

# The Influence of Spatial Ability on Gender Differences in Mathematics College Entrance Test Scores Across Diverse Samples

M. Beth Casey, Ronald Nuttall,  
and Elizabeth Pezaris  
Boston College

Camilla Persson Benbow  
Iowa State University

The relationship between mental rotation ability and gender differences in Scholastic Aptitude Test—Math (SAT—M) across diverse samples was investigated. Talented preadolescents, college students, and high- and low-ability college-bound youths, totaling 760, were administered the Vandenberg Mental Rotation Test. Gender comparisons showed male outperforming female students in both mental rotation and SAT—M for all 3 high-ability groups but not for the low-ability group. For all female samples, mental rotation predicted math aptitude even when SAT—Verbal was entered first into the regression. For male samples, the relationship varied as a function of sample. When mental rotation ability was statistically adjusted for, the significant gender difference in SAT—M was eliminated for the college sample and the high-ability college-bound students. This suggests that spatial ability may be responsible in part for mediating gender differences in math aptitude among these groups.

The interpretation of cognitive gender differences is often contentious within the scientific as well as lay communities. One area of vigorous debate involves the finding of a male advantage in certain mathematical domains (Benbow, 1988; Rosser, 1989). This gender difference has been demonstrated cross-culturally, with the largest male advantage found in geometry and word problems. Moreover, gender differences in math tend to be largest in countries with the highest achieving students and in countries with the largest gender gap in experience and training (Harnisch, Steinkamp, Tsai, & Walberg, 1986; Steinkamp, Harnisch, Walberg, & Tsai, 1985). The smallest differences, on the other hand, are found in countries with lower achieving students in mathematics, which includes the United

States. Within the United States, poorly educated female students outperform their male peers, but as the level of education increases, the male advantage in mathematics emerges (Moore & Smith, 1987). Thus, the male advantage in mathematical performance is found primarily for high-ability samples and on measures of mathematical problem-solving, such as the mathematics section of the Scholastic Aptitude Test (SAT—M; Benbow, 1988; Burnett, Lane, & Dratt, 1979; Hyde, Fennema, & Lamon, 1990; Lubinski & Benbow, 1992).

Scores on college entrance mathematics tests, such as the SAT—M, are of particular importance, because SAT—M scores can affect college opportunities and are predictive of later performance in math- and science-related college curricula (Benbow & Lubinski, 1993). Given the female disadvantage on the SAT—M, it is especially useful to identify the skills that contribute to the success of those women who do well on these math exams. It is also important to determine whether gender differences in these “underlying” skills do, in fact, contribute to gender differences in math performance. These are the goals of the present study. Specifically, we attempted to establish evidence for a relationship between one type of spatial skill (mental rotation) and performance on the SAT—M for female students. If such a relationship can be documented for female students, the goal then becomes the determination of whether significant male–female differences in mathematics can be eliminated when this spatial skill is statistically adjusted for.

M. Beth Casey, Ronald Nuttall, and Elizabeth Pezaris, Program in Developmental and Educational Psychology, Boston College; Camilla Persson Benbow, Department of Psychology, Iowa State University.

This article is based in part on a paper presented in February 1994 by M. Beth Casey at the 22nd Annual Meeting of the International Neuropsychological Society, Cincinnati, Ohio, in conjunction with the present authors and M. L. Harwood.

We particularly acknowledge the help of a group of doctoral students at Boston College in the literature review and data analysis for the findings on the college sample: Maureen Buckley, Kathleen Gallivan, Barbara Gray, Kevin Gillespie, Mary Lou Harwood, Pam Healey, Geri Hull, Gail Martino, Patty Murphy, Ximena Perez De Castillo, Susan Kayhoe, Jennifer Rapkin, Christopher Recupero, and Stephen Rossetti, as well as Sheri Berenbaum for her statistical suggestions and Edwin Tucker for his helpful comments on the article. We also thank the Melrose (Massachusetts) school system for their support in making this research possible.

Correspondence concerning this article should be addressed to M. Beth Casey, Program in Developmental and Educational Psychology, 201 Campion Hall, Boston College, Chestnut Hill, Massachusetts 02167.

## Gender Differences in Mental Rotation Ability

In the past, a major battle was fought over whether men do, in fact, excel in spatial ability compared with women (Burnett, 1986; Caplan, MacPherson, & Tobin, 1985). The answer now appears to be that it depends on the type of spatial ability, with mental rotation ability showing strong gender differences favoring men (Casey & Brabeck, 1989, 1990). This skill involves

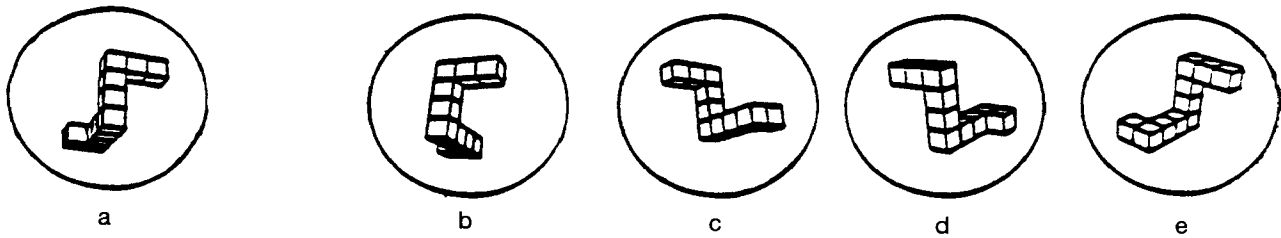


Figure 1. Sample mental rotation figures. Choices a and b are identical (but rotated) to the choice on the left (Vandenberg, 1978).

mentally rotating a two-dimensional representation of an object in three-dimensional space (see Figure 1). Linn and Petersen (1985) conducted a meta-analysis of the literature wherein they documented both large and consistent average gender differences, and a more recent meta-analysis confirmed that this large gender difference has remained stable over the last 17 years (Masters & Sanders, 1993).

