Psychological Bulletin
1994, Vol. 116, No. 3, 429–456

0033-2909/94/$3.00

Gender Differences in Personality: A Meta-Analysis

Alan Feingold

Four meta-analyses were conducted to examine gender differences in personality in the literature (1958–1992) and in normative data for well-known personality inventories (1940–1992). Males were found to be more assertive and had slightly higher self-esteem than females. Females were higher than males in extraversion, anxiety, trust, and, especially, tender-mindedness (e.g., nurturance). There were no noteworthy sex differences in social anxiety, impulsiveness, activity, ideas (e.g., reflectiveness), locus of control, and orderliness. Gender differences in personality traits were generally constant across ages, years of data collection, educational levels, and nations.

Do men and women differ in personality characteristics? Research on gender differences was begun by scientists who believed that individual differences in traits were biologically determined, and that findings of gender differences supported their view (Fausto-Sterling, 1985; Feingold, 1992d; Shields, 1975). Contemporary research on gender differences has focused on cognitive abilities (for recent reviews, see Feingold, 1993a; Linn, 1992; Wilder & Powell, 1989) and social behavior (e.g., Feingold, 1988b, 1992c; see also review by Eagly, 1987), including mate selection preferences (e.g., Buss, 1989; Feingold, 1990, 1991, 1992b). Yet, gender differences in personality (operationally defined as "what personality scales measure"[1]) are as worthy of inquiry as cognitive gender differences (and may be a proximal cause of gender differences in social behavior), but they have received scant attention.

Interest in gender differences in personality initially followed the formulation of theories of personality and the development of methods of assessment. *Personality inventories* are self-report, pencil-and-paper questionnaires on which respondents report their own feelings and behaviors, generally by responding to a series of yes-or-no or true-or-false test items. Personality inventories yield one or more scores that measure such traits as assertiveness, extraversion, and anxiety, and gender differences in personality traits were first examined by psychometricians to determine whether separate norms were needed for males and females.

Past Reviews of Gender Differences in Personality

Discussions of gender differences of any kind often begin with the conclusions from Maccoby and Jacklin's (1974) landmark review of sex differences in cognition, temperament, and social behavior. Maccoby and Jacklin used the formerly popular *narrative method* of review: Studies were grouped by area, the significance or nonsignificance of each sex difference was noted by study, and conclusions were drawn subjectively from both the number and the consistency of significant gender differences. Maccoby and Jacklin's review of temperamental gender differences—which mixed studies that used personality inventories with studies that measured behaviors thought to reflect personality traits—found males to be more assertive (dominant), more aggressive, and less anxious than females. No sex difference was found for self-esteem. Gender differences in locus of control were concluded to vary by age, with a gender difference (greater male internality) emerging only in the college years.

Shortly after the Maccoby and Jacklin (1974) review appeared, Glass (1976; Glass, McGaw, & Smith, 1981) popularized the use of meta-analysis, or quantitative methods, to cumulate research findings. Meta-analysis consists of computing an effect size for each between-groups difference, categorizing effect sizes by domain (e.g., personality trait), and quantitatively combining and comparing effect sizes across studies by domain (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Mullen, 1989; Rosenthal, 1991). The usefulness of meta-analysis in summarizing findings of sex differences was quickly recognized (e.g., Cooper, 1979; Eagly & Carli, 1981; Hall, 1978), and it has become the most popular method of reviewing gender differences (see reviews by Eagly, in press; Eagly & Wood, 1991; Feingold, 1993a; Hyde & Linn, 1986; Linn & Hyde, 1989).

This article is based, in part, on a dissertation submitted to Yale University in partial fulfillment of the requirements for the doctorate of philosophy degree in psychology.

I want to thank my thesis advisor, Jerome L. Singer, and the other members of my dissertation committee—Alan Kazdin, Bonnie Leadbeater, William McGuire, and Peter Salovey—for their contributions to the research. I want to thank Ronald Mazzella, who served as a rater in the categorization of personality scales, and Alice Eagly, Janet Hyde, Carol Jacklin, and an anonymous reviewer for their comments on drafts of this article. Finally, I am grateful to Lisa Lee (Educational and Industrial Testing Service), Mark Reike (Institute for Personality and Ability Testing), and Robert Tett (Research Psychologists Press) for providing me with unpublished, out-of-print, and in-press normative data on the Revised Eysenck Personality Questionnaire, the Sixteen Personality Factor Questionnaire, and the Personality Research Form, respectively.

