# How Large Are Cognitive Gender Differences?

## *A Meta-Analysis Using $\omega^2$ and* d

JANET SHIBLEY HYDE *Denison University*

ABSTRACT: *Maccoby and Jacklin concluded that the following cognitive gender differences are well-established: verbal ability, quantitative ability, and visual-spatial ability. The present study applied meta-analysis techniques to the studies cited by Maccoby and Jacklin, assessing the magnitude of gender differences using both $\omega^2$ and* d *statistics. The results indicated that gender differences in all of these abilities are very small: For verbal ability, the median $\omega^2$ was .01 and the median* d *was .24; for quantitative ability, the median values of $\omega^2$ and* d *were .01 and .43, respectively; for visual-spatial ability, they were .043 and .45, respectively; and for field articulation, $\omega^2$ was .025 and* d *was .51. Discussion focused on the practical implications of the finding that these "well-established" differences were in fact very small. Concerns about sampling were raised. The problem was also discussed in the context of a larger issue in psychological research: the limitations of the hypothesis-testing approach and the need to estimate the magnitude of effects.*

Maccoby and Jacklin (1974) reviewed the enormous literature on psychological gender differences. In particular, they concluded that three cognitive gender differences were "well-established": Girls have greater verbal ability than boys, and boys have better visual–spatial ability and better mathematical ability than girls.

Sherman (1978) re-reviewed the evidence on cognitive gender differences and pointed out that even for these supposedly well-established differences, the magnitude of the gender difference was very small. For example, Maccoby and Jacklin (1974, Table 3.4) computed the magnitude of gender differences in verbal ability for a subset of studies providing sufficient data. Typically, the magnitude of the difference was only about .25 standard deviations. Sherman (1978, p. 43) noted that the proportion of variance ($\omega^2$) accounted for by gender differences in verbal ability for the 1955 standardization of the WAIS (Matarazzo, 1972) was less than .01.

Meta-analysis is a technique for analyzing a body of research on a particular topic by statistical analysis of the analyses of the individual studies (Glass, 1976). It is becoming increasingly popular as a technique for evaluating a given area of research (for examples, see Hall, 1978; Kulik, Kulik, & Cohen, 1979; Smith, 1980; Smith & Glass, 1977). Typically, these studies use the measure $d = \frac{M_1 - M_2}{SD}$ as a measure of the magnitude of differences between two groups, that is, as a measure of effect size.

The purpose of the present article is to reanalyze the studies on cognitive gender differences considered to be well-established by Maccoby and Jacklin (1974) and to determine the magnitude of these gender differences. Two measures of effect size were used: $\omega^2$ (Hays, 1963) and $d$. Such an analysis is particularly important because the term "well-established" is often taken to mean "large." Maccoby and Jacklin's review has had a widespread influence, and their conclusions are cited in many introductory-level psychology texts. Thus, large numbers of people may have the impression that there are large gender differences in cognitive abilities. This, in turn, may affect practices such as vocational counseling. For example, a girl might be discouraged from a career in mathematics or science because of the "well-established" and "large" superiority of males in quantitative ability (despite repeated caveats in many texts about the great overlap in male and female distributions). Thus, it seems important to determine whether

Requests for reprints should be sent to Janet Shibley Hyde, Department of Psychology, Denison University, Granville, Ohio 43023.



Vol. 36, No. 8, 892–901


0003-066X/81/3608-0892$00.75

these "well-established" differences are in fact "large."

This study can be viewed as one example of a larger issue in psychological research: criticism of the hypothesis-testing approach and a suggested alternative, estimating the magnitude of effects. Cognitive gender differences are good examples because they are widely believed to be well-established, yet the size of the gender difference has rarely been estimated.

## *Method*

### SELECTION OF THE DATA SET

The studies analyzed here are those listed in the Maccoby and Jacklin (1974) review. Thus, the studies reviewed on verbal ability are those listed in Tables 3.3 and 3.4 in Maccoby and Jacklin's book. The studies of quantitative ability are those in Maccoby and Jacklin's Table 3.5. The studies of visual–spatial ability are from Maccoby and Jacklin's Table 3.7. And, finally, the studies of field articulation (visual–analytic spatial ability) are from Maccoby and Jacklin's Table 3.8. These studies were chosen because they are precisely the studies that have led to the conclusion that these cognitive gender differences are "well-established."

