Psychological Bulletin
1995, Vol. 117, No. 2, 250–270

0033-2909/95/$3.00

Magnitude of Sex Differences in Spatial Abilities: A Meta-Analysis and Consideration of Critical Variables

Daniel Voyer and Susan Voyer
St. Francis Xavier University

M. P. Bryden
University of Waterloo

In recent years, the magnitude, consistency, and stability across time of cognitive sex differences have been questioned. The present study examined these issues in the context of spatial abilities. A meta-analysis of 286 effect sizes from a variety of spatial ability measures was conducted. Effect sizes were partitioned by the specific test used and by a number of variables related to the experimental procedure in order to achieve homogeneity. Results showed that sex differences are significant in several tests but that some intertest differences exist. Partial support was found for the notion that the magnitude of sex differences has decreased in recent years. Finally, it was found that the age of emergence of sex differences depends on the test used. Results are discussed with regard to their implications for the study of sex differences in spatial abilities.

In 1974, Maccoby and Jacklin published an extensive review of the literature that clearly established the existence of sex differences in spatial abilities favoring males. That review stimulated a considerable amount of research in the ensuing years. However, Maccoby and Jacklin simply summarized the data on the topic and did not provide a precise quantification of the magnitude of these sex differences. Since then, meta-analytic techniques, which allow one to summarize the results of several studies with a single estimate of effect size (see Rosenthal, 1984), have been developed. Hyde (1981) used meta-analysis to estimate the magnitude of the sex differences reported by Maccoby and Jacklin (1974) and found that sex accounted for only 5% of the variance in the spatial tasks they sampled. This would suggest that sex has only a minimal influence on determining spatial test scores, thus supporting Fairweather's (1976) claim that other factors such as handedness, maturation rate, and birth order might be better predictors of spatial performance than is sex.

However, a subsequent meta-analysis carried out by Linn and Petersen (1985) suggested a different explanation for the small percentage of variance in spatial performance accounted for by sex of study participant. These authors performed a meta-analysis on the results of those studies that had been conducted since the Maccoby and Jacklin (1974) review. They used a psychometric as well as a cognitive rationale to classify spatial tests into categories showing homogeneous (or close to homogeneous) effect sizes. Effect size was defined as the mean standardized difference between the scores of males and females on a specific test. Using this procedure, they found three distinct categories of spatial tests, which they labeled spatial perception, mental rotation, and spatial visualization. *Spatial perception* was defined as the ability to determine spatial relations despite distracting information. Tests in this category produced a mean effect size of 0.44 ($p < .05$). *Mental rotation* was defined as the ability to rotate quickly and accurately two- or three-dimensional figures, in imagination. Tests in this grouping showed a mean effect size of 0.73 ($p < .05$). Finally, *spatial visualization* was defined as the ability to manipulate complex spatial information when several stages are needed to produce the correct solution. This type of spatial ability yielded a mean effect size of 0.13 ($p > .05$). Linn and Petersen's (1985) analysis thus indicated that sex differences were significant only in the first two categories and failed to reach significance in the spatial visualization grouping. This result suggests that Hyde (1981) may have underestimated the percentage of variance accounted for by sex in spatial performance by failing to consider the possibility that different types of spatial tests produce sex differences of differing magnitudes. The Linn and Petersen (1985) meta-analysis indicated that sex differences in spatial perception and mental rotation are robust.

Caplan, MacPherson, and Tobin (1985) argued that this conclusion was not warranted. Although most of the issues discussed by Caplan et al. (1985) have been criticized (see Burnett, 1986; Eliot, 1986; Halpern, 1986; Hiscock, 1986; Sanders, Cohen, & Soares, 1986), two of their points are of particular relevance to the present article. They claimed, first of all, that there is little agreement as to what spatial abilities really are. Because many studies do not define the concept of spatial ability, it is virtually impossible to compare the results of several studies

Daniel Voyer and Susan Voyer, Department of Psychology, St. Francis Xavier University, Antigonish, Nova Scotia, Canada; M. P. Bryden, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Part of this study was presented in Daniel Voyer's doctoral dissertation and was supported by a grant from the Natural Science and Engineering Research Council of Canada.

We thank Ann Bigelow, G. P. Brooks, Barbara Bulman-Fleming, Gina Grimshaw, R. W. Johnson, and E. E. Ware for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Daniel Voyer, Department of Psychology, St. Francis Xavier University, P.O. Box 5000, Antigonish, Nova Scotia, Canada B2G 2W5. Electronic mail may be sent via Internet to dvoyer@juliet.stfx.ca.

using different tests, as Linn and Petersen (1985) did, because we do not know how these tests relate to one another and what aspects of spatial skill are measured in a given study. Second, they argued that sex differences are too small and too inconsistent to allow any clear conclusions. On the basis of such reasoning, Caplan et al. (1985) concluded that the existence of sex differences in spatial abilities has not been convincingly demonstrated.

These two important issues raised by Caplan et al. (1985) are addressed in the present study. In order to define spatial ability more clearly, we consider each test as an operational definition of one specific component of spatial abilities. This makes it possible to determine whether sex differences in spatial abilities really do exist or if they exist only in some aspects of spatial performance. We examine Caplan et al.'s (1985) claim that sex differences are small and inconsistent by computing the magnitude and variability of sex-related differences in spatial performance. On the basis of Linn and Petersen's (1985) study, one would expect to find that sex differences in tests that would be classified in the mental rotation and spatial perception groupings would prove to be robust but that no significant sex differences would be found in tests of spatial visualization. However, we also expected that within the groupings defined by Linn and Petersen (1985), the magnitude of sex differences would vary depending on the test used.

Even though it may be demonstrated that sex differences are statistically significant in several spatial tests, some authors believe that such differences are disappearing. For instance, Feingold (1988) presented data showing that sex differences on the Scholastic Aptitude Test (SAT) and the Differential Aptitude Test (DAT) have been getting smaller in recent years. More specifically, Feingold (1988) found that year of study and effect sizes for sex differences in cognitive abilities are negatively related. He suggested that his data indicate that changes in attitudes toward women over the past several decades have reduced sex differences in cognitive abilities.

However, Feingold's (1988) analysis does not allow clear conclusions to be drawn concerning the relation between year of study and magnitude of sex differences in spatial abilities, because there was only one spatial test included in his analysis (the Space Relations subtest of the DAT). The analysis performed by Feingold thus allowed estimation of the relation between year of study and spatial performance from only a limited perspective. An examination of several spatial tasks in the present study allows a better assessment of Feingold's findings as far as spatial performance is concerned. Furthermore, Feingold (1988) showed that when sex differences are robust (such as in the SAT Quantitative), they remain unchanged across time. It is plausible that the same pattern exists for spatial abilities. Therefore, the present analysis also examines the changes in the magnitude of sex differences in different measures of spatial ability as a function of the year in which the study was conducted. If our analysis is correct, there should be no temporal change in those measures of spatial ability for which sex differences are robust, such as tests measuring mental rotation abilities.

In the present study we use meta-analysis to determine the existence of sex differences in different types of spatial tasks. The meta-analysis includes the results of published studies investigating sex-related differences in spatial abilities, including those analyzed by Maccoby and Jacklin (1974) and Linn and Petersen (1985), and thus provides an extension of Linn and Petersen's findings. The two working hypotheses of the present study are as follows: (a) Sex differences are robust only in tests with an important mental rotation component or with a kinesthetic–gravitational component (as defined by Linn & Petersen, 1985), and (b) where significant sex differences are observed, the magnitude of the difference is not correlated with the year of study.

Definitional Problems

Caplan et al. (1985) suggested that some inconsistencies in the study of sex differences in spatial abilities are caused by the absence of clear definitions of spatial skills. To define spatial abilities, one must initially determine whether spatial ability is a unitary concept or involves a number of diverse components. Unfortunately, there is little agreement among authors as to how spatial abilities should be classified. At first glance, it seems that this lack of agreement could be avoided by the use of a factor analytic approach in order to group and define spatial abilities. However, even this approach does not necessarily produce converging definitions in the literature. For example, Burnett, Lane, and Dratt (1979) defined only a single spatial visualization factor: the ability to manipulate visual images mentally. On the other hand, Very (1967) suggested three distinct spatial factors: spatial ability, spatial visualization I, and spatial visualization II. Very (1967) defined these factors only by their test loadings and provided no verbally formulated definitions. This disagreement about definitions of spatial abilities is further illustrated in Petersen's (1976) treatment. She defined spatial ability as a broad category of skills: the ability to manipulate images visually with a minimum of verbal mediation. The problem with this definition is that its scope is too large to provide a classification of spatial tasks into meaningful groupings.

The lack of a universally accepted definition of spatial ability may be due to the large variety of tests used in the psychometric studies or to the lack of replicability of the factor structures found when several tests are used. Carpenter and Just (1986) attempted to avoid these problems by using an information-processing approach to the study of spatial ability. They defined spatial ability as the ability to generate "a mental representation of a two or three dimensional structure and then assessing its properties or performing a transformation of the representation" (Carpenter & Just, 1986, p. 221). Such an information-processing approach is more concerned with distinguishing the processes underlying different spatial tasks than with the classification of the tasks. Thus, Carpenter and Just's framework leads to a classification of tests in terms of their component processes. This procedure still does not allow a consensus to be reached as to the classification of the common psychometric tests of spatial ability, because it is possible for a study participant to use different processing strategies to solve different items on the same test (Barratt, 1953). Therefore, a classification of spatial tests based on the underlying processes could lead to classifying the same test in several different categories. This approach does not represent a practical way to distinguish among spatial tasks.

Linn and Petersen (1985) provided definitions based on an empirical classification of spatial tests stemming from the meta-analysis they performed. As mentioned earlier, they proposed three categories of spatial tasks: spatial perception, mental rotation, and spatial visualization. Any test in which the task corresponds to the definition of one of these groupings can be classified in one of the three categories these authors defined. The assumption underlying the use of Linn and Petersen's (1985) system of classification is that spatial ability is not a unitary concept, but only a general label used to describe a grouping of several different types of ability.

From the perspective that spatial ability is not a unitary concept, an extreme way to define spatial ability would be to consider that each distinct spatial test provides its own operational definition. This approach has the advantage of taking into account the differences among the tasks used to measure spatial skills. However, in doing so, it fails to recognize the similarities among different tests, and as a result produces more categories of spatial ability than are required, given the overlap in content among the myriad of tests measuring such concepts as mental rotation skills. Although spatial abilities may be defined more parsimoniously with the approach suggested by Linn and Petersen (1985), we used a test-by-test approach in the present analysis for a number of reasons. First, the definitions provided by Linn and Petersen (1985) are vague. In particular, the spatial visualization grouping appears to be a catchall category for tests that do not fit the definition of mental rotation or spatial perception tasks. The definition of spatial visualization presented by these authors provides useful guidelines for classifying spatial tasks. Nevertheless, there are tasks classified in this cluster that involve components related to the other groupings. For example, the Identical Blocks Test (Stafford, 1961) is classified as a spatial visualization task despite its having an important mental rotation component (Linn & Petersen, 1985). Such a problem emphasizes the somewhat arbitrary nature of Linn and Petersen's (1985) classification.

Another problem concerns the criterion (or criteria) used by Linn and Petersen (1985) to decide that effect sizes are close to homogeneity. In Table 1 (p. 1,486) of their 1985 article, Linn and Petersen reported that 4 out of 8 (50%) of the clusters were "close to homogeneity." Because the test of homogeneity is similar to the chi-square test, for a given significance level, the effect is either significant (heterogeneous) or nonsignificant (homogeneous). How did Linn and Petersen (1985) decide that a test was "close to being nonsignificant"? They did not discuss the criteria they used to make such a decision. This suggests that further partitioning was required in their analysis, because homogeneity of effect sizes was not obtained. However, Linn and Petersen (1985) argued that "[as] a wide range of meta-analyses reveal, homogeneity is infrequently achieved" (p. 1,481). Thus, it is justifiable to classify some clusters of effect sizes as close to homogeneity given that statistical homogeneity is rarely achieved. Nevertheless, the fact remains that Linn and Petersen (1985) did not indicate what criteria they used to say that a cluster was close to homogeneity. In the present analysis, groups of effect sizes are classified as close to homogeneity when the obtained chi-square is significant at the .05 level but not at the .01 level. Even though this range is arbitrary, it sets clear limits to what can be considered satisfactorily homogeneous.

Although the categorization of tests focuses on similarities among them, it also neglects potentially critical differences among them. Thus, this approach oversimplifies the underlying pattern of correlations between the tests. Our claims in the present analysis are that spatial ability is not a unitary concept and that the classification proposed by Linn and Petersen (1985) oversimplifies the relation among measures of spatial performance. In the present analysis we attempt to correct this problem by examining the magnitude of sex differences separately in various measures of spatial skills and by grouping these tests on the basis of the factors that affect the magnitude of sex differences. This should result in an equally parsimonious but more practical approach than that suggested by Linn and Petersen (1985), because it takes into account the factors affecting the magnitude of sex differences along with the actual size of these differences.

Meta-Analytic Procedure

Selection Criteria for Inclusion in the Meta-Analysis

The present meta-analysis includes only published studies of sex differences on spatial tasks that have been used with at least five different samples of participants. The selection of a minimum of five studies allows for a meaningful test of the homogeneity of effect sizes for sex differences in the different tasks and for an examination of age effects through regression analysis. We performed a *Psyc-Lit* CD-ROM search for the years 1974 to 1993 to access as many studies of sex differences in spatial performance as possible. Furthermore, we searched recent volumes of journals likely to publish studies on sex differences or spatial abilities. Published studies included in the reviews conducted by Linn and Petersen (1985) and Maccoby and Jacklin (1974) were also selected. Thus, the present meta-analysis provides a nearly exhaustive review of the published literature on sex-related differences in spatial abilities up to 1993. It is worth noting that, in a large number of those studies, the primary purpose was not to test the existence of sex differences in spatial skills; rather, sex of participants was examined only because it was easily observable.