Correspondence concerning this article should be addressed to Alan Feingold, Department of Psychiatry, Yale University, Substance Abuse Center, 34 Park Street, New Haven, Connecticut 06519.

[1] This definition may be extended to include behavioral measures that are reflective of traits measured by personality scales (e.g., behavioral displays of assertiveness are indicants of the personality trait of assertiveness).

Gender differences in personality have been subjected to only limited meta-analytic scrutiny. Meta-analysts (e.g., Eagly & Steffen, 1986; Hyde, 1984) have examined the findings of sex differences in aggression and confirmed Maccoby and Jacklin's (1974) conclusion of greater male aggressiveness. Meta-analysis has also found that females score higher than males on ego development but that the advantage fades with age (Cohn, 1991), which suggests that the sex difference may be a result of earlier female maturation in ego development.

With the exception of Hyde's meta-analysis, findings of sex differences in personality from research conducted in the 1960s and early 1970s (i.e., the general period reviewed by Maccoby and Jacklin, 1974) have not been examined quantitatively. However, Hall (1984) conducted a meta-analysis of findings from later (1975–1983) research by retrieving studies from four journals (*Journal of Personality, Journal of Personality and Social Psychology, Journal of Personality Assessment,* and *Sex Roles*) and quantitatively combining sex differences for several personality dimensions, including traits examined narratively by Maccoby and Jacklin. Hall found that there was essentially no sex difference in either self-esteem or assertiveness but that females were more anxious and less internally controlled than males, although the effect sizes were small for both of these gender differences.

Theoretical Issues

Why might men and women score differently on traits found in standardized personality inventories? At least three models—biological, sociocultural, and biosocial—address the proximal causes of sex differences.

The biological model posits that observed gender differences in personality test scores reflect innate temperamental differences between the sexes. Contemporary research has suggested that there is a strong biological basis underlying individual differences in personality traits. Much of this work has consisted of (a) twins studies on the heritability of personality traits and (b) studies correlating personality traits with hormonal–chemical substances or physiological measures (Eysenck, 1992; Zuckerman, 1991). Zuckerman has suggested that gender differences in the traits of dominance and aggression may be caused by biological sex differences in gonadal hormones.

It has also been hypothesized that sex differences in chromosomes may make women more prone to depression than men (Nolen-Hoeksema, 1987). Women have two X chromosomes, in comparison with one for men, and major affective illnesses may be caused by a mutant gene on the X chromosome (Perris, 1966; Winokur & Tanna, 1969). A greater female vulnerability to depression would be manifested in higher scores for women than for men on measures of depression, anxiety, and neuroticism.

The sociocultural model of gender differences posits that social and cultural factors directly produce gender differences in personality traits. Eagly (1987; Eagly & Wood, 1991) developed a *social role model* positing that sex differences in social behavior stem from gender roles, which dictate the behaviors that are appropriate for males and females. One's behavior may shape one's personality and one's responses to items in personality scales.

Another example of a sociocultural model is the *expectancy model,* which contends that social and cultural factors eventuate in gender stereotypes, which cause sex differences in personality because holders of stereotypical beliefs treat others in ways that result in others conforming to the prejudices of the perceivers. This would be a classic example of a self-fulfilling prophecy (Jussim, 1986; Miller & Turnbull, 1986). In a seminal experiment on the consequences of people's expectations, Rosenthal and Jacobson (1968) induced teachers to believe that some of their students were gifted; the designated students (who actually had been selected randomly) were later found to be more academically successful than other students. In everyday life, however, people's expectations of strangers come mainly from stereotypes, and it has been speculated that stereotype-based expectancies produce self-fulfilling prophecies (Ambady & Rosenthal, 1992; Deaux & Major, 1987; Feingold, 1992c; Hall & Briton, 1993; McArthur, 1982). Moreover, robust gender stereotypes have been documented (e.g., Swim, 1994; see also reviews by Ashmore, Del Boca, & Wohlers, 1986; Deaux & Kite, 1993; Feingold, 1993b; Ruble & Ruble, 1982).

How might stereotype-related expectancies produce differences in personality traits among groups, thereby confirming the perceived differences among groups? Self-concept may mediate expectancy outcomes (Darley & Fazio, 1980). If, for example, assertiveness is a trait seen to be characteristic of men, then people may respond to men in a manner that causes men to first internalize assertiveness as part of their self-concept (e.g., Cooley, 1900; Mead, 1934) and then to behave assertively to bring their behaviors in line with their self-image (Swann, 1984).