A number of researchers have proposed that gender differences in spatial skills may contribute to or be related to the gender differences on the SAT-M (Benbow, 1988; Burnett et al., 1979; Casey, Nuttall, Harwood, Pezaris, & Benbow, 1994). Its relevance for gender differences in mathematics has not yet been established, however. Linn and Petersen's conclusion in 1986 is appropriate today as then: "Evidence for direct relationships between gender differences in spatial ability and gender differences in mathematics and science is largely lacking" (Linn & Petersen, 1986, p. 67). The present article was designed to examine this issue specifically for the spatial skill of mental rotation.

### Criteria for Addressing Gender Issues

Several researchers have proposed criteria that need to be addressed to make the connection between a possible mental rotation and math relationship for female students and gender differences in these abilities. Rosenthal (1988) proposed a series of critical statistical tests. These require demonstrating a significant correlation between the mediator (in this case, mental rotation ability) and the dependent measure (SAT-M). Next, it is necessary to document significant gender differences for both the mediator and for the dependent measure and, then, to document that the significant gender differences in the dependent measure are eliminated when the mediator is covaried out.

In this study, we addressed these criteria by investigating three questions: (a) Does mental rotation ability significantly predict for math aptitude in women? (b) Are there gender differences in both mental rotation ability and math aptitude, favoring male students? and (c) Are these significant gender differences in math aptitude eliminated when mental rotation ability is statistically controlled through an analysis of covariance?

Linn and Petersen (1986) suggested additional criteria for evaluating research in this area. We incorporated these into our design as well. They consist of (a) clarity in the types of measures used, (b) statistically adjusting for general scholastic apti-

tude, and (c) documenting findings across diverse samples. These issues are discussed in more detail below.

### Choice of Measures

In the past, studies that have examined the relation between spatial ability and math aptitude for both male and female students have shown no consistent pattern of results (Burnett et al., 1979; Casey, Pezaris, & Nuttall, 1992; Connor & Serbin, 1985; Ethington & Wolfe, 1984; S. A. Gallagher, 1989; Pattison & Grieve, 1984; Tartre, 1990; Weiner & Robinson, 1986). This is not surprising, given that the studies vary widely in age and ability level of students as well as choice of predictors entered into the regression equation. Most important, the instruments used for measuring both spatial and math abilities have not been consistent across studies.

In the present study, we selected three-dimensional mental rotation ability as the spatial skill to be investigated primarily for empirical reasons: It reveals the largest gender differences in spatial ability. As our dependent measure of mathematics aptitude, we selected the SAT-M, because it serves as a gatekeeper into college and taps mathematical reasoning, which is the type of math ability exhibiting the largest gender differences favoring male students (Lubinski & Benbow, 1992).

### Controlling for Scholastic Aptitude

Linn and Petersen (1986) argued that even a strong correlation between mental rotation ability and mathematical aptitude does not provide convincing evidence that spatial ability makes an important contribution to math aptitude. They proposed that this relationship could be mediated by more general ability requirements shared by both kinds of tests and pointed out that similar correlations are found between both these measures and tests of vocabulary. Thus, deciphering the role of verbal ability becomes another critical test in this type of research. The present study addresses this issue by entering verbal aptitude scores into the regression equation first, and then examining the contribution made by mental rotation ability to math aptitude beyond the effects of verbal ability.

### Generality of the Findings

Linn and Petersen (1986) noted a dearth of data in the field on this issue and pointed out that it is important to determine

how widespread the relationship between mental rotation ability and college entrance math test scores might be. This concern is not new. For example, a prior study by Burnett et al. (1979) directly examined the mental rotation and SAT-M relationship and found a significant relationship for both male and female students. Yet, Burnett et al. noted that their college sample was a high-ability group mainly majoring in science and engineering (with combined SAT scores of over 1,300) and, thus, were cautious in generalizing their results.

Consequently, in the present investigation, the relationship between mental rotation ability and SAT-M was explored for 760 students (327 male and 433 female) across diverse samples. Four types of samples in all were investigated. First, a college sample was specifically chosen so as to include students with diverse majors, not just the science and engineering students that were included in the Burnett et al. (1979) study. A talented sample also was included and consisted of mathematically talented youths who were invited to participate in a math/science summer training program on the basis of their scores on college aptitude tests given at an early age as part of a talent search (i.e., at least the top 1% nationally at age 13). Finally, we decided to examine both high- and low-ability students within another population because Hyde et al.'s (1990) meta-analysis showed that ability level is an important variable to consider when assessing gender differences in math aptitude. We did this for statistical reasons as well (cf. Lubinski & Humphreys, 1990). College-bound students in a high school sample from a middle-income suburban community were divided into high- and low-ability groups on the basis of level of performance on the verbal score on the SAT.

## Method

### Sample

One sample consisted of 274 undergraduates (195 women and 79 men) at two liberal arts colleges in the Northeast. They were recruited from classes offered by different departments to ensure a range of majors and were tested on their mental rotation ability in groups of 10 to 20. (Note: There were more women recruited than men, based on the needs of prior research on this sample [Casey & Brabeck, 1989, 1990].) The majors in the study included math/science and applied math/science fields, the social sciences, education, nursing, and the humanities. High school SAT scores were obtained from the student files.