Initially, we identified 310 effect sizes; 286 were entered into the meta-analysis. Studies were eliminated mainly because they reported insufficient information for the computation of effect sizes or because they reported inconsistent information. Studies were also eliminated if they reported results obtained with a spatial ability measure used with fewer than five different samples. Although the choice of five as a criterion is purely arbitrary, it appeared sufficient to provide a reasonable estimate of effect size for each test. Only studies in which the relevant information (N, statistics, age of participants, etc.) was reported were included in the meta-analysis. Other studies were eliminated because they reported composite spatial scores but no separate analysis of test scores. This resulted in the elimination of some classic studies such as those of Petersen (1976) and Waber (1976). The effect sizes entered in the meta-analysis drawn from those studies that were not included in Linn and Petersen's (1985) meta-analysis are presented in Tables 1, 2, and 3. Notice that effect sizes from studies reviewed by Maccoby and Jacklin (1974) can be found in these tables. The effect size estimates are represented by g to indicate that they are biased estimates of effect size (Hedges & Becker, 1986). We transformed them into unbiased estimates of effect size before the analysis, following the approach suggested by Hedges and Becker (1986). The effect sizes analyzed by Linn and Petersen (1985) are presented in detail in Linn and Petersen (1986), and the studies from which they were drawn are indicated by double asterisks in our reference list.[1]

One problem associated with the inclusion of only published studies in a meta-analysis is the "file drawer problem" (see Rosenthal, 1979)—the possibility that published studies are a biased sample of the studies that are actually carried out because it is presumed that only experiments with significant results are published. Because of this problem, the exclusive use of published studies in a meta-analysis may result in an overestimation of the effects under study. The conventional way to cope with the file drawer problem is to calculate the number of studies averaging zero effect size that would be necessary to offset the significance of the findings at the .05 level (Rosenthal, 1980). This value is called the fail-safe number, and a computational formula for this statistic is provided by Rosenthal (1980). No clear-cut guidelines are available for the determination of what is an unlikely number of unpublished or unretrieved studies, although it is evident that the larger the fail-safe number is, the less likely the file drawer problem becomes as a plausible rival hypothesis to significant results in a given meta-analysis. Rosenthal (1980) suggested that a fail-safe number reaching a value of $5k + 10$ (k being the number of sampled studies) indicates combined results resistant to the file drawer problem. Thus, given that the present meta-analysis includes only published studies, we used this criterion to assess the resistance of the meta-analytic results to the file drawer hypothesis.

Analysis Procedure

Cohen's d was used as the measure of effect size (Cohen, 1977). This index represents the standardized difference between the mean of the groups under study (females and males, in the present study). We computed effect sizes using the formula presented by Cohen (1977) when means and standard deviations were available or using the formulae presented by Wolf (1986) when only a t, χ^2, p, or F statistic was available. In the analysis we followed the procedure presented by Hedges and Becker (1986), who developed meta-analytic techniques designed for the assessment of cognitive sex differences and for the evaluation of the homogeneity of the effect sizes. Homogeneity of the effect sizes allows for the conclusion that the effect sizes in a specific sample of studies were drawn from the same population. In other words, homogeneity of the effects sizes indicates that the studies included in a specific meta-analysis can be considered replications of each other and that a pooled estimate of effect size provides a valid summary of the results from the sample of studies. However, when heterogeneity is detected, it is likely that the pooled estimate is not representative of the state of affairs in a sample.

In general, in the present meta-analysis we followed a hierarchical approach similar to that of Linn and Petersen (1985).[2] Thus, we first conducted an overall analysis examining the magnitude and the homogeneity of sex differences among the 286 effect sizes selected for the analysis. Following this procedure, we partitioned the effect sizes into the categories of spatial ability defined by Linn and Petersen (1985) and by age of the participants. The present approach differed from Linn and Petersen's (1985) analysis in that when homogeneity was not achieved with those groupings, studies were further partitioned on the basis of the specific tests used. Finally, in tests in which homogeneity was not achieved, a partition based on age of participants or on one of a number of procedural variables (scoring procedure, individual vs. group testing, sex of the experimenter, selectivity of the sample; see Hedges & Olkin, 1985) was performed. Because there is no systematic way to decide which variables are important (Wolf, 1986), these partitions were based on an examination of the best way to reduce heterogeneity.

Meta-Analytic Results

The overall analysis of 286 studies revealed a mean weighted d of 0.37 ($z = 2.61$, $p < .01$), which demonstrates that sex differences in spatial abilities favoring males are highly signifi-

---

[1] Studies selection requirements in the present meta-analysis resulted in the elimination of a number of studies included in the Maccoby and Jacklin (1974) review and the Linn and Petersen (1985) analysis. Specifically, 7 of the 38 studies reviewed by Linn and Petersen (1985) were not included in the present analysis. Five of these were not published, one reported results obtained with tests used in less than five studies, and one did not have a separate analysis for each test used. Three of the eliminated studies presented results relevant to the spatial perception category, three studies were relevant to the spatial visualization grouping, and one study applied to these two groupings. None of the mental rotation studies were eliminated. From the Maccoby and Jacklin (1974) review, 22 of the 72 studies were eliminated. Five of these were unpublished studies. The remaining 17 studies were eliminated because they used tests that had been used in fewer than 5 studies. It is worth noting that several of these tests had a questionable spatial content. Furthermore, most of them did not fit the definitions proposed by Linn and Petersen (1985) or any of the definitions given in the Definitional Problems section of the present article. Their elimination is thus legitimate and should not affect an estimation of the magnitude of sex differences in spatial performance.

[2] In the Linn and Petersen (1985) analysis, the sampling procedure resulted in the violation of the assumption that the effect sizes are independent (see Hedges & Becker, 1986). Because the present analysis examines the same studies as Linn and Petersen (1985), this assumption is also violated in the present study when the analysis is based on the classification proposed by these authors. However, the analysis based on a test-by-test partition does not violate this assumption. Results obtained in this analysis are thus valid. Furthermore, Linn and Petersen (1985) used a modified version of the Hedges and Becker (1986) procedure that results in a more conservative estimation of significance levels. This approach was also followed in the present analysis to make our results comparable to Linn and Petersen's.

Table 1
Studies on Sex Differences in Mental Rotation

Study	Difference in favor of	Test	*N* Males	*N* Females	Statistic	Age group (years)	*g*
Allen (1974)	Males	CRT	46	47	$p < .005$	19–25	0.602
Beatty & Duncan (1990)	Males	MRT	59	61	$F = 48.08$	19–25	1.277
Brawn & Kimball (1988)	Males	MRT	100	169	$t = 7.00$	19–25	0.857
Bryden, George, & Inch (1990)	Males	GMR	48	48	$F = 4.68$	19–25	0.446
Caldwell & Hall (1970)	None	GMR	73	71	$p = .5$	4–8	0.113
Casey, Colon, & Goris (1992)	Males	MRT	25	46	$F = 18.51$	13–18	1.036
Corballis & Sergent (1989)	None	GMR	45	88	$p = .5$	18–45	0.118
Cronin (1967)	Males	GMR	108	108	$F = 5.12$	5–6	0.309
Freedman & Rovegno (1981)	Males	MRT	40	40	$F = 36.9$	19–25	1.295
Gladue, Beatty, Larson, & Staton (1990)	Males	MRT	32	31	$F = 8.24$	27	0.741
Goldstein et al. (1990)	Males	MRT	23	23	$F = 5.88$	19–25	0.731
Hakstian & Cattell (1975)	Males	GMR	138	142	$p < .01$	15–19	0.626
Harshman, Hampson, & Berenbaum (1983)	Males	CRT	215	215	$p = .001$	19–25	0.290
Herzberg & Lepkin (1954)	Males	PMA–SR	101	54	$t = 2.63$	16–18	0.425
Hobson (1947)	Males	PMA–SR	720	716	$t = 5.16$	15	0.469
Jahoda (1979)	None	GMR	72	72	$F = 1.0$	7–14	0.113
Johnson & Harley (1980)	Males	GMR	60	60	$F = 5.84$	19–25	0.445
Jones & Anuza (1982)	None	GMR	30	30	$p = .5$	19–25	0.177
Kaess (1971)	None	GMR	27	27	$p = .5$	6, 8, 10	0.187
Kail, Carter, & Pellegrino (1979)	Males	GMR	51	53	$F = 6.20$	19–25	0.493
Kail, Stevenson, & Black (1984)	Males	PMA–SR	36	32	$F = 14.87$	19–25	0.949
Kerns & Berenbaum (1991)	Males	GMR	60	63	$F = 2.73$	19–25	0.493
Keyes (1983)	Males	PMA–SR	48	48	$F = 4.58$	13–16	0.441
Mann, Sasanuma, Sakuma, & Masaki (1990)	Males	MRT	151	142	$F = 44.55$	13–15	0.783
McGilligan & Barclay (1974)	None	PMA–SR	20	20	$p = .5$	15	0.221
McGlone & Davidson (1973)	Males	PMA–SR	58	58	$\chi^2 = 11.00$	16–25	0.616
Meyer & Bendig (1961)	None	PMA–SR	49	51	$t = 1.06$	16	0.241
	None	PMA–SR	49	51	$t = 1.18$	13	0.238
Olson & Eliot (1986)	None	CRT	45	53	$t = .47$	19–25	0.138
Olson, Eliot, & Hardy (1988)	None	CRT	45	53	$p = .5$	18–22	0.138
Ozer (1987)	Males	MRT	49	51	$t = 3.47$	18	0.701
Pezaris & Casey (1991)	None	MRT	78	102	$F = 1.94$	13–16	0.209
Sanders & Soares (1986)	Males	CRT	80	194	$F = 6.13$	17–22	0.300
	Males	MRT	80	194	$F = 37.50$	17–22	0.743
Signorella, Jamison, & Krupa (1989)	Males	CRT	354	278	$p = .04$	19	0.159
Stericker & LeVesconte (1982)	Males	MRT	38	45	$t = 2.60$	19–25	0.551
	Males	PMA–SR	38	45	$t = 2.41$	19–25	0.511
Tapley & Bryden (1977)	Males	GMR	20	20	$F = 6.00$	19–25	0.790
Tracy (1990)	Males	GMR	143	139	$F = 12.60$	10	0.424
Vandenberg & Kuse (1978)	Males	MRT	115	197	$t = 6.04$	14–60	0.686
Vandenberg, Kuse, & Vogler (1985)	None	MRT	27	29	$t = 1.41$	19–25	0.384
Very (1967)	Males	CRT	193	162	$t = 5.82$	18–21	0.620
Voyer & Bryden (1990)	Males	MRT	65	65	$F = 5.21$	19–25	0.404
Waber, Carlson, & Mann (1982)	None	GMR	30	30	$p = .5$	12	0.179
	None	GMR	30	30	$p = .5$	14	0.179
	None	PMA–SR	30	30	$p = .5$	12	0.179
	None	PMA–SR	30	30	$p = .5$	14	0.179
Wilson et al. (1975)	Males	MRT	1,457	1,521	$F = 334.54$	14–60	0.640
Yen (1975)	Males	MRT	1,343	1,343	$p = .001$	14–17	0.116
Yeo & Cohen (1983)	Males	MRT	62	62	$F = 17.28$	19–25	0.753

Note. CRT = Cards Rotation Test; GMR = generic mental rotation task; MRT = Mental Rotations Test; PMA–SR = Primary Mental Abilities—Spatial Relations.

cant. The fail-safe analysis indicated that 178,205 studies with nonsignificant results would be needed to offset the significance of the results at the .05 level. However, as should be expected from Linn and Petersen's (1985) findings, the effect sizes were not homogeneous, $\chi^2(285, N = 286) = 1{,}370.49, p < .001$. This suggests that the studies included in the present analysis are not all drawn from the same population and that the pooled estimate of effect size does not provide a representative summary of the sample of effect sizes. Thus, whereas sex differences in spatial tasks are significant, they are also heterogeneous. Partitioning of the effect sizes into homogeneous clusters is thus indicated.

