariation among the observed measures. The very low GFI value (.48) shown in Table 3 indicates that the null model was clearly rejectable.

Model 1, the measurement model, involved the estimation of five latent variables and their associated covariances, Arithmetical Computations, Arithmetical Reasoning, Spatial Cognition, IQ, and Sex. These relations are represented in Fig. 1. The first latent variable—Arithmetical Computations—was

TABLE 3
Goodness-of-Fit Indexes for Nested Structural Equation Models

Model	*df*	χ^2	GFI	NFI
Overall fit of nested models				
Null	36	869.43	.48	—
Model 1: Measurement model	21	82.71	.93	.90
Model 2: Fix factor covariances to 0 and paths from Sex to AC, AR, and SC	28	249.94	.81	.71
Model 3: Paths from IQ to AC, AR, and SC	25	151.77	.89	.83
Model 4: Path from SC to AR	24	148.46	.89	.83
Model 5: Drop path from Sex to AR	25	149.31	.89	.83
Model 6: Path from AC to AR	24	84.06	.93	.90

Note. GFI = Goodness of Fit Index; NFI = Normed Fit Index; AC = Arthrimetical Computations; AR = Arthmetical Reasoning; SC = Spatial Cognition.

defined by the three computational tests, while the Arithmetical Reasoning latent variable was defined by the NAO and AA tests. The Spatial Cognition latent variable was defined by the two forms of the MRT test, and the IQ and Sex latent variables were defined by the IQ and sex (coded 1 for males and 0 for females) variables, respectively. To improve the empirical identification of the model, the two factor loadings defining the Arithmetical Reasoning and the Spatial Cognition latent variables were constrained to equality. Without such an equality constraint, the single covariance between the two variables defining each of these constructs would be used to estimate two factor loadings. Although such factor loadings can be mathematically identified, the values of unconstrained loadings are often unstable (Jöreskog & Sörbom, 1996). Finally, the unique variances for the IQ and sex variables were fixed at .01 and all latent variable variances were fixed at 1.0, to mathematically identify the model. The estimation of Model 1 resulted in a significant improvement in model fit, relative to the null model, $\Delta \chi^2(15) = 786.72$, $p < .001$, and yielded acceptable GFI (.93) and NFI (.90) values, as shown in Table 3.

In order to test hypothesized relations among the latent variables, Model 2 involved fixing the 10 covariances among the latent variables at 0 and then estimating directed paths from Sex to the Arithmetical Computations, Arithmetical Reasoning, and Spatial Cognition latent variables, as shown in Fig. 1. A path from Sex to IQ was not estimated because there was no sex difference on the IQ test. The estimation of these three directed paths enabled a test of whether there were significant sex differences for these latent constructs; for instance, a sex difference on the variance common to the NAO and AA tests. In keeping with the mean sex differences described above, the standardized coefficients for all of the directed paths were significant (*p*s $< .05$) and positive in value (indicating a male