Two restrictions were placed on the studies from those tables analyzed here. First, studies based on non-U.S. samples were omitted. It was seen as a reasonable goal simply to determine the size of gender differences in the United States. It seems quite possible that owing to differences in factors such as socialization practices or school curricula, gender differences present in one culture may not be present in another. Thus, inclusion of non-U.S. samples might only have clouded the picture. Second, a restriction of studies was made according to the age of subjects in the studies, based on the developmental nature of Maccoby and Jacklin's conclusions. In particular, they concluded that gender differences in verbal ability did not emerge until subjects were about 11 years of age. Thus, in an assessment of the magnitude of the established gender difference, it would be inappropriate to include studies based on subjects younger than 11 years of age. Hence, in Maccoby and Jacklin's Table 3.3, the study by Achenbach (1969) with 11-year-olds and all studies following it in the table are included. Similarly, Maccoby and Jacklin concluded that the gender difference in quantitative ability does not emerge until children are 12 to 13 years of age. Accordingly, the study by Hilton and Berglund (Note 1) including 10-, 12-, 14-, and 16-year-olds and all succeeding studies in Table 3.5 were included in the present review. Maccoby and Jacklin concluded that the gender difference in spatial ability was found consistently in adolescence and adulthood but not in childhood. Accordingly, the study by Nash (1973) with 11- and 14-year-olds and all succeeding studies in Table 3.7 were included in the analysis of visual–spatial ability here, and the study by Nash (1973) and all succeeding studies in Table 3.8 were included in the analysis of field articulation here.

Within the analysis of research on a particular cognitive ability, each study cited represents independent data. In cases in which longitudinal developmental data were presented, data from only one age-group are presented here. Only three studies were in this category, and in two of the three, the statistics were nearly identical at the different ages. In cases of cross-sectional developmental data, results from all appropriate ages were included. In addition, some studies included several different measures of the construct in question (e.g., verbal ability). In such cases, only one measure was included in the analysis in order to preserve the independence of the items in the data set. Only four studies were in this category, and one of them showed identical results for two measures. Of the remaining three studies, the measure showing intermediate results was reported.

It should be noted, however, that some studies are included in the analysis of two or more different abilities. For example, the study by Droege (1967) provided data on verbal ability, quantitative ability, and visual–spatial ability and thus is included in those three separate analyses.

### STATISTICAL ANALYSES

A number of measures of effect size are available (e.g., Craig, Eison, & Metze, 1976; Fleiss, 1969; Friedman, 1968; Hays, 1963; Tresemer, 1975). For each study, an attempt was made to compute two statistical measures of the magnitude of the gender difference. The first was $\omega^2$ (Hays, 1963), which measures the proportion of the total variance in the population that is accounted for by gender differences. This measure is advocated by Fleiss (1969); see also Craig et al., 1976 for estimating the magnitude of effects. Its meaning is similar to the $r^2$ (proportion of variance accounted for) statistic in correlation and regression. Just as $r^2$ gives the proportion of variance in the criterion (Y) explained by the predictor (X), so $\omega^2$ gives the proportion of variance in the entire distribution of scores that can be accounted for by gender differences (the remaining variance being due to within-gender variation and error of measurement). In the commonest cases, in which a $t$ value or group means and standard deviations were presented, $\omega^2$ was calculated using the following formula:

$$\omega^2 = \frac{t^2 - 1}{t^2 + n_1 + n_2 - 1}.$$

In the case of a two-factor design, the following formula was used:

$$\omega^2 = \frac{SS \text{ gender} - (c - 1)\, MS \text{ error}}{MS \text{ error} + SS \text{ total}},$$

where $c$ = the number of levels of the gender factor, that is, $c = 2$ (see Hays, 1963, p. 407). In the case of $t < 1$ or $F < 1$, $\omega^2$ was set equal to 0. The statistic $d$ is defined as the ratio of the difference between group means to the standard deviation of a group (the standard deviations of the two groups are assumed to be equal). It has been recommended for use in meta-analyses (e.g., Glass, 1976). In the present study, the standard deviation was defined as the average of the standard deviation for males and the standard deviation for females. Since the statistic assumes the standard deviations to be equal, the best estimate of that standard deviation is the average of the standard deviations of the two samples, male and female. Thus,

$$d = \frac{M_1 - M_2}{SD}.$$

The sign of $d$ was set to be positive if the study yielded findings in the direction of the "well-established" difference and to be negative if the findings were in the opposite direction. For purposes of interpretation, Cohen (1969) considers a $d$ of .20 small, a value of .50 medium, and a value of .80 large.

In cases in which the original reference provided insufficient information to compute either of these statistics (e.g., if only a significance level was reported), a letter was sent to the author at the address provided in the original paper or to a more recent address if provided by the APA Directory. The letter requested means and standard deviations for males and females and the $t$ or $F$ statistics. It was necessary to write such letters for 18 of the 53 different studies analyzed here. Thus, 34% of the studies provided insufficient information for a standard meta-analysis. Only seven responses to letters were received, and only two of these were able to provide the information requested. The remaining studies are listed as having no available $\omega^2$ and $d$ values in the tables.