Table 2
Studies on Sex Differences in Spatial Perception

Study	Difference in favor of	Test	N: Males	N: Females	Statistic	Age group (years)	*g*
Allen & Hogeland (1978)	Males	RFT	50	50	$t = 2.70$	18–45	0.545
Beatty & Duncan (1990)	Males	WLT	59	61	$F = 8.63$	19–25	0.541
Bogo, Winget, & Gleser (1970)	Males	RFT	45	52	$t = 3.69$	18–21	0.757
Cohen (1978)	None	WLT	50	50	$p = .5$	4, 5, 7	0.136
Fiebert (1967)	Males	RFT	45	45	$F = 4.65$	12, 15, 18	0.460
Gladue, Beatty, Larson, & Staton (1990)	None	WLT	32	32	$F < 1.0$	27	0.172
Golbeck (1986)	Males	WLT	32	32	$F = 10.42$	18–24	0.820
Gross (1959)	Males	RFT	71	69	$p < .01$	17–25	0.506
Gruen (1955)	Males	RFT	30	30	$t = 2.89$	17–40	0.759
Kalichman (1986)	Males	WLT	97	97	$\chi^2 = 21.3$	20	0.666
Kalichman (1989)	Males	WLT	125	125	$\chi^2 = 26.6$	19–25	0.652
Kato (1965)	Males	RFT	20	20	$p < .05$	18–21	0.551
Kenyon (1984)	None	WLT	15	15	$p = .5$	19–25	0.255
Keogh & Ryan (1971)	Males	RFT	22	22	$F = 9.04$	7	0.928
Liben (1991)	Males	WLT	50	54	$F = 6.39$	19	0.501
Liben & Golbeck (1984)	Males	WLT	80	80	$F = 23.00$	19–21	0.763
Liben & Golbeck (1986)	Males	WLT	80	80	$\chi^2 = 6.67$	19–21	0.408
Maxwell, Croake, & Biddle (1975)	Males	WLT	18	17	$\chi^2 = 17.44$	8–9	1.412
McGillicuddy-DeLisi, DeLisi, & Youniss (1978)	None	WLT	30	30	$p = .5$	6, 8, 10	0.177
	Males	WLT	10	10	$t = 2.33$	19–25	1.098
McGilligan & Barclay (1974)	None	RFT	20	20	$t = 1.91$	15	0.620
Meehan & Overton (1986)	Males	WLT	36	61	$F = 9.30$	19–25	0.516
Morell (1976)	Males	RFT	57	57	$F = 7.48$	11–18	0.517
Morf, Kavanaugh, & McConville (1971)	None	RFT	41	41	$p = .5$	18–21	0.151
Morf & Howitt (1970)	None	RFT	22	22	$p = .5$	18–21	0.219
Morris (1971)	Males	WLT	64	64	$F = 6.26$	17–24	0.446
Mower-Popeil & DeLisi (1984)	Males	WLT	39	86	$t = 2.57$	16–30	0.463
Myer & Hensley (1984)	Males	WLT	41	44	$\chi^2 = 5.11$	19–21	0.490
Okonji (1969)	None	RFT	31	34	$p = .5$	19–50	−0.002
Olson & Eliot (1986)	None	WLT	45	53	$t = -1.07$	19–25	0.218
Olson, Eliot, & Hardy (1988)	None	WLT	45	53	$p = .5$	18–22	0.138
Oltman (1968)	None	RFT	80	83	$p = .5$	18–21	0.169
Pennings (1991)	None	WLT	24	24	$F = 2.31$	7–8	0.448
Rebelsky (1964)	Males	WLT	59	69	$\chi^2 = 5.40$	19–25	0.411
Robert (1990)	Males	WLT	59	274	$\chi^2 = 16.16$	18–46	0.441
Robert & Morin (1993)	Males	WLT	69	126	$F = 9.85$	19–21	0.452
Robert & Ohlmann (1991)	Males	WLT	140	140	$F = 13.74$	18–30	0.445
Robert & Tanguay (1990)	None	RFT	17	15	$t = 1.11$	50	0.404
	None	RFT	13	15	$t = 0.93$	65	0.363
	Males	RFT	15	15	$t = 3.43$	75	1.293
Ruble & Nakamura (1972)	None	RFT	28	28	$p = .5$	7–10	−0.183
Saarni (1973)	Males	RFT	32	32	$F = 5.67$	10–15	0.605
Scholan & Smith (1990)	Males	RFT	10	10	$F = 10.57$	18–20	1.533
Schwartz & Karp (1967)	Males	RFT	23	23	$F = 7.68$	17	0.899
	Males	RFT	20	20	$F = 4.87$	30–39	0.716
	None	RFT	17	17	$F = 2.08$	58–82	−0.468
Sherman (1974)	None	RFT	25	25	$t = 1.37$	19–25	0.395
Signorella, Jamison, & Krupa (1989)	Males	WLT	354	278	$p = .0247$	19	0.183
Silverman, Buchsbaum, & Stierlin (1973)	Males	RFT	7	8	$t = 2.18$	13–20	1.210
	Males	RFT	15	15	$p < .02$	18–22	0.945
Spradlin & Hensley (1979)	Males	WLT	60	60	$F = 30.98$	19–21	1.025
Thompson, Harris, & Mann (1981)	Males	WLT	38	98	$t = 2.74$	17–33	0.473
Treagust (1980)	Males	WLT	54	54	$p < .05$	14–17	0.389
Vaught (1965)	Males	RFT	90	90	$F = 10.39$	18–21	0.483
Voyer & Bryden (1993)	Males	RFT	24	24	$F = 12.47$	18–25	1.041
Willemsen (1974)	None	WLT	24	24	$p = .5$	6, 8	0.199
Willemsen & Reynolds (1973)	Males	WLT	30	30	$p = .01$	19–25	0.698
Witkin, Goodenough, & Karp (1967)	Males	RFT	255	260	$F = 41.17$	8, 10–13, 15	0.567

Note. RFT = rod-and-frame test; WLT = Water Level Test.

Table 3
Studies on Sex Differences in Spatial Visualization

Study	Difference in favor of	Test	N Males	N Females	Statistic	Age group (years)	g
Allen (1974)	Males	PFB	46	47	$p > .05$	19–25	0.141
Andrieux (1955)	Males	EFT	28	32	$t = 2.35$	13–18	0.617
Balistreri & Busch-Rossnagel (1989)	Males	EFT	50	50	$F = 5.93$	19–25	0.492
	Males	EFT	14	14	$t = 2.13$	19–25	0.835
Bieri (1960)	None	EFT	30	30	$p = .09$	18–21	0.177
Bieri, Bradburn, & Galinksy (1958)	Males	EFT	50	62	$t = 3.25$	18–21	0.625
Bergan, MacManis, & Melchert (1971)	Males	BD	24	24	$F = 6.32$	9	0.741
Bernard, Boyle, & Jackling (1990)	Males	EFT	140	180	$F = 5.83$	15–19	0.270
Berry (1966)	None	EFT	61	61	$p = .5$	10–adult	0.123
Bigelow (1971)	None	CEFT	80	80	$F = .009$	5–10	−0.107
Birkett (1980)	None	PFB	54	71	$p = .5$	16–42	0.122
Blum, Fosshage, & Jarvik (1972)	None	BD	27	27	$p = .5$	64, 84	−0.187
Brainerd & Huevel (1974)	None	DAT-SR	60	60	$p = .5$	5–6	0.124
Burnett, Lane, & Dratt (1982)	Males	IBT	183	81	$p = .01$	19–25	0.319
Bush & Coward (1974)	Males	EFT	20	20	$F = 7.38$	18	0.881
Coates (1972)	Females	PEFT	123	124	$p < .05$	3–5	−0.250
Cochran & Wheatley (1988)	None	DAT-SR	16	49	$F = 3.25$	20	0.454
Connor, Serbin, & Schackman (1977)	None	EFT	66	67	$p < .10$	6–10	0.118
Corah (1965)	None	CEFT	30	30	$p = .5$	8–11	−0.028
	Males	EFT	60	60	$t = 3.99$	adults	0.735
Corley, DeFries, Kuse, & Vandenberg (1980)	Males	IBT	871	872	$p < .001$	19–25	0.144
Crandall & Lacey (1972)	None	EFT	28	22	$p = .5$	6–12	0.092
Crandall & Sinkeldam (1964)	None	EFT	28	22	$p = .5$	6–12	0.195
DeLucca, Burright, & Donovick (1990)	Males	BD	40	40	$F = 3.39$	19–25	0.417
Embretson (1987)	Males	DAT-SR	43	29	$F = 5.62$	19–25	0.567
Fiebert (1967)	Males	CEFT	45	45	$F = 6.41$	12, 15, 18	0.540
Fralley, Eliot, & Dayton (1978)	Males	IBT	44	49	$F = 8.17$	parents	0.590
Gainer (1962)	None	BD	100	100	$t = .65$	$M = 8.87$	−0.096
Goldstein & Chance (1965)	Males	EFT	13	13	$t = 2.40$	18–21	0.980
Goodenough & Eagle (1963)	None	PEFT	48	48	$p = .5$	5, 8	−0.139
Gruen (1955)	None	EFT	30	30	$t = 1.44$	17–40	0.378
Hakstian & Cattell (1975)	None	EFT	138	142	$p = .5$	15–19	0.081
Harshman, Hampson, & Berenbaum (1983)	Males	DAT-SR	98	97	$p = .001$	19–25	0.432
Hult & Brous (1986)	Males	DAT-SR	133	56	$F = 3.87$	19–21	0.288
Immergluck & Mearini (1969)	Females	EFT	60	60	$\chi^2 = 5.01$	9	−0.409
Johnson, Flinn, & Tyer (1979)	None	EFT	41	41	$t = .95$	19–25	0.151
Kagan, Rosman, Day, Albert, & Phillips (1964)	None	HFT	90	90	$p = .5$	7	0.101
Keogh & Ryan (1971)	None	CEFT	22	22	$p = .5$	7	0.213
Kershner (1971)	None	SR	80	80	$t = 1.25$	6	0.199
Levinson (1960)	None	BD	29	29	$t = .09$	5–6	0.017
Lusk & Wright (1981)	Males	EFT	257	152	$p = .031$	19–25	0.167
MacArthur (1967)	None	EFT	98	69	$p = .5$	9–15	0.105
Marino & McKeever (1989)	Males	IB	40	34	$F = 4.50$	19–25	0.500
Mayes (1982)	None	PFB	40	40	$F = 1.39$	19–25	0.267
McKeever (1986)	Males	IB	147	212	$F = 13.18$	19–25	0.384
Miller & Santoni (1986)	Males	DAT-SR	31	28	$F = 4.5$	19–25	0.562
Mower-Popeil & DeLisi (1984)	Males	PFB	39	86	$t = 2.63$	16–30	0.474
Mumbauer & Miller (1970)	None	CEFT	32	32	$p = .5$	4–5	0.171
Myer & Hensley (1984)	None	EFT	41	44	$\chi^2 = 1.45$	19–21	0.264
Nash (1975)	None	DAT-SR	45	59	$p = .5$	12	0.127
	Males	DAT-SR	38	56	$t = 2.27$	16	0.471
Newcombe, Bandura, & Taylor (1983)	Males	DAT-SR	22	23	$t = 1.85$	19–25	0.564
Norman (1953)	Females	BD	85	68	$t = 1.36$	15–29	0.221
Okonji (1969)	None	EFT	14	11	$p = .5$	21–27	0.269
Olson & Eliot (1986)	None	PFB	45	53	$t = 1.24$	19–25	0.253
Olson, Eliot, & Hardy (1988)	None	EFT	45	53	$p = .5$	18–22	0.138
Pande (1970)	Males	EFT	70	70	$t = 5.25$	18–24	0.894
Parlee & Rajagopal (1974)	Males	EFT	47	48	$t = 4.25$	19–25	0.881
Pennings (1991)	None	EFT	24	24	$F = .06$	7–8	−0.072
Persaud (1991)	None	PFB	54	84	$p = .5$	20–56	0.116
Reppucci (1971)	None	PEFT	24	24	$p = .5$	2	0.199
Samuel (1983)	Males	BD	416	416	$F = 8.89$	12–16	0.207

(*table continues*)

Table 3 (*continued*)

Study	Difference in favor of	Test	N Males	N Females	Statistic	Age group (years)	g
Sarason & Minard (1962)	Males	BD	48	48	$p < .01$	18–21	0.543
Schubert & Cropley (1972)	None	BD	96	90	$p = .5$	6–15	0.099
Schwartz & Karp (1967)	Males	EFT	23	23	$F = 5.53$	17	0.763
	Males	EFT	20	20	$F = 8.43$	30–39	0.942
	None	EFT	17	17	$F = .08$	58–82	−0.238
Sherman & Fennema (1978)	None	DAT-SR	161	152	$t = 1.29$	15	0.146
Sherman (1974)	Males	DAT-SR	25	25	$t = 2.46$	19–25	0.710
Stafford (1961)	Males	IBT	58	70	NR	13–17	0.600
	Males	IBT	52	52	NR	adults	1.068
Stericker & LeVesconte (1982)	Males	DAT-SR	38	45	$t = 4.03$	19–25	0.854
Stouwie, Hetherington, & Parke (1970)	None	EFT	75	81	$t = .28$	8–9	0.109
Stuart, Breslow, Brechner, Ilyus, & Wolpoff (1965)	None	EFT	37	75	$t = 0.54$	18–20	−0.129
Tapley & Bryden (1977)	None	DAT-SR	20	20	$F = 0.61$	19–25	0.219
Van Blerkom (1987)	None	EFT	24	24	$F < 1.0$	19–25	0.199
Weinberg & Rabinowitz (1970)	None	BD	28	20	$p = .5$	12–19	0.199
Willoughby (1967)	None	HFT	39	37	$t = 0.93$	18–21	0.157
Witkin, Goodenough, & Karp (1967)	Males	BD	24	24	$F = 6.32$	9	0.741
Witkin (1950)	Males	EFT	51	51	$t = 2.7$	19–21	0.54
Yen (1975)	Males	PFB	1343	1343	$p = .001$	14–17	0.116
Yeo & Cohen (1983)	None	HFT	62	62	$F = .99$	19–25	0.122

Note. BD = Block Design subtest of the Wechsler Adult Intelligence Scale or the Wechsler Intelligence Scale for Children; DAT-SR = Differential Aptitude Test—Spatial Relations; CEFT = children's Embedded Figures Test; EFT = Embedded Figures Test; HFT = Hidden Figures Test; IBT = Identical Blocks Test; PEFT = preschool Embedded Figures Test; PFB = Paper Form Board; NR = not reported.

In an initial attempt to achieve homogeneity, we partitioned the effect sizes following Linn and Petersen's (1985) classification of spatial tasks. Their mental rotation grouping included the following tests: the Spatial Relations subtest of the Primary Mental Abilities Test (PMA; Thurstone, 1958) and the Cards Rotation Test (Ekstrom, French, & Harman, 1976), in which participants are required to perform a mental rotation of two-dimensional objects; the Mental Rotations Test (Vandenberg & Kuse, 1978), which is a paper-and-pencil version of the Shepard and Metzler (1971) mental rotation task, using three-dimensional objects; and what we have termed *generic mental rotation tasks,* which include any variant of the Shepard and Metzler (1971) chronometric task, either presented on slides or on a computer screen. The tests included in the spatial perception category were as follows: the rod-and-frame test (Witkin & Asch, 1948), which requires subjects to adjust a rod to the vertical, despite the distracting information provided by the tilted frame; and the Water Level Test (Piaget & Inhelder, 1956), in which participants are required to indicate the orientation of the liquid in a tilted container. Finally, the spatial visualization cluster included the Paper Form Board (Likert & Quasha, 1941), in which participants must decide which of five, two-dimensional line drawings of shapes can be made out of a set of fragmented parts; the DAT Spatial Relations Subtest (Bennett, Seashore, & Wesman, 1947), in which participants are required to indicate what an unfolded shape would look like when folded; the Identical Blocks Test (Stafford, 1961), in which participants must indicate which block among a number of alternatives is the same as a standard, given a variety of cues (letters and numbers on the faces of the blocks); the Block Design subtest of the Wechsler Adult Intelligence Scale, the Wechsler Adult Intelligence Scale—Revised, and the Wechsler Intelligence Scale for Children (Wechsler, 1946, 1949, 1955, 1974, 1981), in which participants must reconstruct a shape using three-dimensional blocks; and Paper Folding (Ekstrom et al., 1976) in which participants are required to indicate which one among four unfolded pieces of paper is the same as a folded model in which holes were punched. The spatial visualization grouping also incorporated the various adult and children's versions of the Embedded Figures Test (Witkin, 1950), including the Hidden Figures Test (Ekstrom et al., 1976), in which participants must find a simple figure embedded within a complex pattern.