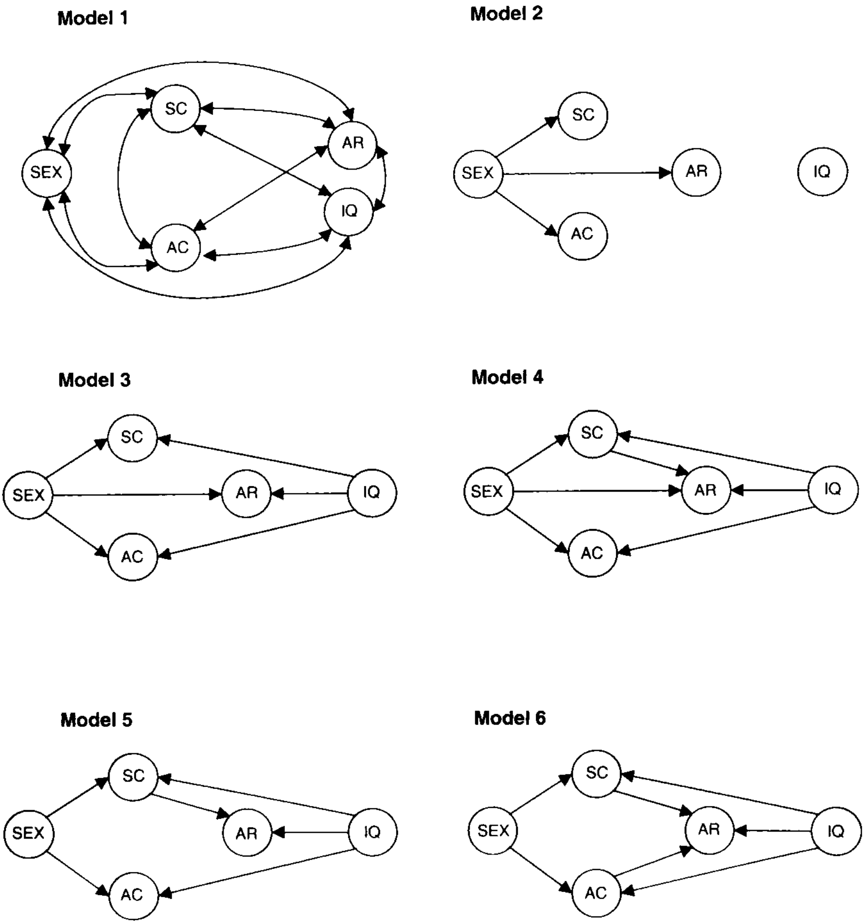

FIG. 1. Patterns of structural relations estimated in Model 1 through Model 6; SC = Spatial Cognition; AC = Arithmetical Computations; AR = Arithmetical Reasoning. See text and Table 3 for corresponding descriptions.

advantage); βs = .49, .18, and .23 for the respective directed paths from Sex to Spatial Cognition, Arithmetical Reasoning, and Arithmetical Computations. However, the specification of Model 2 resulted in a worsening of model fit, $\Delta\ \chi^2(7) = 167.23$, $p < .001$, and unacceptable GFI (.81) and NFI (.71) values. The worsening of model fit indicates significant and unestimated covariation among the latent variables.

To improve the level of model fit and to test substantive hypotheses, a series of additional directed paths were estimated. For Model 3, directed paths from IQ to the Arithmetical Computations, Arithmetical Reasoning, and Spatial Cognition latent variables were estimated, as shown in Fig. 1. The estimation of these paths statistically removed the covariance between IQ and these ability domains from the estimation of the theoretically more substantive relations tested in subsequent models. The resulting model yielded a significant improvement in model fit, $\Delta\ \chi^2(3) = 98.17$, $p < .001$, and improved GFI (.89) and NFI (.83) values. Each of the standardized path coefficients was positive in value—indicating higher IQ scores were associated with better performance in each of the three ability domains (Jensen, 1998)—and differed significantly from 0 (ps $< .05$).

As shown in Fig. 1, model 4 involved the estimation of a directed path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable, based on the hypothesis that spatial abilities contribute to individual differences in the ability to solve mathematical word problems (Geary, 1996; Johnson, 1984; Lewis, 1989). The estimation of Model 4 resulted in a marginally significant improvement in model fit, $\Delta\ \chi^2(1) = 3.31$, $p < .08$. The directed path from Sex to the Arithmetical Reasoning latent variable dropped to nonsignificance ($\beta = .08$, $t < 1$, $p > .25$), whereas the path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable was marginally significant ($\beta = .16$, $t = 1.85$, $p < .08$); GFI $= .89$, NFI $= .83$. A follow-up model, Model 4a, yielded a nonsignificant directed path from the Spatial Cognition latent variable to the Arithmetical Computations latent variable ($\beta = -.01$, $t < 1$, $p > .25$), indicating that the relation between spatial abilities and arithmetical abilities is specific to word problems.