## Results

### VERBAL ABILITY

The results for the 27 studies of verbal ability are shown in Table 1. As can be seen in that table, for those studies for which it was possible to compute $\omega^2$ and $d$, both values tended to be low. The median value of $\omega^2$ for verbal ability was .01; that is, gender differences accounted for only about 1% of the variance in verbal ability. The median value of $d$ was .24. That is, the means for males and females were about one fourth of a standard deviation apart.

### QUANTITATIVE ABILITY

The results from the 16 studies of quantitative ability are shown in Table 2. Once again, the values of $\omega^2$ and $d$ tended to be low. The median value of $\omega^2$ was .01 for quantitative ability. The median value of $d$ was .43.

### VISUAL–SPATIAL ABILITY

The results from the 10 studies of visual–spatial ability are shown in Table 3. Three of those studies provided data for different samples at two different age groups. Hence, Table 3 can be considered to show 13 sets of data. The median value of $\omega^2$ was .043. The median value of $d$ was .45.

### FIELD ARTICULATION (VISUAL–ANALYTIC SPATIAL ABILITY)

The results from the 20 studies of field articulation (typically measured by the Rod-and-Frame Test) are shown in Table 4. Two of the studies provided data for three separate age-group samples each, so that the number of data sets can be considered to be 24. The median value of $\omega^2$ was .025. The median value of $d$ was .51.

## Discussion

### MAGNITUDE OF GENDER DIFFERENCES

The main conclusion that can be reached from this analysis is that the gender differences in verbal ability, quantitative ability, visual–spatial ability, and field articulation reported by Maccoby and Jacklin (1974) are all small. Gender differences appear to account for no more than 1%–5% of the population variance. The difference in means is

TABLE 1
*Studies of Gender Differences in Verbal Ability*

| Study | Age | N | Difference | $\omega^2$ | d | Kind of sample |
|---|---|---|---|---|---|---|
| Achenbach (1969) | 11 | 159 | None | NA | NA | 5th and 6th graders |
| Cicirelli (1967) | 11 | 641 | None | NA | NA | 6th graders, suburban, middle-class, white |
| Weinberg & Rabinowitz (1970) | 12–19 | 48 | None | NA | NA | Hospital patients |
| Flanagan (Note 2) | 14, 17 | 1,545 | Females | NA | NA | Project TALENT data |
| Walberg (1969) | 16–17 | 2,074 | Females | .02 | .33 | High school physics students |
| Backman (1972) | 17 | 2,925 | Females | .33 | 1.40 | Project TALENT sample |
| American College Testing Program (1976–1977) | 18 | 45,222 | Females | .02 | .26 | College freshmen |
| Rosenberg & Sutton-Smith (1966) | 18–20 | 600 | None | NA | NA | College sophomores |
| Bieri et al. (1958) | 18–21 | 76 | None | 0 | .19 | Radcliffe women and Harvard men |
| DeFazio (1973) | 18–21 | 44 | None | 0 | NA | Selected for extreme scores on Hidden Figures Test |
| Feather (1968) | 18–21 | 60 | None | NA | NA | College students selected for extreme internal–external scores |
| Feather (1969) | 18–21 | 165 | None | NA | NA | College students |
| Koen (1966) | 18–21 | 72 | None | NA | NA | Harvard summer school |
| Laughlin (1969) | 18–21 | 528 | None | 0 | .03 | College students |
| Marks (1968) | 18 | 760 | None | NA | NA | Penn State freshmen |
| Mendelsohn & Griswold (1966) | 18–21 | 223 | None | 0 | .16 | College (Berkeley) |
| Mendelsohn & Griswold (1967) | 18–21 | 181 | None | 0 | NA | Berkeley undergraduates |
| Sarason & Minard (1962) | 18–21 | 96 | None | NA | NA | College |
| Very (1967) | 18–21 | 355 | Females | .04 | .41 | Penn State students |
| Rosenberg & Sutton-Smith (1964) | 19 | 377 | Females | .02[a] | NA | College students |
| Rosenberg & Sutton-Smith (1969) | 19 | 1,013 | Females | .04 | NA | College students |
| Sutton-Smith et al. (1968) | 19 | 1,055 | None | NA | NA | College students |
| Bayley & Oden (1955) | 29 | 1,102 | Males | .01 | .25 | Gifted adults and their spouses |
| Blum, Fosshage, & Jarvik (1972) | 84 | 54 | Females | .06 | .58 | Longitudinal development sample |
| Gates (1961) | 13 | 1,657 | Females | .01 | .22 | Schoolchildren |

*(table continues)*

Table 1 (*continued*)

| Study | Age | N | Difference | $\omega^2$ | d | Kind of sample |
|---|---|---|---|---|---|---|
| Droege (1967) | 18 | 6,167 | Females | .01 | .22 | Representative sample of high school students |
| Matarazzo (1972)[b] | 16–64 | 1,700 | Females | .003 | .12 | WAIS 1955 standardization sample |

*Note.* NA = not available (could not be computed from original paper, and author did not respond to request for further information or could not supply it). WAIS = Wechsler Adult Intelligence Scale.
[a] Percentage of variance accounted for was computed as $\phi^2 = \chi^2/N$.
[b] This study is not included in Maccoby and Jacklin's (1974) review; it is an additional large-scale study noted by Sherman (1978), included here because it is a major standardization study.