The second sample consisted of mathematically talented preadolescents participating in a summer math/science training program in the Midwest. They included seventh to ninth graders ( $N = 186$ ; girls:  $n = 84$ ,  $M$  age = 13.25 years,  $SD = 0.81$ ; boys:  $n = 102$ ,  $M$  age = 13.38 years,  $SD = 0.96$ ) who were either recruited from a national talent search program or a statewide teacher selection program to participate in a summer math/science training program. Students who met any of the following aptitude test scoring criteria were invited to participate in the summer program and its associated research studies: (a) at least 20 on the American College Test (ACT) Math or ACT Science Reasoning Tests, (b) at least 500 on the Math SAT (College Board Aptitude Test), (c) 430 on the Verbal SAT (SAT-V), or (d) 930 on the combined SAT scores. (Note: For 34 boys and 23 girls, ACT scores were available in place of SAT scores. These ACT scores were converted to SAT equivalents by the following formula: SAT estimate =  $200 + 20 \times$  [ACT score].)

The criteria for inclusion in the present study consisted of 500 on the

SAT-M, 430 on the SAT-V, or a combined score of 930. Only students who had not participated in the program in prior years were included in the present study (to avoid confounding by prior math training). Students were administered the mental rotation test at the beginning of the summer program before their course work. In the analyses of variance (ANOVAs) for this sample, age was covaried out to adjust for variations in age among the students in the program.

The last two samples were taken from college-bound students who were enrolled in a middle-income suburban high school in the Northeast and who had elected to take the SAT. The mental rotation test was administered at the end of their sophomore year during their geometry class period. The data were collected over 2 consecutive years. At the end of the senior year, the SAT scores were obtained from the student files. These students were divided into two separate groups on the basis of their SAT-V scores: a high-ability and a low-ability group. It is important to have these two subsamples, because the magnitude of the gender difference in mathematics aptitude often varies with overall ability level of the sample (Hyde et al., 1990). Students who scored at or above 480 on the SAT-V were designated the high-ability sample (50 girls and 45 boys). Students scoring below 480 on the SAT-V were designated the low-ability sample (104 girls and 101 boys). With this cutoff, 30% of the sample was included in the high-ability group. (Note: Verbal scores were used to make the cutoff instead of total SAT to avoid overlap with the dependent measure [Lubinski & Humphreys, 1990].)

### Measures

**Mental rotation.** Mental rotation ability was assessed through the Vandenberg Test of Mental Rotation (Vandenberg & Kuse, 1978), which uses figures originally designed by Shephard and Metzler (1971). Standard procedures were followed in administering and scoring the Vandenberg test (maximum score = 40), with the exception that students were allowed 5 min to complete each of the two sections of the test rather than 3 min (10 items in each section). The task involves correctly matching two out of four choices to a standard. The two correct choices are identical to the standard but were presented in rotated positions (see Figure 1). The distractor stimuli vary from the standard either as mirror images or in terms of slight variations in features of the figure.

**SAT scores.** The SAT-M and SAT-V scores developed by the Educational Testing Service were used in the study. Most students take these tests more than once. For the college-bound high school students and for the college students, the highest SAT-M and SAT-V scores were selected for use in the study in accordance with college admissions practices. The initial SAT scores were used for the talented students as these were used as program entrance requirements.

The SAT is a 3-hr test administered under carefully controlled conditions and is divided into three parts: verbal (SAT-V), math (SAT-M), and a test of standard written English (not included here). The SAT-V consists of sections that include analogies, reading comprehension, antonyms, and sentence completion. The SAT-M assesses the understanding and application of mathematical principles, as well as numerical reasoning ability; it assumes knowledge of basic arithmetic, geometry, and algebraic concepts. The technical quality of the test is state of the art, with high reliability of recent (1990s) forms for both SAT-M and SAT-V. Correlations between SAT-M and SAT-V have been in the high 60s, a finding that suggests overlapping scholastic aptitude skills (R. C. Cohen, Swerdilick, & Smith, 1992).

## Results

### *Mental Rotation as a Predictor of Math Aptitude*

For all four samples, a regression analysis was performed for the male and female students separately, using math aptitude

score (SAT-M) as the dependent measure. To control for scholastic aptitude factors, we entered the verbal aptitude score (SAT-V) in the equation before the Vandenberg Mental Rotation score. (For students in the talented preadolescent sample, age was entered into the equation before the verbal ability score.) Because removing cases with residuals  $\pm 2$  standard errors from the regression slope (the outliers) resulted in only a minor effect on the regression parameters, the results reported are for the entire samples without excluding any cases.

*Findings for female samples.* For the female students, the correlations between mental rotation and SAT-M scores were consistent, ranging from .35 to .38 in the four samples (see Table 1). This consistent pattern is further confirmed in the regression analyses for the four samples. Mental rotation scores, when entered second in the equation, made a significant contribution to SAT-M for the female students in every sample above the effects of SAT-V. Table 1 presents regression and correlation information for both genders. The b-weights for mental rotation scores, after verbal ability had been entered into the equations, ranged from 2.62 to 3.43. (Note: A regression slope of 3.43 would indicate that for each additional item correct on the mental rotation test, the SAT-M score improved by 3.43 points.) The increment in  $R^2$  for mental rotation ranged from .06 to .12 for the four female samples and was significant in all four. The confidence intervals for the mental rotation b-weights indicated that there were no significant differences in slope for the regression equations for the four female samples.