Based on the nonsignificant directed path from Sex to the Arithmetical Reasoning latent variable in Model 4, this path was dropped and the model reestimated, resulting in Model 5 (see Fig. 1). Model 5 yielded a nonsignificant change in model fit, relative to Model 4, $\Delta\ \chi^2(1) < 1$, $p > .50$, and the directed path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable now reached conventional significance levels ($\beta = .21$, $t = 2.8$, $p < .05$). Model 6 is also shown in Fig. 1 and involved the estimation of a path from the Arithmetical Computations latent variable to the Arithmetical Reasoning latent variable. This path was estimated based on the finding that speed of executing basic arithmetical operations (e.g., fact retrieval) contributes to performance on arithmetical reasoning tests (Geary & Widaman, 1992) and to test the Royer et al. (1999) hypothesis. Model 6 resulted in a significant improvement in model fit, relative to Model 5, $\Delta\ \chi^2(1) = 65.3$, $p < .001$, and acceptable GFI (.93) and NFI (.90) values. Equally important, the level of statistical fit comparing Model 6 to Model 1 was nonsignificant, $\Delta\ \chi^2(3) = 1.35$, $p > .50$, indicating that the estimated directed paths adequately captured all of the covariances among the latent variables.

The coefficients for all of these directed paths differed significantly from 0 (ps $> .05$) and are shown in Fig. 2; the associated standardized factor loadings and unique variances are shown in Table 4. The paths show a positive relation between IQ scores and performance in each of the three ability domains and that

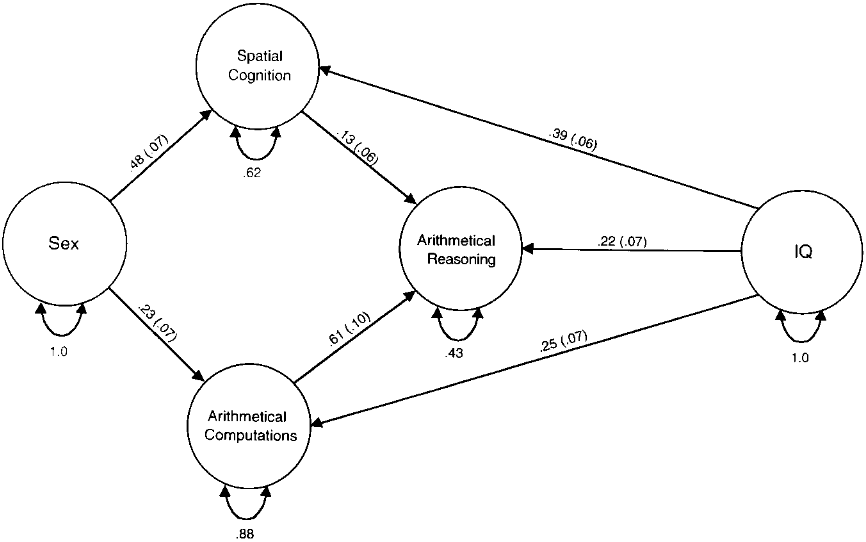

FIG. 2. Standardized estimates from Model 6 of the relations among the latent variables; the parenthetical values are standard errors.

individual differences in arithmetical reasoning are also related to individual differences in computational fluency and spatial abilities. The sex difference in spatial abilities and computational fluency remained significant and favored males. The paths also indicate the male advantage in arithmetical reasoning is mediated by the male advantage in spatial and computational abilities. The product (e.g., .23 × .61) of the indirect paths from Sex to the Arithmetical Reasoning latent variable provides estimates of the relative strength of these two mediators and indicates computational fluency is the stronger of the two mediators.

Finally, to test if sex might now show a direct relation to performance on the arithmetical reasoning tests, Model 6a was estimated. In this model, all of the paths shown in Fig. 2 were retained and the directed path from Sex to the Arithmetical Reasoning latent variable was reestimated. The results showed a nonsignificant, $\Delta \chi^2(1) = 0.59$, $p > .50$, change in model fit. Moreover, the path coefficient from sex to arithmetical reasoning was nonsignificant ($\beta = -.06$, $t < 1$, $p > .10$), and all of the directed paths shown in Fig. 2 remained significant (ps $< .05$) and did not differ significantly in value comparing Model 6 and Model 6a (ps $> .10$). The results provide additional support for the position that the estimation of a direct relation between sex and performance on the arithmetical reasoning tests is not necessary once the indirect relations shown in Fig. 2 are estimated.