TABLE 2
*Studies of Gender Differences in Quantitative Ability*

| Study | Age | N | Difference | $\omega^2$ | d | Kind of sample |
|---|---|---|---|---|---|---|
| Hilton & Berglund (Note 1) | 16 | 539 | Males | NA | NA | College-bound students |
| Cicirelli (1967) | 11 | 641 | None | NA | NA | 6th graders, suburban, white |
| Keating & Stanley (1972) | 12–13 | 396 | Males | NA | NA | Junior high students exceptionally gifted in math |
| Droege (1967) | 18 | 6,167 | Males | <.01 | .06 | High school students |
| Flanagan (Note 2) | 14, 17 | 4,545 | Males | NA | NA | Project TALENT data |
| Walberg (1969) | 16–17 | 1,050 | Males | .01 | .43 | High school physics students |
| Backman (1972) | 17 | 2,925 | Males | .17 | .89 | Project TALENT sample—high school students |
| American College Testing Program (1976–1977) | 18 | 45,222 | Males | .03 | .35 | College freshmen |
| Rosenberg & Sutton-Smith (1964) | 19 | 377 | None | .007[a] | NA | College students |
| Rosenberg & Sutton-Smith (1969) | 19 | 1,013 | Males | .005 | NA | College students |
| Sutton-Smith et al. (1968) | 19 | 1,055 | None | NA | NA | College students |
| Rosenberg & Sutton-Smith (1966) | 19 | 600 | Males | NA | NA | College students |
| Bieri et al. (1958) | 18–21 | 76 | Males | .11 | .72 | Radcliffe women, Harvard men |
| Jacobson et al. (1970) | 18–21 | 136 | None | .01 | .27 | College students |
| Sarason & Minard (1962) | 18–21 | 96 | None | NA | NA | College students |
| Very (1967) | 18–21 | 355 | Males | .06 | .59 | College students |

*Note.* NA = not available.

only about one fourth to one half of a standard deviation (see Figures 1 and 2). Generally, it seems that gender differences in verbal ability are smaller and gender differences in spatial ability are larger, but even in the latter case, gender differences account for less than 5% of the population variance.

TABLE 3
*Studies of Gender Differences in Visual-Spatial Ability*

| Study | Age | N | Difference | $\omega^2$ | d | Kind of sample |
|---|---|---|---|---|---|---|
| Nash (1975) | 11–12 | 105 | None | 0 | .04 | 6th and 9th graders, New York public schools |
| | 14–15 | 102 | Males | .04 | .48 | |
| Stafford (1961) | 14–17 | 128 | Males | .07 | .60 | High school students |
| Droege (1967) | 18 | 6,167 | Males | .04 | .41 | High school seniors |
| Flanagan et al. (Note 2) | 14, 17 | 4,545 | Males | NA | NA | Project TALENT data |
| Backman (1972) | 17 | 2,925 | Males | .14 | .83 | Project TALENT sample—high school students |
| Brissett & Nowicki (1973) | 18–21 | 80 | None | 0 | — | College students |
| Kidd & Cherymisin (1965) | 18–21 | 100 | Males | .045 | NA | College students |
| Very (1967) | 18–21 | 355 | Males | .06 | .52 | College students |
| Davies (1965) | 20–29 | 90 | Males | .57 | NA | Adult volunteers |
| | 30–39 | 90 | Males | .13 | NA | |
| Sherman (1978)[a] | 15 | 1,233 | NA | .01 | .25 | High school math students |
| | 16 | | | .03 | .37 | |

*Note.* NA = not available.
[a] This study was not included in Maccoby and Jacklin's (1974) review.

PRACTICAL IMPLICATIONS

What are the practical implications of the finding that these cognitive gender differences, although statistically reliable and replicable, are small? Two areas of practical concern are discussed below: vocational counseling and explanation of observed gender differences in participation in certain occupations.

As noted in the introduction, there may be a temptation to use the findings of "well-established" gender differences in vocational counseling. The present analysis shows that gender is a poor predictor of one's performance on ability tests in any of these areas. Thus, presumably, gender would also be a poor predictor of performance on jobs requiring these abilities. Because the gender differences are small, one might question the wisdom of mentioning such differences in introductory psychology or child development texts that will be read by students who will eventually become teachers and guidance counselors and who may be prone to make the error of applying the findings of gender differences in their own counseling. At the very least, it would be important to stress how small the differences are and how unwise it would be to apply them in counseling a given individual.