*Findings for male samples.* Though not central to the purpose of this article, the mental rotation and math aptitude relationship was examined for the male as well as for the female samples. In contrast to the female samples, the correlations between mental rotation and SAT-M for the male samples showed no consistent pattern, with correlations ranging from  $-.03$  to  $.54$  (see Table 1). The regression analyses also revealed no consistent relationships. For the high- and low-ability college-

bound samples, mental rotation did make a significant contribution to predicting SAT-M scores beyond the effects of SAT-V: In fact, the male students in the high-ability college-bound sample showed the strongest relationship across all groups. Yet, in the talented and college samples, mental rotation scores did not significantly contribute to the prediction equation above SAT-V. Examination of the confidence intervals for the mental rotation b-weights among the four male samples further documented this inconsistency among the male samples. (Note: The slopes for the regression equations were varied for the male samples with high-ability college-bound students showing a significantly higher b-weight than all the other groups; the low-ability college-bound students showed a significantly higher b-weight only in comparison with the talented sample.)

### Gender Differences in Abilities

*Math aptitude.* There was a significant gender difference in SAT-M scores among all the high-ability samples: college sample,  $F(1, 272) = 4.47, p = .035$ ; talented sample,  $F(1, 183) = 21.39, p < .001$ ; and high-ability college-bound sample,  $F(1, 93) = 4.50, p = .037$ ; but not in the low-ability college-bound sample. This SAT-M difference favored the male samples, with male students scoring higher than female students in all the high-ability groups. Presented in Table 2 are the means and standard deviations for the different samples. (Note: The ranges for SAT-M scores were 310-780 for the college sample men and 300-800 for the college sample women; 390-800 for the talented boys and 320-760 for the talented girls; 390-750 for the high-ability college-bound boys and 360-740 for the high-ability college-bound girls; and 260-690 for the low-ability college-bound boys and 220-680 for the low-ability college-bound girls.)

The effect size,  $d$  (see Hedges & Becker, 1986), for the gender difference was large for the talented sample (.70), moderate for

Table 1  
Vandenberg Mental Rotation Test (MR) as a Predictor of Scholastic Aptitude Test—Math (SAT-M) as a Function of Gender

Sample	$r$ SAT-M with MR	$B$ weights		95% CI for MR	$R^2$ increase in MR	Total $R^2$
		SAT-Verbal	MR			
Female students						
Talented <sup>a</sup>	.35**	0.18	3.43	1.45-5.41	.12**	.20**
College	.35**	0.52	2.78	1.43-4.12	.06**	.32**
High ability <sup>b</sup>	.38**	0.68	3.06	1.08-5.04	.11**	.44**
Low ability <sup>b</sup>	.35**	0.96	2.62	1.03-4.22	.06**	.43**
Male students						
Talented <sup>a</sup>	$-.03$	$-0.01$	$-0.49$	$-2.33-1.36$	.00	.02
College	.13	0.45	0.36	$-2.28-3.00$	.00	.19**
High ability <sup>b</sup>	.54**	0.53	4.62	2.56-6.68	.28**	.43**
Low ability <sup>b</sup>	.25**	0.83	2.16	0.54-3.78	.05**	.30**

Note. CI = confidence interval.

<sup>a</sup> Age was entered first into the regression equation. <sup>b</sup> College-bound high school sample.

\*\*  $p < .01$ .

Table 2  
Scholastic Aptitude Test—Math (SAT-M) Mean Scores Both Alone and Adjusted for Mental Rotation (MR) Ability

Sample	SAT-M						<i>d</i>
	Male students			Female students			
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	
Alone							
Talented <sup>a</sup>	591 <sub>a</sub>	88	102	530 <sub>a</sub>	87	84	0.70
College	574 <sub>b</sub>	92	79	547 <sub>b</sub>	94	195	0.29
High ability <sup>b</sup>	592 <sub>b</sub>	85	45	557 <sub>b</sub>	80	50	0.42
Low ability <sup>b</sup>	450	98	101	439	98	104	0.11
Adjusted for MR scores							
Talented <sup>a</sup>	586 <sub>a</sub>			535 <sub>a</sub>			0.60
College	560			561			-0.01
High ability <sup>b</sup>	581			568			0.16
Low ability <sup>b</sup>	448			440			0.08

Note. Means sharing a similar subscript indicate significant sex differences (subscript a,  $p < .01$ ; subscript b,  $p < .05$ ).

<sup>a</sup> Age adjusted. <sup>b</sup> College-bound high school sample.

the high-ability adolescent sample (.42), and small for the college sample (.29; J. Cohen & Cohen, 1975). In contrast, for the low-ability college-bound sample,  $d$  was only .11 (see Table 2). (Note:  $d$  expresses the size of the gender difference in standard deviation units [Hyde, 1993].)

**Mental rotation ability.** The same pattern of gender differences was found for mental rotation as was found for SAT-M (see Table 3). In every sample, the male students scored higher than did the female students, except for the low-ability college-bound students: college sample,  $F(1, 272) = 57.74, p < .001$ ; talented sample,  $F(1, 183) = 27.84, p < .001$ ; high-ability college-bound sample,  $F(1, 93) = 8.93, p < .004$ ; and low-ability college-bound sample,  $F(1, 203) = .25, p > .05$ . (Note: The ranges for mental rotation scores were between 8–40 for the college men and 0–40 for the college women; 6–40 for the talented boys and 2–38 for the talented girls; 2–40 for the high-ability college-bound boys and 0–40 for the college-bound girls; and 0–40 for the low-ability college-bound boys and 2–40 for the low-ability college-bound girls.)