Within-sex analyses. To determine if the same pattern of structural relations was evident within sex, Model 6 was separately fitted to covariance matrixes

TABLE 4
Standardized Estimates from Structural Equation Model 6

	Latent variables					
Observed measures	Arithmetical computations	Arithmetical reasoning	Spatial cognition	IQ	Sex	Unique factor
Simple subtraction	.76					.42
Complex subtraction	.81					.34
Complex addition	.71					.50
NAO		.65				.58
AA		.62				.62
MRT Form 1			.62			.62
MRT Form 2			.99			.02
IQ				.99		.02
Sex					.98	.04

Note. All reported factor loading greater than .04 are significantly higher than 0 (ps $<$ .001) and all nonreported loadings were fixed at 0. The unique variances for IQ and Sex were fixed at .01 in the covariance metric; the reported values are the associated standardized estimates. The two loadings defining the Arithmetical Reasoning and Spatial Cognition latent variables were forced to equality in the covariance metric, but different unique variances resulted in different standardized loadings.

derived from the male and female samples. The associated models yielded GFI values of .90 and .93 for the male and female samples, respectively, and respective NFI values of .86 and .92. The standardized path coefficients are shown in Fig. 3 (males/females) and all of these differed significantly (ps $<$ .05) or marginally significantly (ps $<$ .10) from zero.[1] With one exception, the standardized path coefficients did not differ in value across the male and female samples (ps $>$ .10). The one exception was a much stronger relation between IQ and Spatial Cognition in the female sample ($\beta = .58$) than in the male sample ($\beta = .27$; $t = 3.1$, $p < .05$).

DISCUSSION

The study focused on the source of the male advantage in the solving of arithmetical word problems, specifically, sex differences in spatial abilities and computational fluency (Casey et al., 1995, 1997; Geary, 1996; Royer et al., 1999). A corollary hypothesis, of course, is that spatial abilities contribute to the ability to solve mathematical word problems, an assumption that has been repeatedly questioned (Friedman, 1995; Linn & Petersen, 1985; Lubinski &

[1] All path coefficients were significantly different from 0 (ps $<$ .05), except for the following, which were marginally significant (ps $<$ .10); the path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable in both samples and the path from IQ to the Arithmetical Reasoning latent variable in the female sample. The probability of finding two marginally significant path coefficients from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable, if such a relation did not exist, is less than .01.

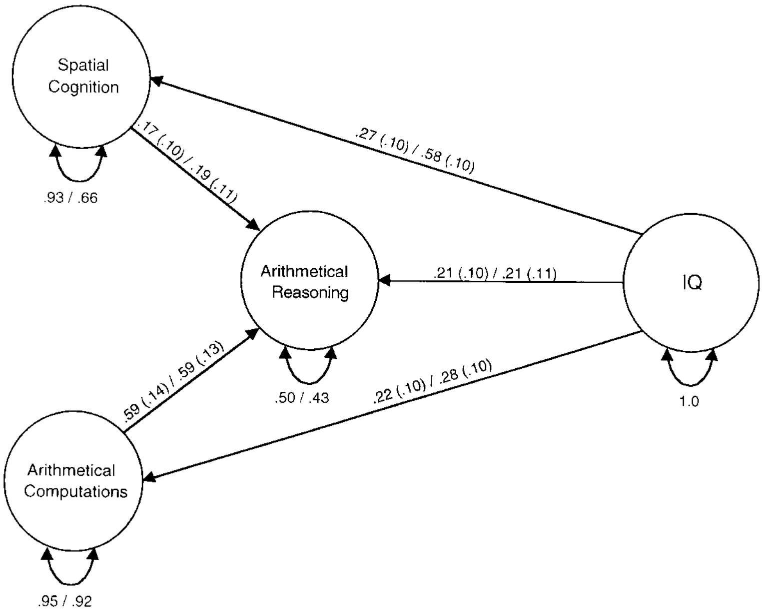

FIG. 3. Standardized estimates from Model 6 of the relations among the latent variables estimated separately for males and females (males/females); the parenthetical values are standard errors.

Humphreys, 1990; Royer et al., 1999). It has been suggested, for instance, that the correlation between spatial and mathematical abilities does not reflect a causal relationship but rather the common influence of intelligence on both spatial and mathematical abilities. By simultaneously estimating paths from IQ to the Spatial Cognition and Arithmetical Reasoning latent variables, the influence of IQ on spatial and arithmetical reasoning abilities was statistically controlled. With IQ controlled, a significant directed path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable was found, suggesting that the correlation between spatial and mathematical abilities is not simply due to the common influence of IQ (see Geary, 1996). However, the results leave the question of whether verbal abilities contribute to the solving of mathematical word problems unanswered, a question that will need to be pursued in future studies (Friedman, 1995).