The findings of gender differences in cognition have also been used in an explanatory way, in particular to account for gender segregation in some occupations. For example, only about 1% of engineers in the United States are women (Bird, 1968). Assuming that engineering requires a high level of spatial ability, can the gender difference in spatial ability account for the relative absence of women in this profession? The above findings of such a small gender difference would appear to argue that the answer is no. However, the question has now shifted from a discussion of overall mean differences in the population to differences at the upper end of the distribution. And relatively small mean differences can generate rather large differences at the tails of distributions, as the following sample calculation will show. Assume, conservatively, that the gender difference in spatial ability is .40 *SD*. Using *z* scores, the mean score for males will be .20 and the mean for females will be −.20. Assume also that being a successful engineer requires spatial ability at least at the 95th percentile for the population. A continuation of the *z*-score computation shows that about 7.35% of males will be above this cutoff, whereas only 3.22% of females will be. This amounts to about a 2:1 ratio of males to females with sufficient ability for the profession. This could, therefore, generate a rather large difference although certainly not as large a one as the existing one.

The disparity would become even larger if one considered some occupational feat, such as winning a Nobel prize or a Pulitzer prize, that would

TABLE 4
*Studies of Gender Differences in Field Articulation*

| Study | Age | N | Difference | $\omega^2$ | d | Kind of sample |
|---|---|---|---|---|---|---|
| Nash (1975) | 14 | 102 | Males | NA | NA | 9th graders, New York public schools |
| Weinberg & Rabinowitz (1970) | 12–19 | 48 | None | NA | NA | Hospital patients |
| Fiebert (1967) | 12 | 30 | None | .02 | .49 | Deaf subjects |
| | 15 | 30 | None | 0 | .30 | |
| | 18 | 30 | Males | .06 | .67 | |
| Silverman et al. (1973) | 18–22 | 30 | Males | NA | NA | College students |
| Gross (1959) | 18–25 | 140 | Males | .22 | 1.16 | College students |
| Stuart et al. (1965) | 18–21 | 64 | None | 0 | NA | College students and some noncollege high school graduates |
| Gruen (1955) | 19–35 | 60 | Males | .11 | .77 | Trained dancers |
| Schwartz & Karp (1967) | 17 | 46 | Males | .09 | .85 | Paid volunteers |
| | 30–39 | 40 | Males | .09 | .71 | |
| | 58–80 | 34 | None | .03 | .51 | |
| Bieri (1960) | 18–21 | 60 | None | NA | NA | College students |
| Bieri et al. (1958) | 18–21 | 76 | Males | NA | NA | Radcliffe women and Harvard men |
| Bogo et al. (1970) | 18–21 | 97 | Males | .12 | .78 | College students |
| Goldstein & Chance (1965) | 18–21 | 26 | None | 0 | NA | College students |
| Morf et al. (1971) | 18–24 | 78 | Males | .025 | .40 | College students |
| Morf & Howitt (1970) | 18–27 | 44 | None | 0 | .27 | College students |
| Oltman (1968) | 18–21 | 163 | None | 0 | .17 | College students |
| Sarason & Minard (1962) | 18–21 | 96 | Males | NA | NA | College students |
| Vaught (1965) | 18–21 | 180 | Males | .04 | NA | College students |
| Willoughby (1967) | 18–21 | 76 | None | 0 | .18 | College students |
| Gerace & Caldwell (1971) | 25 | 40 | Males | .08 | NA | Adult volunteers |
| Blum et al. (1972) | 64 | 43 | None | .02 | −.46 | Aging twins (females scored higher) |

*Note.* NA = not available.

require even higher levels of the ability. For example, suppose that spatial ability at the 99.5th percentile is now required. The same $z$-score calculations indicate that .85563% of males and .27375% of females would be above that cutoff, for an approximate 3:1 ratio of males to females. Once again, though, this is not nearly a large enough difference to account for the small proportions of women winning Nobel prizes. (The same computations could be used, recognizing the female superiority in verbal ability, to account for gender differences in the winning of Pulitzer prizes; once again, the gender difference in abilities would not, by itself, account for the observed difference.)

In sum, even small average gender difference

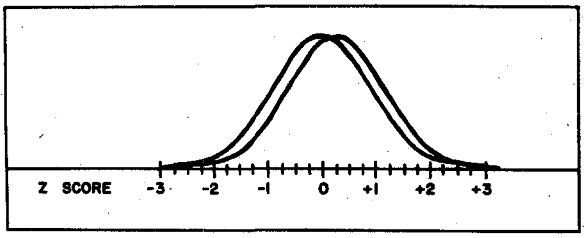


Figure 1. Two normal distributions with means .25 *SD* apart, that is, with a *d* of .25. This is approximately the magnitude of the gender difference in verbal ability.

in abilities can generate rather large differences in proportions of males and females above some high cutoff point that might be required for outstanding performance of some occupation. Nonetheless, the known differences in abilities are still too small to explain the observed occupational differences.