The effect size of the gender difference in mental rotation was large for the talented ( $d = .79$ ) and college samples ( $d = 1.01$ ) and moderate for the high-ability college-bound sample ( $d = .61$ ). However,  $d$  was only .07 for the low-ability college-bound students (see Table 3).

**Verbal aptitude.** For the verbal aptitude scores, the pattern of relationships differed from those obtained for mental rotation ability and SAT-M. An ANOVA revealed no statistically significant differences for the college sample (men,  $M = 504, SD = 89, n = 79$ ; women,  $M = 499, SD = 81, n = 195$ ), the talented sample (boys,  $M = 522, SD = 112, n = 102$ ; girls,  $M = 525, SD = 122, n = 84$ ), and the high-ability college-bound sample (boys,  $M = 554, SD = 60, n = 45$ ; girls,  $M = 553, SD = 64, n = 50$ ). There was, however, a significant gender difference in verbal ability for the low-ability college-bound sample favoring

girls,  $F(1, 203) = 4.44, p = .036$  (boys,  $M = 385, SD = 57, n = 101$ ; girls,  $M = 401, SD = 57, n = 104$ ).

### Gender Differences in Adjusted Scores

**Math aptitude adjusted for mental rotation.** Next, we performed analyses to determine whether the significant gender differences in SAT-M scores for the three high-ability samples would be eliminated when mental rotation ability was covaried out. An analysis of covariance (ANCOVA) was performed using mental rotation score as a covariate. When mental rotation ability was statistically adjusted for, the significant gender difference was eliminated for the college sample,  $F(1, 272) = .02, p > .05$ ; and the high-ability college-bound sample,  $F(1, 92) = .64, p > .05$ . For the very highly select sample of preadolescents selected through a talent search, however, a significant gender difference still remained even after statistically adjusting for mental rotation ability,  $F(1, 183) = 13.97, p < .001$ . Table 2 presents the math aptitude means for the different samples when adjusted for the mental rotation score. (Note: For the low-ability college-bound sample, again there were nonsignificant gender differences in the ANCOVA, as was found in the initial ANOVA.)

We again computed the effect size for the gender difference in math aptitude for each of the samples but with mental rotation ability statistically adjusted for. For the talented sample, the effect size remained sizable at .60 (from .70). In contrast, for the college sample, the effect size was reduced to -.01 (from .29); for the high-ability college-bound students, the effect size was reduced to .16 (from .42). (Note: For the low-ability college-bound students, the effect size remained low at .08.)

**Mental rotation ability adjusted for math aptitude.** Finally, we performed analyses to determine whether the significant gender differences in mental rotation ability would be eliminated for the three high-ability samples when SAT-

Table 3  
Mental Rotation (MR) Mean Scores Both Alone and Adjusted for Scholastic Aptitude Test—Math (SAT-M)

Sample	MR scores						<i>d</i>
	Male students			Female students			
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	
Alone							
Talented <sup>a</sup>	27.43 <sub>a</sub>	9.54	102	20.14 <sub>a</sub>	8.83	84	0.79
College	25.76 <sub>a</sub>	7.35	79	17.53 <sub>a</sub>	8.39	195	1.01
High-ability <sup>b</sup>	26.98 <sub>a</sub>	9.73	45	21.28 <sub>a</sub>	8.86	50	0.61
Low-ability <sup>b</sup>	20.81	10.19	101	20.13	9.28	104	0.07
Adjusted for SAT-M							
Talented <sup>a</sup>	27.01 <sub>a</sub>			20.57 <sub>a</sub>			.72
College	25.42 <sub>a</sub>			17.88 <sub>a</sub>			.93
High-ability <sup>b</sup>	26.05 <sub>a</sub>			22.21 <sub>a</sub>			.41
Low-ability <sup>b</sup>	20.65			20.28			.04

Note. Means showing a subscript indicate significant sex differences,  $p < .01$ .

<sup>a</sup> Age adjusted. <sup>b</sup> College-bound high school sample.

M was covaried out. An ANCOVA was performed using SAT-M as a covariate for the gender comparison of mental rotation ability. When math aptitude was adjusted for, the significant gender difference still remained for all three samples: for the college sample,  $F(1, 271) = 52.24, p < .001$ ; for the high-ability college-bound sample,  $F(1, 92) = 4.85, p = .03$ ; and for the talented sample,  $F(1, 182) = 20.15, p < .001$ . Table 3 presents the mental rotation means for the different samples when adjusted for math aptitude score. (Note: For the low-ability college-bound sample, again there were nonsignificant gender differences in the ANCOVA, as was found in the initial ANOVA.) Furthermore, the effect size for the gender differences in mental rotation ability for each of the samples remained essentially unchanged for mental rotation ability when SAT-M was covaried out. For the talented ( $d = .72$ ) and college samples ( $d = .93$ ), the effect size remained large; for the high-ability college-bound sample, the effect size remained moderate ( $d = .41$ ). (Note: For the low-ability college-bound students, the effect size remained low at .04.)

### Discussion

The present study consisted of a series of critical tests for determining whether mental rotation ability mediates gender differences in SAT-M scores. In the discussion, we first briefly present the findings for the four questions examined in the article, based on the criteria proposed by Rosenthal (1988) when documenting mediators for gender differences. Then, we discuss each finding in greater depth. Finally, we discuss the wider implications.