As can be seen from Table 4, when the tests are classified in this manner, the basic findings of Linn and Petersen (1985) are replicated. Specifically, the largest effect size is found for the mental rotation category (mean weighted $d = 0.56$, $p < .05$; fail-safe no. = 25,304), whereas the smallest effect size is found for the spatial visualization category (mean weighted $d = 0.19$, *ns*). As in Linn and Petersen's review, the spatial perception category is intermediate between these two extremes (mean weighted $d = 0.44$, $p < .05$; fail-safe no. = 16,743). It appears that measures of mental rotation ability show the most reliable sex differences, whereas measures of spatial visualization do not show consistent sex differences. The weighted estimate of effect size can be directly interpreted in standard deviation units. This means that on the average, males outperform females by about 0.6 standard deviation units in mental rotation, 0.4 standard deviation units in spatial perception, and 0.2 standard deviation units in spatial visualization.

Categorizing the tests according to the Linn and Petersen (1985) classification resulted in a significant reduction in heterogeneity, $\chi^2(2, N = 286) = 410.09, p < .001$. However, further

Table 4
Effect Sizes for Sex Differences in Spatial Abilities as a Function of Category of Test and Age of the Study Participants

Category of tests	N	Weighted estimator of effect size: Present study	Linn & Petersen (1985)	Test of significance for effect size (Z)	Homogeneity statistic (χ^2)
Mental rotation					
All ages	78	0.56	0.73	4.63*	560.19*
Under 13 years	13	0.33		2.00*	10.72[a]
13–18 years	23	0.45		4.21*	200.62*
Over 18 years	42	0.66		5.55*	263.70*
Spatial perception					
All ages	92	0.44	0.44	2.25*	158.74*
Under 13 years	21	0.33	0.37	1.73	36.31[b]
13–18 years	18	0.43	0.37	2.18*	16.22[a]
Over 18 years	53	0.48	0.64	2.48*	97.16*
Spatial visualization					
All ages	116	0.19	0.13	1.43	241.47*
Under 13 years	40	0.02		0.12	53.01[b]
13–18 years	20	0.18		1.52	41.58*
Over 18 years	56	0.23		2.00*	112.48*

[a] Homogeneity achieved. [b] Close to homogeneity.
* $p < .05$.

examination of Table 4 indicates that significant heterogeneity of effect size remained in all of the categories. Once again, this indicates that the effect sizes are not drawn from the same population and that further partitioning is warranted. Following the approach used by Linn and Petersen (1985), we next partitioned the effect sizes in terms of the age of the participants. The classification of participants into three age groups (under 13 years of age, between 13 and 18 years inclusive, and above 18 years of age) was the same as that used by Linn and Petersen (1985). It is worth noting that such a classification produces differences in the composition of the groups. For example, the "Over 18" group is primarily composed of undergraduate samples that are presumably selected from a more limited range of intellectual abilities than are elementary school samples ("Under 13") or high school samples (between 13 and 18 inclusive). Furthermore, the "Over 18" grouping includes a few samples of individuals older than 30 years of age or of individuals drawn from the general population. This state of affairs is likely to produce more variability in the older samples.

The results of the partition in terms of age of participants within each of the task categories are also presented in Table 4. It can be seen that the effect sizes increase with the age of the participants sampled in each category. However, effect sizes were homogeneous (or close to homogeneous) in only a few of the subcategories: for subjects under age 13 in the three categories of tests and for subjects between ages 13 and 18 in spatial perception. In those four groupings, the studies can be considered as replications of each other. For each of the three categories, partitioning by age significantly reduced heterogeneity: mental rotation, $\chi^2(2, N = 78) = 85.15, p < .001$; spatial perception, $\chi^2(2, N = 92) = 9.05, p < .02$; and spatial visualization, $\chi^2(2, N = 116) = 34.40, p < .001$. However, there remains significant heterogeneity in many groupings, which indicates that the pooled estimate of effect size is not necessarily a valid reflection of the effect size obtained in individual studies. This heterogeneity demonstrates that Linn and Petersen's (1985) classification is not fully adequate and justifies the partitioning of the effect sizes in terms of the individual tests.

The result of such a partitioning is shown in Table 5. It can be seen that only 4 of the 12 tests sampled (Embedded Figures Test, DAT Spatial Relations subtest, Block Design, and Paper Folding) fail to show significant sex differences. This partitioning also produces a significant reduction in heterogeneity when compared with the overall analysis, $\chi^2(11, N = 286) = 541.73, p < .001$. However, significant heterogeneity remains in 6 of the tests (Mental Rotations Test, rod-and-frame test, Water Level Test, Embedded Figures Test, Identical Blocks Test, and Block Design). This means that further partitioning was indicated for these tests.

As before, the effect sizes for those tests showing heterogeneity were partitioned by age of participants. This analysis revealed that homogeneity or near homogeneity of effect sizes was achieved for the rod-and-frame test when age was taken into account (see Table 6), with a significant reduction in heterogeneity, $\chi^2(2, N = 30) = 12.61, p < .01$. Whereas heterogeneity was achieved for Block Design in the two older age groupings, effect sizes were not even close to homogeneity for subjects under age 13, and further partitioning by procedural variables did not allow homogeneity to be achieved. Furthermore, age partitioning did not significantly reduce heterogeneity for Block Design, $\chi^2(2, N = 15) = 4.05, p > .05$.

For the remaining tests, we used procedural variables to partition the effect sizes in an attempt to achieve homogeneity. The results of these partitions are found in Tables 7 and 8. Homoge-

Table 5
Summary Statistics for Sex Differences in Various Measures of Spatial Abilities

Test	N	Weighted estimator of effect size	Fail-safe no.	Test of significance for effect size (Z)	Homogeneity statistic (χ^2)
Cards Rotation Test	10	0.31	285	2.84*	15.67[a]
Mental Rotations Test	35	0.67	9,795	6.55*	378.00*
Generic mental rotation task	15	0.37	245	2.04*	17.99[a]
Spatial Relations (PMA)	18	0.44	730	2.82*	21.73[a]
Rod-and-frame test	30	0.48	1,273	2.15*	57.54*
Water Level Test	62	0.42	8,395	2.29*	109.56*
Paper Form Board	7	0.18	307	2.52*	3.82[a]
Embedded Figures Test	59	0.18	—	1.23	124.04*
Spatial Relations (DAT)	22	0.27	—	1.38	34.32[b]
Identical Blocks Test	8	0.27	196	2.47*	28.72*
Block Design	15	0.17	—	0.94	32.79*
Paper Folding	5	0.12	—	1.57	4.58[a]

Note. PMA = Primary Mental Abilities Test; DAT = Differential Aptitude Test.
[a] Homogeneity achieved. [b] Close to homogeneity.
* $p < .05$.

neity was achieved on the Mental Rotations Test, along with a significant reduction in heterogeneity, $\chi^2(2, n = 35) = 329.41$, $p < .001$, when effect sizes were partitioned by scoring procedure. Because this test requires two responses for each of the 20 items, it can be scored out of 40 if credit is given for every correctly marked choice or out of 20 if the entire item must be answered correctly. The partitioning presented in Table 7 represents this distinction. The Other category represents those studies in which a modified version of the test was used that produced a maximum score other than 20 or 40.

Results from this partition showed that the Mental Rotations Test produces significant sex differences regardless of how it is scored. However, the magnitude of sex differences is largest when the test is scored out of 20 and smallest when it is administered in an unconventional format. The sex difference is significantly larger when the test is scored out of 20 than when it is scored out of 40, $z^2 = 24.05$, $p < .001$. Because scoring the test out of 40 gives more credit for guessing, the reduced sex difference when it is scored this way may be a consequence of women's guessing more often than men (Voyer & Chisholm, 1992).

On the Identical Blocks Test, homogeneity and a significant reduction in heterogeneity, $\chi^2(1, N = 8) = 18.32$, $p < .001$, were achieved when effect sizes were partitioned in terms of individual versus group testing conditions. As can be seen in Table 7, sex differences are found in both testing conditions. However, the magnitude of the effect is larger when the testing is

Table 6
Summary Statistics for Sex Differences in Tests With Homogeneous Effect Sizes When Partitioned by Age

Test	N	Weighted estimator of effect size	Test of significance for effect size (Z)	Homogeneity statistic (χ^2)
Rod-and-frame test				
Under 13	4	0.46	1.58	10.23[b]
13–18	6	0.58	3.37*	2.46[a]
Over 18	20	0.43	1.80	32.24[b]
Block design				
Under 13	9	0.05	0.21	23.82*[c]
13–18	2	0.21	2.15*	0.01[a]
Over 18	4	0.28	1.56	4.91[a]

[a] Homogeneity achieved. [b] Close to homogeneity. [c] Further partitioning did not achieve homogeneity of effect sizes.
* $p < .05$.

Table 7
Summary Statistics for Sex Differences in Tests With Homogeneous Effect Sizes When Partitioned by an Experimental Procedure Variable

Test	N	Weighted estimator of effect size	Test of significance for effect size (Z)	Homogeneity statistic (χ^2)
Mental Rotations Test				
Scored out of 40	13	0.70	4.73*	19.57[a]
Scored out of 20	19	0.94	8.09*	26.34[a]
Other	3	0.14	2.29*	2.68[a]
Identical Blocks Test				
Individual testing	5	0.54	3.24*	8.47[a]
Group testing	3	0.17	2.22*	1.93[a]
Water Level Test				
Deviation	30	0.44	2.42*	34.55[a]
Criterion: 5°	10	0.62	3.05*	26.47[b]
Criterion: 6°–10°	16	0.58	1.95	16.52[a]
Other	6	0.24	2.18*	1.07[b]

[a] Homogeneity achieved. [b] Close to homogeneity.
* $p < .05$.

Table 8
Summary Statistics for Sex Differences on the Various Versions of the Embedded Figures Test (EFT)

Test	N	Weighted estimator of effect size	Test of significance for effect size (Z)	Homogeneity statistic (χ^2)
Children's EFT	23	0.01	0.01	22.65[a]
Individual EFT	21	0.42	1.88	39.06[b]
Group EFT	15	0.18	1.96*	19.09[a]

[a] Homogeneity achieved. [b] Close to homogeneity.
* $p < .05$.

individual ($z^2 = 4.06$, $p < .05$). The main distinction between these two situations is that in group testing a relative state of anonymity exists in which an individual participant feels less direct pressure from the experimenter and the other participants in the group because she or he is not the only point of attention for the experimenter. In an individual testing situation, pressure is on the participant to perform, because he or she is the only point of attention. Just how this distinction would affect men and women differentially is less clear and warrants investigation.

The Water Level Test also required partitioning by scoring procedure for homogeneity to be achieved (see Table 7). The Water Level Test can be scored either with a measure of the deviation from the horizontal (in degrees) or in terms of an arbitrary criterion that respondents pass or fail. This pass/fail criterion can be set at 5°, between 6° and 10°, or elsewhere. Partitioning by these different scoring criteria significantly reduces heterogeneity, $\chi^2(3, N = 62) = 30.95$, $p < .001$. The sex difference is smallest in those studies in which a poorly defined criterion was used and largest when an objective pass/fail criterion was used.

The Embedded Figures Test category actually comprised several different tests. These included the group Embedded Figures Test (including the Hidden Figures Test), the individual Embedded Figures Test, and the children's Embedded Figures Test (including the preschool Embedded Figures Test). When the effect sizes were partitioned into these three categories, homogeneity was achieved (see Table 8). This partitioning significantly reduced heterogeneity, $\chi^2(2, N = 59) = 43.24$, $p < .001$. As was the case with the Identical Blocks Test, the largest effect size was obtained with individual testing, although this did not reach significance, and a nonsignificant sex difference was found for the children's Embedded Figures Test.

In summary, all of the individual tests can be partitioned into logical categories that yield homogeneous effect sizes. These partitions emphasize the conclusion that sex differences are robust. However, they also point to various problems in the administration and scoring of certain tests that lead to deviant results.

Changes in the Magnitude of Sex Differences With Age

The present study also permits an investigation of the relation between effect size and chronological age, both for the sample of studies as a whole and for the individual tests. This analysis was carried out using weighted regression analysis, as suggested by Hedges and Becker (1986). The weights used in the analysis were the same as those that were used to calculate the weighted estimate of effect size in the primary analysis. Note that this analysis used the actual age of participants as a continuous predictor variable, as opposed to using the classification of participants into one of the three discrete categories as was done in the analysis presented in Table 4. Age of participants was estimated from the mean age reported in many of the individual studies. In other cases, age was calculated from the elementary or high school grade. For example, subjects in Grade 1 usually are 6-year-olds. Finally, we estimated the age of undergraduate student samples by assuming that first-year undergraduates are on average 19 years old. However, given Feingold's (1988) finding that the magnitude of cognitive gender differences is decreasing over time, a possible confounding effect of year of study cannot be ignored. We controlled for this eventuality by partialing out year of publication in the analysis. Finally, we investigated the possibility that the relation between age and effect size might be nonlinear by coding a quadratic and a cubic trend as predictors in the analysis. The results of this analysis are presented in Table 9.

An examination of Table 9 shows that, overall, there is an increase in the magnitude of sex differences with age ($r = 0.263$, $p < .01$), with a significant linear component ($z = 9.74$, $p < .01$) and a nonsignificant quadratic component when the linear component is factored out ($z < 1$, *ns*). This finding corroborates the data presented in Table 4, in which participants below age 13 do not show significant sex differences in any of the categories of spatial tests, participants above age 18 always show sex differences, and those between ages 13 and 18 obtain significant sex differences in the spatial perception and mental rotation groupings.