SAMPLING

In evaluating the existence and magnitude of cognitive gender differences, one important methodological issue deserves further discussion: sampling. It seems reasonable to define a cognitive gender difference as existing at a particular age if it is found to exist in a random sample of males and females of that age from a given culture (in this case, the United States). This definition is in contrast to that of Sherman (1978), who argued that gender differences are best studied on samples matched for prior relevant experience. For example, she has studied visual–spatial ability and quantitative ability in males and females who had taken the same high school math courses. The problem with this approach is that it may involve different percentages of the male and female populations and different cutoff points in their distributions. Essentially it confuses research on the existence of a gender difference with research on the cause of it. It would seem important first to establish the existence of a gender difference in random samples of the population, and then, if it exists, proceed to analyze its causes, whether biological or cultural or both.

With this definition of sampling in mind, brief descriptions of the samples involved in the gender differences studies can be found in Tables 1, 2, 3, and 4. Some of the deviations from ideal sampling are striking. For example, the study of verbal ability by Bieri, Bradburn, and Galinsky (1958) used as subjects Radcliffe female and Harvard male undergraduates. These subjects have surely already been selected for high levels of verbal ability, and the finding of no significant gender difference in the study is not at all surprising. Other obvious examples of problematic sampling occur in a study of verbal ability using Harvard summer school students (Koen, 1966), in another study of verbal ability using Berkeley undergraduates (Mendelsohn & Griswold, 1967), and in another study of verbal ability using high school physics students (Walberg, 1969). Finally, the study of quantitative ability by Keating and Stanley (1972) used junior high students who were exceptionally gifted in math and who initially were self-selected; doubtless, girls are less likely to self-select for high ability, particularly in a gender-inappropriate area such as math. Thus, many talented girls were probably lost from the sample. In defense of these studies, none was designed with the goal of normatively assessing the existence and magnitude of gender differences. Nonetheless, they have counted in the major tallies.

Indeed, the tough-minded might argue that all studies of college students are rendered irrelevant because the subjects have already had some prior selection on the basis of their cognitive abilities. If this rule is followed, few studies are left on the basis of which one can assess cognitive gender differences.

The more general point is that if one is to assess the existence and magnitude of cognitive gender differences, careful attention must be paid to sampling, and this has not been done in prior research or reviews.

A second point about sampling also needs to be raised. Some of the best studies from the point of view of random sampling are those that involve representative sampling of all high school students or all students in a given district. Such studies often involve enormous sample sizes. For example, the study by Droege (1967), which included measures of verbal ability, quantitative ability, and spatial ability, had a sample size of 6,167. Such a sample

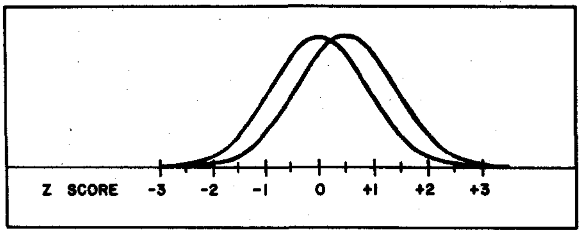


Figure 2. Two normal distributions with means .50 *SD* apart, that is, with a *d* of .50. This is approximately the magnitude of the gender differences in quantitative ability, visual–spatial ability, and field articulation.

size increases the chance of finding a highly reliable gender difference that is very small and accounts for very little variance. Such sample sizes seldom appear in research in other meta-analyses (e.g., Hall, 1978; Smith, 1980). It seems that research on cognitive gender differences is particularly prone to the problems of enormous samples because cognitive gender differences can so easily be studied in large-scale research, such as standardization of tests. Thus, research on cognitive gender differences seems particularly prone to having large sample sizes that produce reliable but tiny differences.

Finally, the present findings may help to explain the lack of replicability in this area. If the true difference in population means for males and females is small, then one would expect that in repeated samplings, many would find no significant gender differences. This is precisely what one sees in Maccoby and Jacklin's tables. Further, the present findings explain why theories designed to explain the differences (e.g., brain lateralization theories) have met with little success and produce such conflicting data (Sherman, 1978). A small phenomenon will produce conflicting data in repeated sampling.

## RECOMMENDATIONS

One problem that occurred in the process of doing the present analyses was the relatively high proportion of studies that provided insufficient information for $\omega^2$ and $d$ to be calculated. To facilitate more systematic assessment of gender differences, some journals have adopted the policy of requiring authors to report results of analyses of gender differences. However, if such analyses are to be complete, the magnitude of the differences also needs to be assessed. Based on the poor yield in response to letters requesting complete information for the present study, it seems important that such data be included in the original paper. Thus, a recommendation that emerges from this study is that journals encourage authors either to report $\omega^2$ and/or $d$ values along with their analyses of gender differences or to report sufficient information (means and standard deviations for males and females and/or $t$ and $F$ statistics) so that $\omega^2$ and $d$ can be calculated. Such policies would seem to be important in order to avoid the assumption that gender differences that are well-established are also large, an assumption that the present review has shown to be false.