#### *Addressing the Questions Examined in the Study*

We first asked, (a) Does mental rotation ability significantly predict math aptitude in female students? We found that the female students in all four samples showed a significant relationship between mental rotation skills and math aptitude. We then asked, (b) Is there a similar relationship in male students? Some samples showed a significant relationship, whereas others did not. Next we asked, (c) Are there gender differences in both mental rotation ability and math aptitude, favoring male students? Yes there are. We found that for both the mental rotation and math abilities, male students excelled over female students, but only for the three high-ability samples. Finally, we asked, (d) When mental rotation ability is statistically adjusted for, are the significant gender differences in math aptitude eliminated? For both the high-ability college-bound students and the college students, the significant gender differences in SAT-M were eliminated when scores were adjusted for the ability to mentally rotate images. This was not the case for the highly talented sample.

Thus, we applied the criteria set up by Rosenthal (1988) for establishing evidence for mediating variables contributing to gender differences in math ability. In doing so, we found evidence supporting the contention that some of the mental processes tapped by a test of mental rotation abilities can account for gender differences in mathematical aptitude. Next, we turned to the criteria proposed by Linn and Petersen (1986) and

addressed the major concerns they expressed regarding research on this type of gender difference. We address these concerns as we discuss the findings in greater depth.

#### *Mental Rotation as a Predictor of Math Aptitude*

*Findings for female samples.* First, on the basis of Linn and Petersen's (1986) criteria, we documented that for female students, the mental rotation and math aptitude relationship is, indeed, a widespread phenomenon, not limited to science and engineering majors (cf. Burnett et al., 1979). We found that the female students across all four samples showed a strikingly consistent relationship between mental rotation skills and math aptitude (with all correlations for these two measures clustering between .35 and .38, a medium effect size). Thus, the present study has revealed the generality of this relationship for a wide range of both high- and low-aptitude female students who were among the 433 women and girls participating in the study.

Second, in accordance with Linn and Petersen's (1986) criteria, we demonstrated that the association between mental rotation and mathematical abilities among female students cannot be attributed solely to verbal or general scholastic aptitude factors. For female students across all samples, mental rotation skills predicted math aptitude scores (accounting for 6% to 12% of the additional variance) even when verbal ability was statistically covaried out.

Linn and Petersen (1986) further recommended that longitudinal data need to be collected. In the two college-bound samples, mental rotation ability was assessed in the sophomore year and SAT scores were obtained 2 years later in the senior year. Thus, in this study, we have provided some preliminary longitudinal evidence supporting the long-term predictive power of mental rotation ability for female students.

*Findings for male samples.* The reasons for variations in math performance within the male samples were less central to the goals of the study. Our purpose was to identify the skills affecting female variations in math performance, because the ultimate question of interest is how to intervene to improve math performance in female students. For comparison purposes, we did, nevertheless, obtain the data and found that the results for the male students were inconsistent: (a) In both subgroups of male college-bound students (both high and low ability), mental rotation ability significantly contributed to the prediction of math aptitude beyond the effects of verbal ability. (b) For the two other high-ability male samples (the talented and college samples), no such significant relationship was found.

The reasons for this inconsistency in pattern of results for the male students are not clear. It cannot be attributed to ability levels or ceiling effects. First, the high-ability group within the college-bound sample had levels of SAT-M and mental rotation scores that were equivalent to the other two high-ability groups. Second, Burnett et al. (1979) documented a similar relationship for their high-ability male sample. We checked the data in relation to the issue of restriction of range and found that this effect also cannot explain the findings. There was restriction of range among the college male sample but not among the college-bound high-ability group nor among the talented preadolescents. The highest score possible was 40 on the Vandenberg

Mental Rotation Test. Even for the highest scoring male group, with a mean of 27, there were less than 10% of the male students who had the maximum possible value of 40. Furthermore, the standard deviations remained reasonably similar across all female and male samples, ranging from 7.35 to 10.19, with no consistent pattern based on gender or a significant mental rotation and math relationship. Finally, the pattern of findings for the male students also does not appear to be an artifact of the 30%–70% cutoff point used to divide the high- and low-ability samples. We reanalyzed the data using a 33%–67% and a 25%–75% cutoff and still obtained the same results for both male and female students. Perhaps the longitudinal nature of the data for the college-bound sample contributed to the differences in findings.

### *Gender Differences in Math Aptitude and Mental Rotation Ability*

On the basis of the prior meta-analysis of Hyde et al. (1990), as well as the findings from cross-cultural research (Harnisch et al., 1986; Steinkamp et al., 1985), one might have predicted that we would find gender differences in math aptitude for our higher ability but not our lower ability students. Indeed, this pattern was reflected in the present findings for both mathematical and mental rotation abilities.

Burnett et al. (1979) found that the male advantage on the math aptitude test was reduced to nonsignificant levels for their highly select science and engineering students when they covaried out mental rotation ability. Their initial finding supported the importance of spatial ability in influencing gender differences in math aptitude. As Burnett et al. pointed out, despite the fact that causative statements cannot be conclusively made with this type of correlational data (J. Cohen & Cohen, 1975), these findings clearly are consistent with a spatial hypothesis.