Table 9
Results of the Weighted Regression Analysis With Age of Subjects as Predictor, Effect Size as Dependent Variable, and Year of Birth Partialed Out

Test	N	R^2	Z	Age range	Quadratic trend (Z)
Overall	286	0.069	9.74*	2–75	<1.00
Cards Rotation Test	10	0.241	−1.95	13–20	<1.00
Mental Rotations Test	35	0.245	9.63*	14–58	−13.19*
Generic Mental Rotation task	15	0.061	1.05	6–25	−1.98*
Spatial Relations (PMA)	18	0.252	2.35*	10–20	1.26
Rod-and-frame test	30	0.024	−1.06	7–65	−1.32
Water Level Test	62	0.012	1.18	5–75	−2.08*
Paper Form Board	7	0.002	0.09	12–30	<1.00
Embedded Figures Test	59	0.012	1.23	2–65	−2.28*
Spatial Relations (DAT)	22	0.221	2.75*	6–20	1.82
Identical Blocks Test	8	0.043	1.11	15–25	2.58*
Block Design	15	0.001	0.21	6–65	−2.55*
Paper Folding	5	0.754	1.86	10–25	<1.00

Note. PMA = Primary Mental Abilities Test; DAT = Differential Aptitude Test.
* $p < .05$.

Table 9 also shows the relation between age and effect size for the individual tests. Three of the 12 tests, the Mental Rotations Test, the PMA Spatial Relations subtest, and the DAT Spatial Relations subtest, showed a significant linear positive relation between age and magnitude of sex differences. Even though the trend analyses showed significant quadratic components for the Mental Rotations Test, the generic mental rotation task, the Water Level Test, the Embedded Figures Test, the Identical Blocks Test, and the Block Design subtest, an examination of the scatterplots relating age to effect size indicated that virtually all of these quadratic trends were the result of either outliers or very restricted age ranges. Although some of the increase in effect size as a function of age may be attributed to the small sex differences obtained in prepubescent children, it is noteworthy that one of the tests showing a significant age effect is the Mental Rotations Test, which was not administered to children younger than 14.

We considered the magnitude of sex differences in individual age groups in an attempt to discover at what age sex differences first appear. However, because some of the tests were not used with subjects younger than 13, an examination of the age of emergence of sex difference in the whole sample of studies would be misleading. For this reason, we conducted an analysis only on tests in which samples of subjects younger than 13 were studied. This analysis showed that the earliest age at which sex differences were consistently reported is 7 years for the rod-and-frame test, 9 years for the Water Level Test, 10 years on the generic mental rotation task and the PMA Spatial Relations subtest, 13 years on the DAT Spatial Relations subtest, and 14 years on the Embedded Figures Test. It appears from this that sex differences in early childhood have not been convincingly established, even though they appear at a relatively young age on the rod-and-frame test and the Water Level Test (but see Kerns & Berenbaum, 1991). It is worth noting that most measures of spatial performance presented in the present study were developed for use with adults. This makes these tasks very difficult for young children, which thus produces floor effects that might mask sex differences at young ages.

Changes in the Magnitude of Sex Differences Over Time

Feingold (1988) claimed that cognitive sex differences are getting smaller because of societal changes in attitudes toward women. We examined the hypothesis that the magnitude of sex differences is negatively correlated with the year of publication in the present data by using the weighted regression analysis approach proposed by Hedges and Becker (1986). Overall, the analysis showed a small positive, rather than negative, relation between the year in which the study was published and the magnitude of the sex difference ($r = .079$, *ns*). However, this particular analysis confounds year of study and the age of the participants. In his study, Feingold isolated the effect of year of publication by classifying participants by age groups. This was made possible because of the systematic sampling used in the standardization data he utilized. In the present analysis, sampling is not systematic, and this is likely to produce cohort effects that are confounded with the effect of year of study on the magnitude of sex differences. For example, 75-year-olds who participated in an experiment in 1992 were not raised in the same social environment as 20-year-olds who participated in an experiment in 1992. If Feingold is correct in stating that the decrease in the magnitude of sex differences in cognitive abilities is due to changes in attitudes in the social environment, it is more important to take into account the environment the participants were raised in than the year in which a study was published. This means that year of birth, not year of study, is the variable that must be taken into account to allow a proper test of Feingold's argument.

Following this rationale, participants' year of birth was computed by subtracting their age from the year in which each study was published. However, even though this procedure allows an examination of the influence of the environment participants were raised in, it does not control for age of participants because there is a strong correlation between age and year of birth ($r = -.736$, $p < .001$). To control for this possible confound, we partialed out age of participants from the analysis. We thus performed a weighted regression analysis with year of birth as the predictor, magnitude of sex differences as the dependent variable, and age of participants partialed out. As in the age analysis, we investigated the possibility that the relation between age and effect size might be nonlinear by coding a quadratic and a cubic trend as predictors in the analysis. Results of this analysis are presented in Table 10. Overall, the weighted regression analysis revealed a nonsignificant negative relation between year of birth and magnitude of sex differences ($z = -1.36$, $p > .05$). This indicates that participants who were born more recently tend to show smaller sex differences in spatial abilities when compared with participants who were born earlier but that this tendency is not significant at the .05 level.

Examining the different tests of spatial ability separately allows for a more detailed investigation of Feingold's (1988) hypothesis. This approach revealed that the Cards Rotation Test, the Water Level Test, the Embedded Figures Test, and the Iden-

Table 10
Results of the Weighted Regression Analysis With Year of Birth as Predictor, Effect Size as Dependent Variable, and Age Partialed Out

Test	N	R^2	Z	Year range
Overall	286	0.001	−1.36	1902–1983
Cards Rotation Test	10	0.920	−3.80*	1947–1968
Mental Rotations Test	35	0.103	6.26*	1920–1977
Generic mental rotation task	15	0.007	0.36	1957–1980
Spatial Relations (PMA)	18	0.001	0.05	1932–1976
Rod-and-frame test	30	0.012	0.77	1902–1973
Water Level Test	62	0.053	−2.41*	1915–1983
Paper Form Board	7	0.124	−0.68	1953–1971
Embedded Figures Test	59	0.043	−2.32*	1902–1983
Spatial Relations (DAT)	22	0.001	0.22	1954–1972
Identical Blocks Test	8	0.341	−3.13*	1936–1969
Block Design	15	0.006	0.43	1907–1971
Paper Folding	5	0.074	0.58	1954–1970

Note. PMA = Primary Mental Abilities Test; DAT = Differential Aptitude Test.
* $p < .05$.

tical Blocks Test all showed a significant negative linear relation between year of birth and magnitude of sex differences. The Paper Form Board demonstrated a negative but nonsignificant effect. Only one test, the Mental Rotations Test, showed a significant positive relation between year of birth and magnitude of sex differences. Finally, the analysis revealed a positive but nonsignificant relation between year of birth and magnitude of the effect sizes on the remaining tests. The nonlinearity was nonsignificant in all of the tests.

Although those tests showing significant negative relations might be viewed as providing limited support for Feingold's (1988) hypothesis, it is noteworthy that Feingold's analysis involved the DAT Spatial Relations subtest, one of the tests that does not show a significant relation between year of study and effect size in the present analysis. Thus, using a somewhat different approach we failed to fully replicate Feingold's (1988) findings.

Discussion

In general, the meta-analysis succeeded in showing how the various tests of spatial ability that we have examined can be partitioned to achieve homogeneity of effect size. A consideration of the steps necessary to accomplish this provides both an indication of how different tests of spatial ability might be grouped and some cautionary evidence concerning administration and scoring procedures. The analysis also permitted us to examine changes in effect size with age and changes that may be occurring over time.

Magnitude and Classification of Sex Differences

A summary of the results of the meta-analysis is presented in Table 11. In this table, the various tests have been grouped by effect size and the clusters suggested by Linn and Petersen (1985), with the partitions necessitated to achieve homogeneity indicated. This table suggests that the classification proposed by Linn and Petersen is at least reasonably successful and also indicates some of the places where caution is required.

The results of the present analysis strongly suggest that sex differences in spatial abilities do exist. The astronomical overall fail-safe number of 178,205 is much larger than the criterion suggested by Rosenthal (1980), that is, $5(286) + 10 = 1{,}440$. This finding makes it obvious that the file drawer problem is not plausible as an alternate explanation. It is unlikely that over 170,000 studies on sex differences in spatial abilities with nonsignificant results are gathering dust in file drawers around the world. Furthermore, the fail-safe numbers associated with the effect size obtained in individual tests with significant effects (see Table 5) also exceed the criterion set by Rosenthal (1980).

As Table 11 shows, the effect sizes vary considerably from test to test, indicating that different measures of spatial ability measure somewhat different processes. When the different effect sizes are grouped according to Linn and Petersen's (1985) categories, it can be seen that their grouping is replicated reasonably well, in that sex differences are most compelling for tests in their mental rotation category, large but less consistent for their spatial perception category, and highly variable and frequently nonsignificant for tests in their spatial visualization category. It is important to note that the majority of studies in the present analysis were not included in the Linn and Petersen (1985) investigation. Furthermore, the present analysis spans nearly 50 years of research, whereas Linn and Petersen (1985) considered studies only between 1974 and 1982. Thus, the present meta-analysis reports further evidence for the existence of sex differences in favor of males in tests that assess mental rotation and spatial perception skills.

It is interesting to note that the only two so-called spatial visualization tests showing significant sex differences (Paper Form Board, Identical Blocks Test) have an important mental rotation component (Linn and Petersen, 1985). This finding suggests that Linn and Petersen's (1985) spatial visualization category is a catchall grouping and that these two tests might better be classified in the mental rotation category. This might also be interpreted as indicating that Linn and Petersen's (1985) claim that sex differences in spatial visualization are not significant was premature, because two tests in this category show such differences. However, one is then faced with the question of why these two spatial visualization tests show significant sex differences when no other tests in this category do. Although both tests have a mental rotation component, so do other spatial visualization tests that do not show significant sex differences (Block Design and Paper Folding, for example). An informal analysis of the processing required in these two tests suggests that they require the ability to transform a spatial problem presented in two dimensions to a solution in three dimensions. In Horan and Rosser's (1984) terms, this type of problem requires dimensionality crossing, a condition that was found by these authors to result in sex differences in favor of males. However, we are still faced with the problem of explaining why other spatial visualization tests that involve dimensionality crossing (Block Design, for example) do not show significant sex differences. It appears that the present analysis cannot provide a satisfactory explanation of the inconsistent findings in the spatial visualization grouping. More empirical work focusing on the task demands of various tests might provide a solution to this problem.

Procedural Issues

For many tests, heterogeneity of effect size was achieved only by partitioning according to some aspect of the way the test was administered or scored. The fact that such procedural factors have a strong influence on effect size indicates that serious attention should be given to the way in which particular tests are administered.

On the Mental Rotations Test, for example, effect sizes were smaller when the test was scored out of 40 than when it was scored out of 20. The original instructions (Vandenberg & Kuse, 1978) call for scoring it out of 20 by requiring that a respondent select both correct alternatives to be given credit, although many researchers apparently feel that the score is more precise when it is scored out of 40. Giving credit for each correct choice permits guessing to have a greater influence. It would be worthwhile to investigate sex differences in the patterns of guessing and omissions on this test.

Table 11
Summary of the Results of the Present Meta-Analysis

Effect size range	Linn & Petersen's (1985) categories: Mental rotation	Spatial perception	Spatial visualization
0.75–1.00	*MRT (scored out of 20)		
0.50–0.74	*MRT (scored out of 40)	*WLT (5°) WLT (6°–10°) *RFT (ages 13–18)	*IBT (individual)
0.40–0.49	*PMA-SR	RFT (under age 13) RFT (over age 18) *WLT (deviation)	EFT (individual)
0.30–0.39	*GMR *CRT		
0.20–0.29		*WLT (other)	BD (over age 18) DAT-SR BD (ages 13–18)
0.10–0.19	*MRT (other)		*EFT (group) PFB *IBT (group) PF
0.00–0.09			BD (under age 13) EFT (child)

Note. CRT = Cards Rotation Test; MRT = Mental Rotations Test; GMR = generic mental rotation task; PMA-SR = Primary Mental Abilities—Spatial Relations; RFT = rod-and-frame test; WLT = Water Level Test; PFB = Paper Form Board; EFT = Embedded Figures Test; DAT-SR = Differential Aptitude Test—Spatial Relations; IBT = Identical Blocks Test; BD = Block Design; PF = Paper Folding.
* Clusters showing sex differences significant at the .05 level.

On two other tests, the Identical Blocks Test and the Embedded Figures Test, larger effect sizes were obtained when the test was administered individually than when it was given in a group. This would suggest that there are meaningful sex differences in the way men and women respond to the differences between these two testing situations. For example, women may be more stressed by the individual testing situation and perform more poorly. Alternatively, men may be less attentive in the group situation and perform poorly in that context. However, inconsistencies in the present analysis suggest that factors extraneous to the testing situation may be involved. For instance, the Block Design subtest is always administered individually, but it revealed no significant sex differences in the present study. This observation indicates that even though the magnitude of sex differences in different tests may be affected by procedural factors, the importance of task content should not be forgotten. Again, this is an issue that warrants further investigation.

For some of the tests, most notably the Mental Rotations Test and the Water Level Test, a grouping we have termed "Other" was needed to achieve homogeneity of effect size and led to the smallest effect sizes. In general, this grouping was used to classify studies in which the scoring deviated in some major way from the conventional method. The fact that such studies are outliers in terms of effect size simply serves to emphasize that one cannot expect to find meaningful effects if one arbitrarily alters the metric.

Emergence of Sex Differences

The effect of age is difficult to evaluate in a large number of studies because of the variety of approaches used. Nevertheless, it was possible for us to estimate the age of emergence of sex differences in spatial abilities in the present analysis in two ways. First, when the three Linn and Petersen (1985) clusters were subdivided by age (Table 4), the effect sizes increased with age for all categories. For subjects under age 13, only mental rotation tests yielded significant effect sizes, whereas for adults, all categories led to significant sex differences.

We accomplished a more detailed analysis of age effects by correlating age with effect size and controlling for the year of birth of the participants (Table 9). This analysis indicated an overall linear association between age and effect size, with sex differences showing a significant increase with age. Of the individual tests, the Mental Rotations Test, the PMA Spatial Relations subtest, and the DAT Spatial Relations subtest all showed significant positive correlations with age.