A more radical recommendation is also suggested by the present study. The results indicate that it is possible to have a moderately reliable psychological phenomenon (e.g., one that appears in 50% or more of published studies on the topic) that is nonetheless small. How many other areas of psychological research and theory are based on small effects that are assumed to be large and reliable because the size of the effect has never been evaluated? The radical recommendation, then, is that journals require reports of effect size such as $\omega^2$ or $d$ with all reports of significant results. Of course, a small effect might still be an important one. But at least the reader would have the option of deciding whether a significant effect was large enough to merit further attention, either in teaching or in research.

## REFERENCE NOTES


1. Hilton, T. L., & Berglund, G. W. *Sex differences in mathematics achievement—a longitudinal study.* (Educational Testing Service Research Bulletin). Available from Educational Testing Service, Princeton, N.J.
2. Flanagan, J. C., et al. *Counselor's technical manual for interpreting test scores* (Project TALENT). Palo Alto, Calif., 1961.


## REFERENCES


Achenbach, T. M. Cue learning, associative responding, and school performance in children. *Developmental Psychology*, 1969, *1*, 717–725.

American College Testing Program. *College student profiles: Norms for the ACT assessment.* Iowa City: ACT Publications, 1976–1977.

Backman, M. E. Patterns of mental abilities: Ethnic, socioeconomic, and sex differences. *American Educational Research Journal*, 1972, *9*, 1–12.

Bayley, N., & Oden, M. The maintenance of intellectual ability in gifted adults. *Journal of Gerontology*, 1955, *10*, 91–107.

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

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

Bird, C. *Born female.* New York: McKay, 1968.

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

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

Brissett, M., & Nowicki, S., Jr. Internal versus external control of reinforcement and reaction to frustration. *Journal of Personality and Social Psychology*, 1973, *25*, 35–44.

Cicirelli, V. G. Sibling constellation, creativity, IQ, and academic achievement. *Child Development*, 1967, *38*, 481–490.

Cohen, J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press, 1969.

Craig, J. R., Eison, C. L., & Metze, L. P. Significance tests and their interpretation: An example utilizing published research and $\omega^2$. *Bulletin of the Psychonomic Society*, 1976, *7*, 280–282.

Davies, A. D. The perceptual maze test in a normal population. *Perceptual and Motor Skills*, 1965, *20*, 287–293.