The present study not only provides further support for this hypothesis but also extends these findings to two types of high-ability students who also showed male–female differences on the math aptitude test: (a) the diverse college sample that included students across a wide range of majors and (b) the high-ability college-bound seniors. Thus, for these high-ability samples, as well as for the Burnett et al. (1979) high-ability sample, when mental rotation skills were statistically adjusted for, the male math advantage was reduced to nonsignificant levels. It is interesting to note that the reverse was not found. When math aptitude scores were covaried out of mental rotation scores, the significant male advantage in mental rotation ability still remained for all three high-ability samples.

It is important to point out, however, that there was a different pattern of results for the talented youths compared with the other high-ability samples. For this precocious sample, gender differences in math aptitude remained even when mental rotation scores were statistically adjusted. Other factors may have been operating, however. Because these seventh to ninth graders represent the top 1% of students in the United States, this is clearly a select group, much more select than the other high-ability samples. Furthermore, there was at least one additional way in which the talented youths differed from the other samples. Because most were tested at 13 years of age, they could not

draw on the algorithms taught to them in school for solving the math problems as could the students in the other samples. That is, the talented youths took the SAT at a point before exposure to the high school math curriculum on which the math SAT is partially based. Therefore, when the talented students solve problems on SAT–M, they may be required to invent their own solutions rather than to recall them from memory, and hence the SAT–M may tap different skills for this group.

A recent report by the Educational Testing Service (A. M. Gallagher, 1992) examined gender differences among mathematically talented high school examinees who scored above 650 on the SAT–M. An item analysis revealed that the majority of items that favored male students required the use of mathematical insight, whereas all the items favoring female students required standard algorithmic solutions. Thus, it is possible that, for the male advantage in math skills among this younger talented sample, gender differences in insight may contribute above and beyond the contributions of gender differences in mental rotation ability (see Kimball, 1989). Alternatively, Benbow (1988) has suggested that gender differences among talented youths are due to the overabundance of high-scoring male students. That is, the distribution of SAT–M scores is skewed for the talented male but not for the talented female students.

### *Wider Implications of the Findings*

It is important to recognize that, at most, only 12% of the variance in math aptitude scores was explained by mental rotation skills across the female samples. In the regression analyses, mental rotation ability explained 6% to 12% of the variance beyond the effects of verbal ability, which in turn accounted for 5% to 26% of the variance for the different female samples. Irrespective, there is a substantial amount of variance still unaccounted for. Individual differences in other factors, such as number of math courses taken (Pallas & Alexander, 1983), math attitudes (Eccles & Jacobs, 1986), and autonomous learning (Fennema & Petersen, 1985; Kimball, 1989) may have contributed as well.

Moreover, it is not clear what the mechanism might be that explains why gender differences in three-dimensional mental rotation skills serves as a mediator for gender differences in math aptitude among high-ability college and college-bound students. There are several possible underlying mechanisms. Such a relationship could occur simply from an overlap in the specific skills required for the two tests (Linn & Petersen, 1986). This is clearly a possibility, because one quarter to one third of the items on the SAT–M are geometry items. However, this is unlikely to be the whole story, because mental rotation ability also is related to skill at other types of mathematics tasks, such as word problems (Johnson, 1984).

Alternatively, three-dimensional mental rotation ability may be an index for more general, complex spatial reasoning skills (McGuinness, 1993) as well as the general predisposition to use these spatial skills when solving different types of problems (Casey, in press; Pezaris & Casey, 1991). McGuinness (1993) proposed that boys' orientation tends to be more visually based, which allows them to use visual images to make sense of symbolic mathematical formulations. In fact, using a selective inter-

ference paradigm, Pezaris and Casey (1991) documented evidence for such a male–female difference in processing style: When solving mental rotation problems, boys had a greater dependence on spatial strategies than did the majority of girls.

Notably, in this study, we found a subset of female students who, like the male students, differed from the majority of the female samples. This female group also depended to a greater extent on spatial strategies (Pezaris & Casey, 1991). Thus, investigating individual differences within female students may provide the best strategy in the future for understanding how to improve female math skills. We suggest that those female students who have the added flexibility to apply spatial as well as verbal strategies when problem-solving may have an advantage when solving both mental rotation and math problems. Thus, strategy differences within female students may account for the mental rotation and math relationship found for all the female samples. If spatial thinking does provide an edge when solving problems, then skill at using spatial thinking may well be the underlying mechanism for the pattern of relationships and gender differences found here.