The magnitude of the relation between spatial abilities and skill at solving arithmetical word problems was somewhat smaller than that found in studies of the relation between spatial abilities and performance on the SAT-M and on algebraic word problems (Casey et al., 1997; Johnson, 1984). In the current study, the path coefficients from spatial cognition to arithmetical reasoning ranged from .13 to .19. The coefficient, based on a path analysis, between spatial abilities and SAT-M performance was .23 in the Casey et al. (1995) study and the partial correlation (controlling for SAT-M performance) between spatial abilities and algebraic problem solving was .29 in Johnson's (1984) study. Differences across studies are likely to reflect differences in the types of items used to assess mathematical problem solving, as it appears that the relation between spatial abilities and mathematical problem solving abilities varies from one word problem to another (Johnson, 1984). Presumably word problems differ in the extent to which spatial representations will facilitate performance (i.e., reduce problem solving errors) and in the ease with which the quantitative relations conveyed in the problems can be translated into spatial representations (Geary, 1996; Lewis, 1989).

The use of structural equation modeling to assess sex differences has an important advantage over the assessment of mean differences on single tests. By defining latent constructs with multiple measures, the influence of error variance associated with individual measures is reduced (Jöreskog & Sörbom, 1996). For instance, the path from Sex to the Arithmetical Reasoning latent variable provides an estimate of the magnitude of the sex difference in variance common to both arithmetical reasoning tests. The associate path coefficient, in turn, should provide a more reliable estimate of the sex difference in arithmetical reasoning abilities than will performance on any single test. The use of this method revealed sex differences, favoring males, for performance on the MRT and on the tests of arithmetical reasoning and arithmetical computations, but not for the IQ test. The latter finding is in keeping with Jensen's (1998) and Geary's (1998) conclusions, based on reviews of the literature, that there are no mean sex differences in general intelligence (but see Rushton & Ankney, 1996). The magnitude of the male advantage on the MRT was in the range found in previous studies (Geary, 1998; Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995), as was the male advantage on the arithmetical reasoning tests (Geary, 1996; Johnson, 1984; Lummis & Stevenson, 1990).

The male advantage on the computational tests is consistent with the results of Royer et al. (1999) as well as with other recent studies. In these studies, sex differences, favoring males, have been found for children's and young adults' strategic approaches to solving arithmetic problems and in the speed of executing some of the underlying component processes (Carr & Jessup, 1997; Carr et al., 1999; Fennema et al., 1998; Geary et al., 1993). The results are, however, inconsistent with Hyde et al.'s (1990) meta-analytic finding of no sex difference in computational abilities for 15- to 18-year-olds and a female advantage through early adolescence. The current results are also inconsistent with the earlier described finding of no consistent sex differences in strategic approaches for solving arithmetic problems or in the speed of processing for older Americans or Chinese children. The reason for the discrepancy between these patterns, the current results, and those of Royer et al. and others are unclear, but might involve sampling issues or cohort effects. At all ages, the cognitive performance of males, including performance on computational tests, is often more variable than that of females, with more males at the higher and lower end of the ability distributions (Geary, 1998; Willingham & Cole, 1997). One result would be a sex difference, favoring males, for high ability samples, and this possibility cannot be ruled out for the current sample. Another possibility is that the male advantage in computational abilities has emerged recently in the United States and thus would not have been found in the studies reviewed by Hyde et al. or in samples of older American adults or Chinese children.