DeFazio, V. J. Field articulation differences in language abilities. *Journal of Personality and Social Psychology*, 1973, *25*, 351–356.
Droege, R. C. Sex differences in aptitude maturation during high school. *Journal of Counseling Psychology*, 1967, *14*, 407–411.
Feather, N. T. Change in confidence following success or failure as a predictor of subsequent performance. *Journal of Personality and Social Psychology*, 1968, *9*, 38–46.
Feather, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology*, 1969, *13*, 129–144.
Fiebert, M. Cognitive styles in the deaf. *Perceptual and Motor Skills*, 1967, *24*, 319–329.
Fleiss, J. L. Estimating the magnitude of experimental effects. *Psychological Bulletin*, 1969, *72*, 273–276.
Friedman, H. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 1968, *70*, 245–251.
Gates, A. I. Sex differences in reading ability. *Elementary School Journal*, 1961, *61*, 431–434.
Gerace, T. A., & Caldwell, W. E. Perceptual distortion as a function of stimulus objects, sex, naivete, and trials using a portable model of the Ames distorted room. *Genetic Psychology Monographs*, 1971, *84*, 3–33.
Glass, G. V. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 1976, *10*, 3–8.
Goldstein, A. G., & Chance, J. E. Effects of practice on sex-related differences in performance on Embedded Figures. *Psychonomic Science*, 1965, *3*, 361–362.
Gross, F. The role of set in perception of the upright. *Journal of Personality*, 1959, *27*, 95–103.
Gruen, A. The relation of dancing experience and personality to perception. *Psychological Monographs*, 1955, *69*(14, Whole No. 399).
Hall, J. A. Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 1978, *85*, 845–857.
Hays, W. L. *Statistics*. New York: Holt, Rinehart & Winston, 1963.
Jacobson, L. I., Berger, S. E., & Milham, J. Individual differences in cheating during a temptation period when confronting failure. *Journal of Personality and Social Psychology*, 1970, *15*, 48–56.
Keating, D. P., & Stanley, J. C. Extreme measures for the exceptionally gifted in mathematics and science. *Educational Researcher*, 1972, *1*(9), 3–7.
Kidd, A. H., & Cherymisin, D. G. Figure reversal as related to specific personality variables. *Perceptual and Motor Skills*, 1965, *20*, 1175–1176.
Koen, F. Codability of complex stimuli: Three modes of representation. *Journal of Personality and Social Psychology*, 1966, *3*, 435–441.
Kulik, J. A., Kulik, C. C., & Cohen, P. A. A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist*, 1979, *34*, 307–318.
Laughlin, P. R., Branch, L. G., & Johnson, H. H. Individual versus triadic performance on a unidimensional complementary task as a function of initial ability level. *Journal of Personality and Social Psychology*, 1969, *12*, 144–150.
Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.
Marks, E. Personality factors in the performance of a perceptual recognition task under competing incentives. *Journal of Personality and Social Psychology*, 1968, *8*, 69–74.
Matarazzo, J. D. *Wechsler's measurement and appraisal of adult intelligence* (5th ed). Baltimore, Md.: Williams & Wilkins, 1972.
Mendelsohn, G. A., & Griswold, B. B. Assessed creative potential, vocabulary level, and sex as predictors of the use of incidental cues in verbal problem solving. *Journal of Personality and Social Psychology*, 1966, *4*, 423–431.
Mendelsohn, G. A., & Griswold, B. B. Anxiety and repression as predictors of the use of incidental cues in problem solving. *Journal of Personality and Social Psychology*, 1967, *6*, 353–359.
Morf, M. E., & Howitt, R. Rod-and-Frame Test performance as a function of momentary arousal. *Perceptual and Motor Skills*, 1970, *31*, 703–708.
Morf, M. E., Kavanaugh, R. D., & McConville, M. Intratest and sex differences on a portable rod-and-frame test. *Perceptual and Motor Skills*, 1971, *32*, 727–733.
Nash, S. C. *Conceptions and concomitants of sex-role stereotyping*. Unpublished doctoral dissertation, Columbia University, 1973.
Nash, S. C. The relationship among sex-role stereotyping, sex-role preference, and the sex difference in spatial visualization. *Sex Roles*, 1975, *1*, 15–32.
Oltman, P. R. A portable rod-and-frame apparatus. *Perceptual and Motor Skills*, 1968, *26*, 503–506.
Rosenberg, B. G., & Sutton-Smith, B. The relationship of ordinal position and sibling sex status to cognitive abilities. *Psychonomic Science*, 1964, *1*, 81–82.
Rosenberg, B. G., & Sutton-Smith, B. Sibling association, family size, and cognitive abilities. *Journal of Genetic Psychology*, 1966, *109*, 271–279.
Rosenberg, B. G., & Sutton-Smith, B. Sibling age spacing effects upon cognition. *Developmental Psychology*, 1969, *1*, 661–668.
Sarason, I. G., & Minard, J. Test anxiety, experimental instructions, and the Wechsler Adult Intelligence Scale. *Journal of Educational Psychology*, 1962, *53*, 299–302.
Schwartz, D. W., & Karp, S. Field dependence in a geriatric population. *Perceptual and Motor Skills*, 1967, *24*, 495–504.
Sherman, J. *Sex-related cognitive differences*. Springfield, Ill.: Charles C Thomas, 1978.
Silverman, J., Buchsbaum, M., & Stierlin, H. Sex differences in perceptual differentiation and stimulus intensity control. *Journal of Personality and Social Psychology*, 1973, *25*, 309–318.
Smith, M. L. Sex bias in counseling and psychotherapy. *Psychological Bulletin*, 1980, *87*, 392–407.
Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 1977, *32*, 752–760.
Stafford, R. E. Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and Motor Skills*, 1961, *13*, 428.
Stuart, I. R., Breslow, A., Brechner, S., Ilyus, R. B., & Wolpoff, M. The question of constitutional influence on perceptual style. *Perceptual and Motor Skills*, 1965, *20*, 419–420.
Sutton-Smith, B., Rosenberg, B. G., & Landy, F. Father-absence effects in families of different sibling compositions. *Child Development*, 1968, *39*, 1213–1221.
Tresemer, D. Measuring "sex differences." *Sociological Inquiry*, 1975, *45*, 29–32.
Vaught, G. M. The relationship of role identification and ego strength to sex differences in the Rod-and-Frame Test. *Journal of Personality*, 1965, *33*, 271–283.
Very, P. S. Differential factor structures in mathematical ability. *Genetic Psychology Monographs*, 1967, *75*, 169–207.
Walberg, H. J. Physics, femininity, and creativity. *Developmental Psychology*, 1969, *1*, 47–54.
Weinberg, S., & Rabinowitz, J. A sex difference in the Wechsler IQ Vocabulary scores as a predictor of strategy in a probability-learning task performed by adolescents. *Developmental Psychology*, 1970, *3*, 218–224.
Willoughby, R. H. Field-dependence and locus of control. *Perceptual and Motor Skills*, 1967, *24*, 671–672.