## 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*, 169–183.
- Benbow, C. P., & Lubinski, D. (1993). Consequences of gender differences in mathematical reasoning ability and some biological linkages. In M. Haug, R. E. Whalen, C. Aron, & K. L. Olsen (Eds.), *The development of sex differences and similarities in behavior* (pp. 87–110). Boston: Kluwer Academic.
- Burnett, S. A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist*, *41*, 1012–1014.
- Burnett, S. A., Lane, D. L., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence*, *3*, 345–354.
- Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist*, *40*, 786–799.
- Casey, M. B. (in press). Understanding individual differences in spatial ability within females: A nature/nurture interactions framework. *Developmental Review*.
- Casey, M. B., & Brabeck, M. M. (1989). Exceptions to the male advantage on a spatial task: Family handedness and college major as a factor identifying women who excel. *Neuropsychologia*, *27*, 689–696.
- Casey, M. B., & Brabeck, M. M. (1990). Women who excel on a spatial task: Proposed genetic and environmental factors. *Brain & Cognition*, *12*, 73–84.
- Casey, M. B., Nuttall, R., Harwood, M. L., Pezaris, E., & Benbow, C. P. (1994, February). *The impact of mental rotation ability on math aptitude: A comparison of males and females across diverse samples*. Paper presented at the 22nd Annual Meeting of the International Neuropsychological Society, Cincinnati, OH.
- Casey, M. B., Pezaris, E., & Nuttall, R. (1992). Spatial ability as a predictor of math achievement: The importance of sex and handedness patterns. *Neuropsychologia*, *30*, 35–45.
- Cohen, J., & Cohen, P. (1975). *Applied multiple regression for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, R. C., Swerdlik, M. E., & Smith, D. K. (1992). *Psychological testing and assessment: An introduction to test and measurement*. Mountain View, CA: Mayfield.
- Connor, J. M., & Serbin, L. A. (1985). Visual–spatial skill: Is it important for mathematics? Can it be taught? In F. S. Shipman, L. R. Brush, & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation* (pp. 151–174). Hillsdale, NJ: Erlbaum.
- Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs*, *11*, 367–380.
- Ethington, C. A., & Wolfe, L. M. (1984). Sex differences in a causal model of mathematics achievement. *Journal for Research in Mathematics Education*, *15*, 361–377.
- Fennema, E., & Petersen, P. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L. S. Wilkinson & C. B. Marret (Eds.), *Gender influences in classroom interaction* (pp. 17–36). New York: Academic Press.
- Gallagher, A. M. (1992). *Sex differences in problem-solving used by high-scoring examinees on the SAT-M* (College Board Report No. 92–2, ETS RR No. 92–33). New York: College Board Publications.
- Gallagher, S. A. (1989). Predictors of SAT mathematics scores of gifted male and gifted female adolescents. *Psychology of Women Quarterly*, *13*, 191–203.
- Harnisch, D. L., Steinkamp, M. W., Tsai, S., & Walberg, H. (1986). Cross-national differences in mathematics attitude and achievement among seventeen-year-olds. *International Journal of Educational Development*, *6*, 233–244.
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 14–50). Baltimore: Johns Hopkins University Press.
- Hyde, J. S. (1993). Gender differences in mathematics ability, anxiety, and attitudes: What do meta-analyses tell us? In L. A. Penner, G. M. Batshe, H. M. Knoff, & D. L. Nelson (Eds.), *The challenge in mathematics and science education: Psychology's response* (pp. 241–250). Washington, DC: American Psychological Association.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.
- Johnson, E. S. (1984). Sex differences in problem solving. *Journal of Educational Psychology*, *76*, 1359–1371.
- Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin*, *105*, 198–214.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.
- Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67–101). Baltimore: Johns Hopkins University Press.
- Lubinski, D., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math–science pipeline. *Current Directions in Psychological Science*, *1*, 61–66.
- 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.
- Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavioral Genetics*, *23*, 337–341.
- McGuinness, D. (1993). Gender differences in cognitive style: Implications for mathematics performance and achievement. In L. A. Penner, G. M. Batsche, H. M. Knoff, & D. L. Nelson (Eds.), *The challenge in mathematics and science education: Psychology's response* (pp. 251–274). Washington, DC: American Psychological Association.
- Moore, E. G., & Smith, A. W. (1987). Sex and ethnic group differences in mathematics achievement: Results from the national longitudinal study. *Journal for Research in Mathematics Education*, *18*, 25–36.

- Pallas, A. M., & Alexander, K. L. (1983). Sex differences in quantitative SAT performance: New evidence on the differential coursework hypothesis. *American Educational Research Journal*, 20, 165-182.
- Pattison, P., & Grieve, N. (1984). Do spatial skills contribute to sex differences in different types of mathematical problems? *Journal of Educational Psychology*, 76, 678-689.
- Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain & Cognition*, 17, 1-22.
- Rosenthal, R. (1988). Evaluating explanations of sex differences in mathematical reasoning scores. *Behavioral and Brain Sciences*, 11, 207-208.
- Rosser, P. (1989). *The SAT gender gap: Identifying the causes*. Washington, DC: Center for Women's Policy Studies.
- Shephard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three dimensional objects. *Science*, 171, 701-703.
- Steinkamp, M. W., Harnisch, D. L., Walberg, H. J., & Tsai, S. (1985). Cross-national gender differences in mathematics attitude and achievement among 13-year-olds. *Journal of Mathematical Behavior*, 4, 259-277.
- Tartre, L. A. (1990). Spatial skills, gender, and mathematics. In E. Fenema & G. C. Leder (Eds.), *Mathematics and gender* (pp. 27-59). New York: Teachers College Press.
- Vandenberg, S. G. (1978). [Vandenberg Test of Mental Rotation]. Unpublished test.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three dimensional spatial visualization. *Perceptual and Motor Skills*, 48, 599-604.
- Weiner, N. C., & Robinson, S. E. (1986). Cognitive abilities, personality and gender differences in math achievement of gifted adolescents. *Gifted Child Quarterly*, 20, 83-87.

Received March 14, 1994

Revision received November 1, 1994

Accepted November 1, 1994 ■

## Low Publication Prices for APA Members and Affiliates

**Keeping You Up-to-Date:** All APA members (Fellows; Members; Associates, and Student Affiliates) receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*.

High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they can subscribe to the *American Psychologist* at a significantly reduced rate.

In addition, all members and affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential Resources:** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the APA*, the *Master Lectures*, and *Journals in Psychology: A Resource Listing for Authors*.

**Other Benefits of Membership:** Membership in APA also provides eligibility for low-cost insurance plans covering life, income protection, office overhead, accident protection, health care, hospital indemnity, professional liability, research/academic professional liability, student/school liability, and student health.

**For more information,** write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242, USA

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.