In any case, the structural relations shown in Fig. 2 are consistent with the position that the male advantage in spatial cognition contributes to the male advantage in mathematical problem solving (Casey et al., 1997; Geary, 1996; Johnson, 1984). Once the path from the Spatial Cognition latent variable to the Arithmetical Reasoning latent variable was estimated, the path from Sex to the Arithmetical Reasoning latent variable was no longer significant. In other words, an indirect relation between sex and arithmetical reasoning, mediated by spatial abilities, provided a better fit to the data than did modeling a direct relation between sex and arithmetical reasoning. At the same time, the results are also consistent with those of Royer et al. (1999); that is, that a male advantage in computational fluency contributes to the male advantage in mathematical achievement. In fact, the structural relations shown in Fig. 2 suggest that the male advantage in arithmetical reasoning is mediated by sex differences, favoring males, in both spatial cognition and computational fluency. The latter was the stronger of these two mediators, but this result might be due to the arithmetical content of the word problems used in this study. It is possible that spatial abilities will be a stronger mediator than computational abilities for reasoning tests that include word problems that are less heavily dependent on arithmetic abilities (Johnson, 1984), but, again, this is a yet-to-be-tested hypothesis.

The last issue to be addressed is the finding that the relation between IQ and performance on the MRT was stronger in females than in males. The pattern suggests that the MRT is a more difficult test for females than for males, in keeping with the mean sex difference and the finding that the magnitude of the correlation between IQ and tests of cognitive ability increases with increases in the difficulty of the cognitive test (Jensen, 1998; Linn & Petersen, 1985; Voyer et al., 1995). If replicated, the pattern might also be interpreted as suggesting that the male advantage in three-dimensional spatial abilities reflects a fundamental sex difference (see Geary, 1998, for further discussion). In other words, most males, regardless of their IQ level, develop relatively—in comparison to the average female—good three-dimensional spatial abilities, whereas the development of the same level of competency might require higher levels of intelligence in females. If so, then this would suggest that complex three-dimensional spatial competencies may reflect more of a learned skill for many females, but a more inherent skill for most males.

REFERENCES

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences,* **11,** 169–232.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin,* **88,** 588–606.

Brown, D. G. (1958). Sex-role development in a changing culture. *Psychological Bulletin,* **55,** 232–242.

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., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. *Journal for Research in Mathematics Education,* **30,** 20–46.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills and internalized beliefs and anxieties. *Developmental Psychology,* **33,** 669–680.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology,* **31,** 697–705.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests 1976.* Princeton, NJ: Educational Testing Service.

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,** 6–11.

Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research,* **65,** 22–50.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, DC: American Psychological Association.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences,* **19,** 229–284.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences.* Washington, DC: American Psychological Association.

Geary, D. C. (1999). Sex differences in mathematical abilities: Commentary on the math-fact retrieval hypothesis. *Contemporary Educational Psychology,* **24,** 267–274.

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., Fan, L., & Bow-Thomas, C. C. (1992). Numerical cognition: Loci of ability differences comparing children from China and the United States. *Psychological Science,* **3,** 180–185.

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging,* **8,** 242–256.

Geary, D. C., & Widaman, K. F. (1987). Individual differences in cognitive arithmetic. *Journal of Experimental Psychology: General,* **116,** 154–171.

Geary, D. C., & Widaman, K. F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence,* **16,** 47–80.

Harnisch, D. L., Steinkamp, M. W., Tsai, S. L., & Walberg, H. J. (1986). Cross-national differences in mathematics attitude and achievement among seventeen-year-olds. *International Journal of Educational Development,* **6,** 233–244.

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

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Johnson, E. S. (1984). Sex differences in problem solving. *Journal of Educational Psychology,* **76,** 1359–1371.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8.* Chicago, IL: Scientific Software.

Lewis, A. B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology,* **81,** 521–531.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development,* **56,** 1479–1498.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin,* **107,** 385–393.

Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology,* **26,** 254–263.

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist,* **45,** 513–520.

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

Marshall, S. P., & Smith, J. D. (1987). Sex differences in learning mathematics: A longitudinal study with item and error analyses. *Journal of Educational Psychology,* **79,** 372–383.

Raven, J. C., Court, J. H., & Raven, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* London: H. K. Lewis.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Marchant, H., III. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology,* **24,** 181–266.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin & Review,* **3,** 21–36.

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

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development,* **59,** 833–851.

Stevenson, H. W., Lee, S. Y., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the United States. *Child Development,* **61,** 1053–1066.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills,* **47,** 599–604.

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.

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

Received June 3, 1999; revised October 15, 1999