It was also possible for us to determine the age at which sex differences appear on tests in which subjects younger than 13 years of age were studied. This analysis showed mixed results, with sex differences emerging as young as age 7 on the rod-and-frame test and as old as age 14 on the Embedded Figures Test. It thus appears that a single value is not representative of the age of emergence of sex differences in spatial tasks. Instead, it is better described in relation to the task used. This finding supports the notion that spatial ability is not a unitary construct, but consists of a collection of specific skills that are affected differently by the age of the participants. However, several different lines of evidence converge on the point that standard measures of spatial ability do not yield strong support for sex differences in prepubertal children. This does not necessarily mean that such differences do not exist; it simply reflects the fact that many of the tests considered in the present analysis may not be appropriate for use with young children. For example, children may find it difficult to comprehend just what is expected of them on some tests, which may thus produce a floor effect in which sex differences cannot be observed.

It is also important to keep in mind that young children do not necessarily use the same processes as adults to solve specific tasks. It is possible that the age of emergence of sex differences in a given task represents the transition between modes of processing. Such transition might be linked to stages of cognitive development as described by Piaget (1952). Furthermore, the processing requirements vary with the task under study. It is thus possible that two factors combine to produce the emergence of sex differences at different ages on different tests. These factors are (a) changes in children's mode of processing, and (b) component processes required in a specific task. This approach can explain why sex differences appear at age 7 on the rod-and-frame test. The age of 7 marks the beginning of the concrete–operational stage. It is possible that during this stage, children's understanding of the reversibility of actions and conservation allows them to better handle the processes required in the task. However, the presence of sex differences in rod-and-frame test performance at that age requires the assumption that sex differences exist in either the age of attainment of the concrete–operational stage or in the performance on tasks assessing cognitive development associated with this stage. Even though Meehan (1984) reported sex differences in the performance of tasks assessing formal thinking, such data are not available as far as concrete–operational tasks are concerned. Furthermore, the age of emergence of sex differences on other tests does not correspond with any obvious developmental landmarks. It is probable that the appearance of sex differences at various ages on different tests is due to the effect of cumulative experience (Baenninger & Newcombe, 1989). From this perspective, sex differences appear at a specific age on a given test because boys get more relevant experiences than girls as they age (Etaugh, 1983). For this reason, boys tend to reach the mode of processing required to solve a specific task earlier than do girls. Furthermore, children and adults may well use different approaches to the same task, because children are often facing quite unfamiliar testing situations, whereas adults can draw on established strategies. Again, any detailed explanation of the differences between adults and children in the way they respond to specific test situations would require more empirical studies to investigate specific hypotheses.

Analysis of the Year of Birth

The present analysis indicated a significant negative relation between year of birth and magnitude of sex differences for four tests, the Cards Rotation Test, the Water Level Test, the Embedded Figures Test, and the Identical Blocks Test, whereas one test also showed a negative but nonsignificant correlation (Paper Form Board). The remaining tests showed a positive relation between year of birth and sex differences. This relation was significant only for the Mental Rotations Test. This pattern of results partially supports Feingold's (1988) claim that cognitive sex differences are decreasing in magnitude, and it argues for the plausibility of the interpretation that changes in attitudes toward sex differences in cognitive abilities, changes in attitudes toward the sexes, or changes in educational practices have an effect on the size of sex-related differences in specific tasks.

The presence of a positive relation between year of birth and magnitude of effect sizes on the Mental Rotations Test contradicts Feingold's (1988) thesis. This test showed the largest mean effect size of all the tests sampled, and the correlation with year of birth suggests that when sex differences are large on a specific test, changes in attitudes or educational practices are not sufficient to reduce the magnitude of such differences. The failure of social changes to reduce the magnitude of sex differences on the Mental Rotations Test supports the claim that basic biological differences between females and males may play a role in determining cognitive sex differences on this test and should not be discarded in favor of exclusively environmental explanations. However, biological factors alone cannot account better than environmental factors for an increase in the magnitude of sex differences in recent years. More empirical work is needed if we are to disentangle the influence of social, environmental, and biological factors on the magnitude of sex differences in spatial performance in general and on the Mental Rotations Test in particular.

Causes of Sex Differences

Several explanations have been offered to account for the prevalence of sex differences in spatial abilities. For example, variables such as choice of strategy (Bryden, 1980), rate of maturation (Sanders & Soares, 1986; Waber, 1976), cerebral lateralization (Bryden, 1979; Levy, 1971), genetic complement (McGee, 1982), sex hormones (Imperato-McGinley, Pichardo, Gautier, Voyer, & Bryden, 1991; McGee, 1979), differential experience and socialization (Baenninger & Newcombe, 1989), and sex role identification (Nash, 1975; Signorella & Jamison, 1986) have all been proposed as possible causes. A meta-analysis such as that presented here cannot provide any clear answer as to which of these factors are important, although it may serve to guide further research in more productive directions.

First of all, we have specified a number of tests that show highly significant sex differences that are stable across age, at least after puberty, and have not decreased in recent years. These include the PMA Spatial Relations subtest, the generic mental rotation task, the Mental Rotations Test, and the rod-and-frame test. It would seem important to understand the processes underlying these tasks, to develop appropriate measures for use with young children, and to relate developmental changes in such processes to hormonal and experiential factors.

Second, we have uncovered a number of ways in which the method of administration affects performance (Mental Rotations Test, Identical Blocks Test, and Embedded Figures Test) and other cases in which the effect sizes seem to be decreasing in recent years (Cards Rotation Test, Water Level Test, Embedded Figures Test, and Identical Blocks Test). This suggests that a detailed examination of the way in which these tests are scored and administered might lead to some valuable information about the way in which test factors affect men and women differently. Year-of-birth effects, on the other hand, call for more thorough and systematic investigations of spatial test performance in the general population throughout the life span. Our analysis uncovered relatively few studies examining performance in participants older than 30, and this lack needs to be remedied.

Conclusions

The present analysis shows that sex differences, favoring males, are clearly significant and homogeneous on the Cards Rotation Test, the generic mental rotation tasks, the Spatial Relations subtest of the PMA, and the Paper Form Board. Sex differences on the Spatial Relation subtest of the DAT and Paper Folding are homogeneous but not significant. The rod-and-frame test and the Block Design subtest of the various Wechsler intelligence scales show sex differences in some age groups but not others. Finally, scoring and testing procedures proved to have an important influence on the magnitude of sex differences on the Mental Rotations Test, the Water Level Test, the Identical Blocks Test, and the Embedded Figures Test. The size of the fail-safe numbers associated with the different analyses suggests that the file drawer phenomenon (Rosenthal, 1984) is not sufficient to account for the prevalence of significant sex differences.

Given that sex differences are clearly established in some areas of spatial abilities, the next step is to determine their cause. Unfortunately, the nature of meta-analysis does not allow specific conclusions to be drawn concerning the causes of sex differences. It only provides pointers as to what factors should be examined in future research.

The present analysis indicates that certain tests are more appropriate than others for studying the cause of sex differences in spatial abilities, because they are more likely to demonstrate such differences. Most particularly, the Mental Rotations Test appears to produce the most robust sex differences among all tests included in the present analysis. This is especially obvious when this test is scored out of 20, in which case the average effect size represents a mean difference of 0.94 standard deviation units between the means of men and women. Thus, those conducting future research should concentrate on specific spatial tasks and give special consideration to the Mental Rotations Test in order to determine the factors underlying sex differences in spatial performance.

Another important aspect of the present study is that it partially supports the hypothesis that sex differences in spatial skills have decreased in magnitude in recent years. Nevertheless, this does not necessarily mean that sex differences are disappearing, because they remain significant in a number of tests. It is not clear whether the gap between women and men in terms of their spatial performance will ever be filled. However, the decrease in the magnitude of sex differences in recent years argues for the fact that attitudes concerning sex-related cognitive differences have changed. This attitude change is likely to have affected the way children are raised and the way women and men approach different tasks.

The need to partition further Linn and Petersen's (1985) classification of spatial tasks clearly indicates that their groupings require further refinement. Linn and Petersen obtained homogeneous effect sizes using this classification because of the content of their sample of effect sizes. Their clustering of spatial tasks does not hold with a different sample of studies. A classification of effect sizes in terms of test used and of the factors affecting the magnitude of sex differences appears to be appropriate because the present study included a nearly exhaustive sample of the published literature on sex differences in spatial abilities and because, after partitioning, homogeneity was achieved (or nearly achieved) in all but one grouping (on the Block Design subtest with participants under 13 years of age). Furthermore, researchers should keep in mind that spatial ability is not a unitary concept, but rather a collection of spatial components. Thus, one single test can allow the assessment of only one aspect of spatial performance. Several tests are required for a complete evaluation of spatial abilities. Finally, factors affecting the magnitude of sex differences should be either controlled for or systematically manipulated.

As for the present meta-analysis, it will, we hope, close the debate concerning the existence of sex differences in spatial abilities and orient research in productive ways to determine their cause. Those doing work in the future should focus on understanding specific components of spatial ability and how those components change developmentally as well as how they are affected by specific situations.

References

References marked with an asterisk indicate studies included in the meta-analysis. References marked with two asterisks indicate studies included in Linn and Petersen's (1985) meta-analysis as well as the present meta-analysis.

**Abravanel, E., & Gingold, H. (1977). Perceiving and representing orientation: Effects of the spatial framework. *Merrill-Palmer Quarterly, 23,* 265–278.

*Allen, M. J. (1974). Sex differences in spatial problem-solving styles. *Perceptual and Motor Skills, 39,* 843–846.

*Allen, M. J., & Hogeland, R. (1978). Spatial problem-solving strategies as functions of sex. *Perceptual and Motor Skills, 47,* 348–350.

*Andrieux, C. (1955). Contribution a l'étude des differences entre hommes et femmes dans la perception spaciale. *L'année Psychologique, 55,* 41–60.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20,* 327–344.

*Balistreri, E., & Busch-Rossnagel, N. A. (1989). Field independence as a function of sex, sex-roles, and the sex-role appropriateness of the task. *Perceptual and Motor Skills, 68,* 115–121.

Barratt, E. S. (1953). An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors. *Journal of Psychology, 36,* 17–25.

*Beatty, W. W., & Duncan, D. (1990). Relationship between performance on the everyday spatial activities test and on objective measures of spatial behavior on men and women. *Bulletin of the Psychonomic Society, 28,* 228–230.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential Aptitude Test.* New York: Psychological Corporation.

*Bergan, A., MacManis, D. L., & Melchert, P. A. (1971). Effects of social and token reinforcement on WISC Block Design performance. *Perceptual and Motor Skills, 32,* 871–880.

*Bernard, M. E., Boyle, G. J., & Jackling, B. F. (1990). Sex-role identity and mental ability. *Personality and Individual Differences, 11,* 213–217.

*Berry, J. W. (1966). Temne and Eskimo perceptual skills. *International Journal of Psychology, 1,* 207–229.

*Bieri, J. (1960). Parental identification, acceptance of authority, and within-sex differences in cognitive behavior. *Journal of Abnormal and Social Psychology, 60,* 76–79.

*Bieri, J., Bradburn, W. M., & Galinsky, M. D. (1958). Sex differences in perceptual behavior. *Journal of Personality, 26,* 1–12.

*Bigelow, G. (1971). Field dependence–field independence in 5 to 10-year-old children. *Journal of Educational Research, 64,* 397–400.

*Birkett, P. (1980). Predicting spatial ability from hemispheric "nonverbal" lateralisation: Sex, handedness and task differences implicate encoding strategy effects. *Acta Psychologica, 46,* 1–14.

*Blum, J. E., Fosshage, J. L., & Jarvik, L. F. (1972). Intellectual changes and sex differences in octogenarians: A twenty-year longitudinal study of aging. *Developmental Psychology, 7,* 178–187.

*Bogo, W., Winget, C., & Gleser, G. C. (1970). Ego defenses and perceptual styles. *Perceptual and Motor Skills, 30,* 599–604.

**Bouchard, T. J., Jr., & McGee, M. G. (1977). Sex differences in human spatial ability: Not an X-linked recessive gene effect. *Social Biology, 24,* 332–335.

*Brainerd, C. J., & Huevel, K. V. (1974). Development of geometric imagery in five- to eight-year-olds. *Genetic Psychology Monographs, 89,* 89–143.

Brawn, B., & Kimball, M. M. (1988). Strategy, experience, and confidence: Performance on a mental rotation task. *Canadian Psychology, 29*(2a). (Abstract 92)

Bryden, M. P. (1979). Evidence for sex-related differences in cerebral organization. In M. A. Wittig & A. C. Petersen (Eds.), *Sex-related differences in cognitive functioning* (pp. 121–143). New York: Academic Press.

Bryden, M. P. (1980). Sex differences in brain organization: Different brains or different strategies? *Behavioral and Brain Sciences, 3,* 230–231.

*Bryden, M. P., George, J., & Inch, R. (1990). Sex differences and the role of figural complexity in determining the rate of mental rotation. *Perceptual and Motor Skills, 70,* 467–477.

Burnett, S. A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist, 41,* 1012–1014.

Burnett, S. A., Lane, D. M., & Dratt, L. M. (1979). Spatial visualisation and sex differences in quantitative ability. *Intelligence, 3,* 345–354.

*Burnett, S. A., Lane, D. M., & Dratt, L. M. (1982). Spatial ability and handedness. *Intelligence, 6,* 57–68.

*Bush, D. F., & Coward, R. T. (1974). Sex differences in the solution of achromatic and chromatic embedded figures. *Perceptual and Motor Skills, 39,* 1121–1122.

*Caldwell, E. C., & Hall, V. C. (1970). Concept learning in discrimination tasks. *Developmental Psychology, 2,* 41–48.

Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? *American Psychologist, 40,* 786–799.

Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 221–253). Hillsdale, NJ: Erlbaum.

*Casey, M. B., Colon, D., & Goris, Y. (1992). Family handedness as a predictor of mental rotation ability among minority girls in a math-science training program. *Brain & Cognition, 18,* 88–96.

*Coates, S. W. (1972). *Preschool Embedded Figures Test.* Palo Alto, CA: Consulting Psychologists Press.

*Cochran, K. F., & Wheatley, G. H. (1988). Ability and sex-related differences in cognitive strategies on spatial tasks. *Journal of General Psychology, 116,* 43–55.

Cohen, H. G. (1978). The scaling of six topological Piagetian groupings, as well as the effect that certain selected variables have on the attainment of these groupings and some of their homologs in the logical domain. *Journal of Research in Science Teaching, 15,* 115–127.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

**Connor, J. M., Schackman, M., & Serbin, L. A. (1978). Sex-related differences in response to practice on a visual-spatial test and generalisation to a related test. *Child Development, 49,* 24–29.

*Connor, J. M., Serbin, L. A., & Schackman, M. (1977). Sex differences in children's response to training on a visual-spatial test. *Developmental Psychology, 13,* 293–294.

*Corah, N. L. (1965). Differentiation in children and their parents. *Journal of Personality, 33,* 300–308.

*Corballis, M. C., & Sergent, J. (1989). Hemispheric specialization for mental rotation. *Cortex, 25,* 15–25.

*Corley, R. P., DeFries, J. C., Kuse, A. R., & Vandenberg, S. G. (1980). Familial resemblance for the identical blocks test of spatial ability: No evidence for X-linkage. *Behavior Genetics, 10,* 211–215.

*Crandall, V. C., & Lacey, B. W. (1972). Children's perceptions of internal–external control in intellectual–academic situations and their Embedded Figures Test performance. *Child Development, 43,* 1123–1134.

*Crandall, V. J., & Sinkeldam, C. (1964). Children's dependent and achievement behaviors in social situations and their perceptual field dependence. *Journal of Personality, 32,* 1–22.

*Cronin, V. (1967). Mirror-image reversal discrimination in kindergarten and first grade children. *Journal of Experimental Child Psychology, 5,* 577–585.

**DeAvila, E. A., Havassy, B., with Pascual-Leone, J. (1976). *Mexican-American school children: A neo-Piagetian analysis.* Washington, DC: Georgetown University Press.

*DeLucca, J., Burright, R. G., & Donovick, P. J. (1990). Manual asymmetries during verbal and spatial block design construction. *Cortex, 26,* 541–554.

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

Eliot, J. (1986). Comment on Caplan, MacPherson, and Tobin. *American Psychologist, 41,* 1011.

*Embretson, S. E. (1987). Improving the measurement of spatial aptitude by dynamic testing. *Intelligence, 11,* 333–358.

**Erdos, G. (1979). Sex differences in feedback: Effects on rod-and-frame performance. *Perceptual and Motor Skills, 48,* 1279–1285.

Etaugh, C. (1983). Introduction: The influence of environmental factors on sex differences in children's play. In M. B. Liss (Ed.), *Social and cognitive skills: Sex roles and children's play* (pp. 1–19). New York: Academic Press.

Fairweather, H. (1976). Sex differences in cognition. *Cognition, 4,* 231–280.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95–103.

**Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal, 14,* 51–71.

*Fiebert, M. (1967). Cognitive styles in the deaf. *Perceptual and Motor Skills, 24,* 319–329.

*Fralley, J. S., Eliot, J., & Dayton, C. M. (1978). Further study of the X-linked recessive gene hypotheses for inheritance of spatial abilities. *Perceptual and Motor Skills, 47,* 1023–1029.

*Freedman, R., & Rovegno, L. (1981). Ocular dominance, cognitive strategy, and sex differences in spatial abilities. *Perceptual and Motor Skills, 52,* 651–654.

*Gainer, W. L. (1962). The ability of the WISC subtests to discriminate between boys and girls of average intelligence. *California Journal of Educational Research, 13,* 9–16.

**Geiringer, E. R., & Hyde, J. S. (1976). Sex differences on Piaget's water-level task: Spatial ability incognito. *Perceptual and Motor Skills, 42,* 1323–1328.

*Gladue, B. A., Beatty, W. W., Larson, J., & Staton, R. D. (1990). Sexual orientation and spatial ability in men and women. *Psychobiology, 18,* 101–108.

*Golbeck, S. L. (1986). The role of physical content in Piagetian spatial tasks: Sex differences in spatial knowledge? *Journal of Research in Science Teaching, 23,* 365–376.

*Goldstein, A. G., & Chance, J. E. (1965). Effects of practice on sex-related differences in performance on Embedded Figures. *Psychonomic Science, 3,* 361–362.

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition, 18,* 546–550.

*Goodenough, D. R., & Eagle, C. J. (1963). A modification of the Embedded Figures Test for use with young children. *Journal of Genetic Psychology, 103,* 67–74.

*Gross, F. (1959). The role of set in perception of the upright. *Journal of Personality, 27,* 95–103.

*Gruen, A. (1955). The relation of dancing experience and personality to perception. *Psychological Monographs: General and Applied, 69,* Whole No. 399.

**Guay, R. B., & McDaniel, E. D. (1977). The relationship of mathematical achievement and spatial abilities among elementary school children. *Journal of Research in Mathematics Education, 8,* 211–215.

*Hakstian, A. R., & Cattell, R. B. (1975). An examination of adolescent sex differences in some ability and personality traits. *Canadian Journal of Behavioral Science, 7,* 295–315.

Halpern, D. F. (1986). A different answer to the question, "Do sex-related differences in spatial abilities exist?" *American Psychologist, 41,* 1014–1015.

**Harris, L. J., Hanley, C., & Best, C. T. (1977). Conservation of horizontality: Sex differences in sixth-graders and college students. In M. C. Smart & P. C. Smart (Eds.), *Readings in child development and relationships* (pp. 375–387). New York: Macmillan.

*Harshman, R. A., Hampson, E., & Berenbaum, S. A. (1983). Individual differences in cognitive abilities and brain organization. *Canadian Journal of Psychology, 37,* 144–192.

Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis* (pp. 14–50). Baltimore: John Hopkins University Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

*Herzberg, F., & Lepkin, M. (1954). A study of sex differences on the Primary Mental Abilities Test. *Educational and Psychological Measurement, 14,* 687–689.

Hiscock, M. (1986). On sex differences in spatial abilities. *American Psychologist, 41,* 1011–1012.

*Hobson, J. R. (1947). Sex differences in primary mental abilities. *Journal of Educational Research, 41,* 126–132.

Horan, P. F., & Rosser, R. A. (1984). A multivariate analysis of spatial ability by sex. *Developmental Review, 4,* 387–411.

**Hughes, R. N. (1978). Sex differences in field dependence: Effects of unlimited time on group Embedded-Figures Test performance. *Perceptual and Motor Skills, 47,* 1246.

*Hult, R. E., & Brous, C. W. (1986). Spatial visualization: Athletic skills and sex differences. *Perceptual and Motor Skills, 63,* 163–168.

Hyde, J. S. (1981). How large are cognitive gender differences? *American Psychologist, 36,* 892–901.

**Hyde, J. S., Geiringer, E. R., & Yen, W. M. (1975). On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. *Multivariate Behavioral Research, 10,* 289–309.

*Immergluck, L., & Mearini, M. C. (1969). Age and sex differences in response to embedded figures and reversible figures. *Journal of Experimental Child Psychology, 8,* 210–221.

Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., & Bryden, M. P. (1991). Cognitive abilities in androgen insensitive subjects—Comparison with control males and females from the same kindred. *Clinical Endocrinology, 34,* 341–347.

*Jahoda, G. (1979). On the nature of difficulties in spatial–perceptual tasks: Ethnic and sex differences. *British Journal of Psychology, 70,* 351–363.

**Jamison, W., & Signorella, M. L. (1980). Sex-typing and spatial ability: The association between masculinity and success on Piaget's water-level task. *Sex Roles, 6,* 345–353.

*Johnson, O., & Harley, C. (1980). Handedness and sex differences in cognitive tests of brain laterality. *Cortex, 16,* 73–82.

*Johnson, S., Flinn, J. M., & Tyer, Z. E. (1979). Effect of practice and training in spatial skills on embedded figures scores of males and females. *Perceptual and Motor Skills, 48,* 975–984.

*Jones, B., & Anuza, T. (1982). Effects of sex, handedness, stimulus and visual field on "mental rotation." *Cortex, 18,* 501–514.

*Kaess, D. W. (1971). Measures of form constancy: Developmental trends. *Developmental Psychology, 4,* 296.

*Kagan, J., Rosman, B. L., Day, D., Albert, J., & Phillips, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs: General and Applied, 78* (Whole No. 578).

*Kail, R., Carter, P., & Pellegrino, J. (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics, 26,* 182–186.

*Kail, R., Stevenson, M. R., & Black, K. N. (1984). Absence of a sex difference in algorithms for spatial problem solving. *Intelligence, 8,* 37–46.

*Kalichman, S. C. (1986). Horizontality as a function of sex and college major. *Perceptual and Motor Skills, 63,* 903–906.

*Kalichman, S. C. (1989). Sex roles and sex differences in adult spatial performance. *Journal of Genetic Psychology, 150,* 93–100.

**Karplus, R., Pulos, S., & Stage, E. (1983). Proportional reasoning of early adolescents. In R. Lesh & M. Landau (Eds.), *Acquisitions of mathematics concepts and processes* (pp. 45–90). New York: Academic Press.

*Kato, N. (1965). A fundamental study of rod-frame test. *Japanese Psychological Research, 7,* 61–68.

**Kelly, J., & Kelly, G. (1977). Perception of horizontality by male and female college students. *Perceptual and Motor Skills, 44,* 724–726.

*Kenyon, J. (1984). Paper-and-pencil test of Piaget's water level test: Sex differences and test modality. *Perceptual and Motor Skills, 59,* 739–742.

*Keogh, B. K., & Ryan, S. R. (1971). Use of three measures of field organisation with young children. *Perceptual and Motor Skills, 33,* 466.

*Kerns, K. A., & Berenbaum, S. A. (1991). Sex differences in spatial ability in children. *Behavior Genetics, 21,* 383–396.

*Kershner, J. R. (1971). Children's acquisition of visuo-spatial dimensionality: A conservation study. *Developmental Psychology, 5,* 454–462.

*Keyes, S. (1983). Sex differences in cognitive abilities and sex-role stereotypes in Hong Kong Chinese adolescents. *Sex Roles, 9,* 853–870.

*Levinson, B. M. (1960). Comparative study of verbal and performance ability of monolingual and bilingual native-born Jewish children of traditional parentage. *Journal of Genetic Psychology, 97,* 93–112.

Levy, J. (1971). Lateral specialization of the human brain: Behavioral manifestations and possible evolutionary basis. In J. A. Kiger, Jr. (Ed.), *The biology of behavior* (pp. 159–180). Corvallis: Oregon State University Press.

**Liben, L. S. (1974). Operative understanding of horizontality and its relation to long-term memory. *Child Development, 45,* 416–424.

**Liben, L. S. (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill-Palmer Quarterly, 24,* 97–110.

*Liben, L. S. (1991). Adults' performance on horizontality tasks: Conflicting frames of reference. *Developmental Psychology, 27,* 285–294.

**Liben, L. S., & Golbeck, S. L. (1980). Sex differences in performance on Piagetian spatial tasks: Differences in competence or performance? *Child Development, 51,* 594–597.

*Liben, L. S., & Golbeck, S. L. (1984). Performance on Piagetian horizontality and verticality tasks: Sex-related differences in knowledge of relevant physical phenomena. *Developmental Psychology, 20,* 595–606.

*Liben, L. S., & Golbeck, S. L. (1986). Adults' demonstration underlying euclidean concepts in relation to task context. *Developmental Psychology, 22,* 487–490.

Likert, R., & Quasha, W. H. (1941). *The revised Minnesota Paper Form Board.* New York: Psychological Corporation.

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

Linn, M. C., & Petersen, A. C. (1986). Gender differences in spatial ability: Implications for mathematics and science performance. In J. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advance through meta-analysis.* Baltimore: John Hopkins University Press.

**Linn, M. C., & Pulos, S. (1983). Male–female differences in predicting displaced volume: Strategy usage, aptitude relationships, and experience influences. *Journal of Educational Psychology, 75,* 86–96.

*Lusk, E. J., & Wright, H. (1981). Differences in sex and curricula on learning the group embedded figures tests. *Perceptual and Motor Skills, 53,* 8–10.

*MacArthur, R. (1967). Sex differences in field dependence for the Eskimo. *International Journal of Psychology, 2,* 139–140.

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

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

*Marino, M. F., & McKeever, W. F. (1989). Spatial processing laterality and spatial visualization ability: Relations to sex and familial sinistrality variables. *Bulletin of the Psychonomic Society, 27,* 135–137.

*Maxwell, J. W., Croake, J. W., & Biddle, A. P. (1975). Sex differences in the comprehension of spatial orientation. *Journal of Psychology, 91,* 127–131.

*Mayes, J. T. (1982). Hemisphere function and spatial ability: An exploratory study of sex and cultural differences. *International Journal of Psychology, 17,* 65–80.

**McGee, M. G. (1978). Effects of training and practice on sex differences in mental rotation test score. *Journal of Psychology, 100,* 87–90.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86,* 889–918.

McGee, M. G. (1982). Spatial abilities: The influence of genetic factors. In M. Potegal (Ed.), *Spatial abilities: Development and physiological foundations* (pp. 199–222). New York: Academic Press.

*McGillicuddy-DeLisi, A. V., DeLisi, R., & Youniss, J. (1978). Representation of the horizontal coordinate with and without liquid. *Merrill-Palmer Quarterly, 24,* 199–208.

*McGilligan, R. P., & Barclay, A. G. (1974). Sex differences and spatial ability factors in Witkin's "differentiation" construct. *Journal of Clinical Psychology, 30,* 528–532.

*McGlone, J., & Davidson, W. (1973). The relationship between cerebral speech laterality and spatial ability with special reference to sex and hand preference. *Neuropsychologia, 11,* 105–113.

*McKeever, W. F. (1986). The influences of handedness, sex, familial sinistrality and androgyny on language laterality, verbal ability, and spatial ability. *Cortex, 22,* 521–537.

Meehan, A. (1984). A meta-analysis of sex differences in formal operational thought. *Child Development, 55,* 1110–1124.

*Meehan, A., & Overton, W. (1986). Gender differences in expectancies for success and performance on Piagetian spatial tasks. *Merrill-Palmer Quarterly, 32,* 427–441.

*Meyer, W. J., & Bendig, A. W. (1961). A longitudinal study of the Primary Mental Ability Test. *Journal of Educational Psychology, 52,* 50–60.

*Miller, L. K., & Santoni, V. (1986). Sex differences in spatial abilities: Strategic and experiential correlates. *Acta Psychologica, 62,* 225–235.

*Morell, J. A. (1976). Age, sex, training, and the measurement of field dependence. *Journal of Experimental Child Psychology, 22,* 100–112.

*Morf, M. E., & Howitt, R. (1970). Rod-and-frame test performance as a function of momentary arousal. *Perceptual and Motor Skills, 31,* 705–708.

*Morf, M. E., Kavanaugh, R. D., & McConville, M. (1971). Intratest and sex differences on a portable rod-and-frame test. *Perceptual and Motor Skills, 32,* 727–733.

*Morris, B. B. (1971). Effects of angle, sex and cue on adults' perception of the horizontal. *Perceptual and Motor Skills, 32,* 827–830.

*Mower-Popeil, E., & DeLisi, R. (1984). An examination of spatial ability in relation to factors from the BEM Sex-Role Inventory. *Perceptual and Motor Skills, 59,* 131–136.

*Mumbauer, C. C., & Miller, J. O. (1970). Socioeconomic background and cognitive functioning in preschool children. *Child Development, 41,* 471–480.

*Myer, K. A., & Hensley, J. H. (1984). Cognitive style, gender, and self-report of principle as predictor of adult performance on Piaget's water level task. *Journal of Genetic Psychology, 144,* 179–183.

*Nash, S. C. (1975). The relationship among sex-role stereotyping, sex-role preference, and the sex difference in spatial visualization. *Sex Roles, 1,* 15–32.

*Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles, 9,* 377–386.

*Norman, R. D. (1953). Sex differences and other aspects of young superior adults performance on the Wechsler–Bellevue. *Journal of Consulting Psychology, 17,* 411–418.

*Okonji, M. O. (1969). The differential effects of rural and urban upbringing on the development of cognitive styles. *International Journal of Psychology, 4,* 293–305.

*Olson, D. M., & Eliot, J. (1986). Relationships between experiences, processing style, and sex-related differences in performance on spatial tests. *Perceptual and Motor Skills, 62,* 447–460.

*Olson, D. M., Eliot, J., & Hardy, R. C. (1988). Relationships between activities and sex-related differences in performance on spatial tests. *Perceptual and Motor Skills, 67,* 223–232.

*Oltman, P. R. (1968). A portable rod-and-frame apparatus. *Perceptual and Motor Skills, 26,* 503–506.

*Ozer, D. J. (1987). Personality, intelligence, and spatial visualization: Correlates of mental rotations test performance. *Journal of Personality and Social Psychology, 53,* 129–134.

*Pande, C. G. (1970). Sex differences in field dependence: Confirmation with Indian sample. *Perceptual and Motor Skills, 31,* 70.

*Parlee, M. B., & Rajagopal, J. (1974). Sex differences on the Embedded-Figures Test: A cross-cultural comparison of college students in India and in the United States. *Perceptual and Motor Skills, 39,* 1311–1314.

*Pennings, A. H. (1991). Altering the strategies in embedded figures and water level tasks via instructions: A neo-Piagetian learning study. *Perceptual and Motor Skills, 72,* 639–660.

*Persaud, G. (1991). Age and sex differences on (4) measures of cognitive ability. *Perceptual and Motor Skills, 72,* 1172–1174.

Petersen, A. C. (1976). Physical androgyny and cognitive functioning in adolescence. *Developmental Psychology, 12,* 524–533.

**Petersen, A. C., & Gitelson, I. B. (in press). *Toward understanding sex-related differences in cognitive performance.* New York: Academic Press. (As cited in Linn & Petersen, 1986)

*Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on spatial tasks: Proposed genetics and environmental factors. *Brain and Cognition, 17,* 1–22.

Piaget, J. (1952). *The origins of intelligence in children.* New York: International Universities Press.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space.* London: Routledge & Kegan Paul.

**Pitblado, C. (1976). Superior performance by women in a visual orienting task: A limit on the concept of field dependence. *Perceptual and Motor Skills, 42,* 1195–1200.

**Pulos, S. M., Stage, E. K., & Karplus, R. (1983). *Cognitive correlates of proportional reasoning in early adolescence.* Berkeley: Lawrence Hall of Science, University of California.

*Rebelsky, F. (1964). Adult perception of the horizontal. *Perceptual and Motor Skills, 19,* 371–374.

*Reppucci, N. D. (1971). Parental education, sex differences, and performance on cognitive tasks among two-year-old children. *Developmental Psychology, 4,* 248–253.

**Richmond, P. G. (1980). A limited sex difference in spatial test scores with a preadolescent sample. *Child Development, 51,* 601–602.

*Robert, M. (1990). Sex-typing of the water-level task: There is more than meets the eye. *International Journal of Psychology, 25,* 475–490.

*Robert, M., & Morin, P. (1993). Gender differences in horizontality and verticality representation in relation to initial position of the stimuli. *Canadian Journal of Experimental Psychology, 47,* 507–522.

*Robert, M., & Ohlmann, T. (1991). *Effect of gravitational cues from stimuli and body orientations on water-level representation.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

*Robert, M., & Tanguay, M. (1990). Perception and representation of the Euclidean coordinates in mature and elderly men and women. *Experimental Aging Research, 16,* 123–131.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rosenthal, R. (1980). Summarizing significance levels. In R. Rosenthal (Ed.), *Quantitative assessment of research domains* (pp. 33–46). San Francisco: Jossey-Bass.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

*Ruble, D. N., & Nakamura, C. Y. (1972). Task orientation versus social orientation in young children and their attention to relevant social cues. *Child Development, 43,* 471–480.

*Saarni, C. I. (1973). Piagetian operations and field independence as factors in children's problem-solving performance. *Child Development, 44,* 338–345.

*Samuel, W. (1983). Sex differences in spatial ability reflected in performance on IQ subtests by black or white examinees. *Personality and Individual Differences, 4,* 219–221.

Sanders, B., Cohen, M. R., & Soares, M. P. (1986). The sex differences in spatial ability: A rejoinder. *American Psychologist, 41,* 1015–1016.

*Sanders, B., & Soares, M. P. (1986). Sexual maturation and spatial ability in college students. *Developmental Psychology, 22,* 199–203.

**Sanders, B., Soares, M. P., & D'Aquila, J. (1982). The sex difference on one test of spatial visualisation: A nontrivial difference. *Child Development, 53,* 1106–1110.

*Sarason, I. G., & Minard, J. (1962). Test anxiety, experimental instructions and the Wechsler Adult Intelligence Scale. *Journal of Educational Psychology, 53,* 299–302.

*Scholan, K., & Smith, M. (1990). A sex difference in field dependence/independence in the absence of vestibular activation and eye movements. *Perceptual and Motor Skills, 71,* 763–768.

*Schubert, J., & Cropley, A. J. (1972). Verbal regulation of behavior and IQ in Canadian Indian and white children. *Developmental Psychology, 7,* 295–301.

*Schwartz, D. W., & Karp, S. A. (1967). Field dependence in a geriatric population. *Perceptual and Motor Skills, 24,* 495–504.

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

**Sherman, J. (1980). Mathematics, spatial visualization and related factors: Changes in boys and girls, grades 8–11. *Journal of Educational Psychology, 72,* 476–482.

*Sherman, J. A. (1974). Field articulation, sex, spatial visualisation, dependency, practice, laterality of the brain, and birth order. *Perceptual and Motor Skills, 38,* 1223–1235.

*Sherman, J. A., & Fennema, E. (1978). Distribution of spatial visualization and mathematical problem solving scores: A test of Stafford's X-linked hypotheses. *Psychology of Women Quarterly, 3,* 157–167.

**Signorella, M. L., & Jamison, W. (1978). Sex differences in the correlation among field dependence, spatial ability, sex-role orientation, and performance on Piaget's water-level task. *Developmental Psychology, 14,* 689–690.

Signorella, M. L., & Jamison, W. (1986). Masculinity, femininity, androgyny, and cognitive performance: A meta-analysis. *Psychological Bulletin, 100,* 207–228.

*Signorella, M. L., Jamison, W., & Krupa, M. H. (1989). Predicting spatial performance from gender stereotyping in activity preferences and self-concept. *Developmental Psychology, 25,* 89–95.

*Silverman, J., Buchsbaum, M., & Stierlin, H. (1973). Sex differences in perceptual differentiation and stimulus intensity control. *Journal of Personality and Social Psychology, 25,* 309–318.

*Spradlin, L. M., & Hensley, J. H. (1979). An extension of evidence on sex differences in the failure of a water level conservation task by adults. *Journal of Genetic Psychology, 135,* 305–306.

*Stafford, R. E. (1961). Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and Motor Skills, 13,* 428.

*Stericker, A., & LeVesconte, S. (1982). Effect of brief training on sex-related differences in visual-spatial skill. *Journal of Personality and Social Psychology, 43,* 1018–1029.

*Stouwie, R. J., Hetherington, E. M., & Parke, R. D. (1970). Some determinants of children's self-reward criteria. *Developmental Psychology, 3,* 313–319.

*Stuart, I. R., Breslow, A., Brechner, S., Ilyus, R. B., & Wolpoff, M. (1965). The question of constitutional influence on perceptual style. *Perceptual and Motor Skills, 20,* 419–420.

*Tapley, S. M., & Bryden, M. P. (1977). An investigation of sex differences in spatial ability: Mental rotation of three dimensional objects. *Canadian Journal of Psychology, 31,* 122–130.

**Taylor, L. J. (1977). Sex and psychological differentiation. *Psychological Reports, 41,* 192–194.

*Thompson, E. G., Harris, L. J., & Mann, I. (1981). Relationships among cognitive complexity, sex, and spatial task performance in college students. *British Journal of Psychology, 72,* 249–256.

Thurstone, T. G. (1958). *Manual for the SRA Primary Mental Abilities.* Chicago: Science Research Associates.

*Tracy, D. M. (1990). Toy-playing behavior, sex-role orientation, spatial ability, and science achievement. *Journal of Research in Science Teaching, 27,* 637–649.

*Treagust, D. F. (1980). Gender-related differences of adolescents in spatial representational thought. *Journal of Research in Science Teaching, 17,* 91–97.

*Van Blerkom, M. L. (1987). Haptic lateralization, field dependence, and sex. *Perceptual and Motor Skills, 64,* 907–914.

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

*Vandenberg, S. G., Kuse, A. R., & Vogler, G. P. (1985). Searching for correlates of spatial ability. *Perceptual and Motor Skills, 60,* 343–350.

**Van Leeuwen, M. A. (1978). A cross-cultural examination of psychological differentiation in males and females. *International Journal of Psychology, 13,* 87–122.

*Vaught, G. M. (1965). The relationship of sex role identification and ego-strength to sex differences in the rod-and-frame test. *Journal of Personality, 33,* 271–283.

*Very, P. S. (1967). Differential factor structures in mathematical abilities. *Genetic Psychology Monographs, 75,* 169–207.

*Voyer, D., & Bryden, M. P. (1990). Mental rotation: A study of gender differences and sex role identity. *Canadian Psychology, 31*(2a), 253.

*Voyer, D., & Bryden, M. P. (1993). Masking and laterality effects in a lateralized rod-and-frame task. *Canadian Journal of Experimental Psychology, 47,* 26–37.

*Voyer, D., & Chisholm, B. D. (1992). *Factors affecting performance in a lateralized embedded figures task.* Poster presented at the 33rd meeting of the Psychonomic Society, St. Louis, MO.

Waber, D. P. (1976). Sex differences in cognition: A function of maturation rate? *Science, 192,* 572–574.

*Waber, D. P., Carlson, D., & Mann, M. (1982). Developmental and differential aspects of mental rotation in early adolescence. *Child Development, 53,* 1614–1621.

**Walker, J., & Krasnoff, A. (1978). The horizontality principle in young men and women. *Perceptual and Motor Skills, 46,* 1055–1061.

Wechsler, D. (1946). *The Wechsler-Bellevue Intelligence Scale, Form II.* New York: Psychological Corporation.

Wechsler, D. (1949). *The Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* New York: Psychological Corporation.

*Weinberg, S., & Rabinowitz, J. (1970). A sex difference in the Wechsler IQ vocabulary scores as a predictor of strategy in a probability-learning task performed by adolescents. *Developmental Psychology, 3,* 218–224.

*Willemsen, E. W. (1974). Judgment of the vertical as a function of stimulus characteristics, age, and sex. *Perceptual and Motor Skills, 38,* 1344.

*Willemsen, E., & Reynolds, B. (1973). Sex differences in adults' judgments of the horizontal. *Developmental Psychology, 8,* 309.

*Willoughby, R. H. (1967). Field-dependence and locus of control. *Perceptual and Motor Skills, 24,* 671–672.

*Wilson, J. R., DeFries, J. C., McClearn, G. E., Vandenberg, S. G., Johnson, R. C., & Rashad, M. N. (1975). Cognitive abilities: Use of family data as a control to assess sex and age differences in two ethnic groups. *International Journal of Aging and Human Development, 6,* 261–276.

**Wilson, J. R., & Vandenberg, S. G. (1978). Sex differences in cognition: Evidence from the Hawaii family study. In T. E. McGill, D. E. Dewsbury, & B. Sachs (Eds.), *Sex and behavior: Status and prospectus* (pp. 317–336). New York: Plenum.

*Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality, 19,* 1–15.

Witkin, H. A., & Asch, S. E. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology, 38,* 762–782.

*Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology, 7,* 291–300.

**Wittig, M. A., & Allen, M. J. (1984). Measurement of adult performance on Piaget's water horizontality task. *Intelligence, 8,* 305–313.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Newbury Park, CA: Sage.

*Yen, W. M. (1975). Sex-linked major-gene influences on selected types of spatial performance. *Behavior Genetics, 5,* 281–298.

*Yeo, R. A., & Cohen, D. B. (1983). Familial sinistrality and sex differences in cognitive ability. *Cortex, 19,* 125–130.

Received September 22, 1993
Revision received August 25, 1994
Accepted August 26, 1994 ■