A third example of a sociocultural model is the *artifact model,* which affords an explanation of sex differences on personality scales rather than in the underlying personality constructs. Gender differences on personality scales are important because such findings are often interpreted as gender differences in personality traits (i.e., constructs). The artifact model posits that sociocultural factors (e.g., gender stereotyping) result in men and women holding different values about the importance of possessing various traits and that these differences differentially bias self-reports of personality characteristics, engendering sex differences in scores on personality inventory traits that do not reflect corresponding sex differences in the personality constructs that the tests purport to measure.

The artifact model, originally proposed by Feingold (1990, 1991, 1992b) as a possible explanation for the findings of sex differences in self-reported mate selection preferences, assumes that personality scales are not perfectly valid measures of their constructs. Women may view nurturance, for example, as a very positive characteristic, and social desirability–related biases may result in women reporting themselves to be more nurturant than they are. Men, by comparison, may have been inculcated with the belief that nurturant males are "wimps" or "sissies" and may underreport their level of nurturance. If so, the gender difference on a personality scale of nurturance would not reflect the gender difference in the nurturance construct. The artifact model would not apply to sex differences in behavioral

measures, such as helping behavior (e.g., Eagly & Crowley, 1986), that are presumed to reflect individual differences in personality traits. However, although sex differences in behavior may reflect differences in personality between men and women, they may also reflect differences in the way men and women are treated, or some combination of sex differences in personality and sex differences in social forces.

The sociocultural model is a theory of proximal rather than of distal causes of gender differences. Even if a pure sociocultural model is valid, biological or evolutionary-related factors may have shaped the sociocultural factors that directly produce gender differences in personality. If so, biology would be the distal cause of sex differences, and culture would be the proximal cause of sex differences. (By comparison, the biological model posits that biology is the proximal cause of sex differences.)

The hypothesis that gender differences have both proximal and distal causes is plausible because social roles, based mainly on distribution of work tasks, may have evolved in preindustrial times as a consequence of physical differences between the sexes that were far more consequential then than in the current technological age. These physical differences include greater male size and strength (which make males superior at hunting, building, and waging unsophisticated warfare) as well as anatomical differences relating to reproduction. Premodern times were characterized by shorter life spans, women bearing children more frequently (and at younger ages) and a lack of medical knowledge and facilities that often made the birthing process life threatening for mothers. Thus, traditional male and female work roles were functional in the preindustrial age and may have resulted in gender differences in personality that maximized males' and females' satisfaction with their respective social roles. Current gender differences may, then, be a consequence of sociocultural factors that are a vestige of bygone eras. Indeed, the biological model is predicated on the same evolutionary ideas, except that it posits that biological factors produce direct temperamental differences between the sexes.

That biological and sociocultural factors are both proximal causes of gender differences in personality is posited by a biosocial model. For example, if men and women are initially perceived differently because of observable male–female differences in behaviors that are linked to innate temperamental sex differences, men and women may be treated differently because of stereotypes that result from these differences in behavior. If social treatment also affects personality development, social factors may augment inherent gender differences. In genetics terminology, phenotypical gender differences may exceed corresponding genotypical gender differences, with the phenotypical sex differences a product of both biological and environmental factors.

Current Meta-Analyses of Gender Differences in Personality

The two studies described here examined gender differences in personality traits through meta-analysis. Study 1 first reexamined the sets of studies—which included research that used behavioral measures of traits as well as studies that used personality scales—that had been reviewed by Maccoby and Jacklin (1974) for gender differences in assertiveness, locus of control, self-esteem, and anxiety. Next, a direct replication of Hall's (1984) meta-analysis was conducted to review sex differences in the same traits using studies published recently in the same journals searched by Hall. Finally, the findings from these three meta-analyses were compared to determine the consistency of sex differences across time and reviewer methodology.

Researchers of cognitive gender differences have examined not only results in the literature but have also combined findings from normative data for standardized tests of cognitive ability (e.g., Feingold, 1988a; Marsh, 1989). Thus, Study 2 examined sex differences in traits in the norms for widely used personality inventories—as classified by Costa and McCrae's (1992) five-factor facet model of personality—and assessed variations in effect sizes across inventories, standardization years, ages, educational levels, and nations.

Study 1: Gender Differences in the Literature

Method

Meta-Analysis of the Maccoby–Jacklin (1974) Studies

Retrieval of studies. The studies used by Maccoby and Jacklin (1974) in their qualitative review of gender differences in self-esteem, internal locus of control, anxiety, and assertiveness were obtained and sorted by category.

Criteria for selection of studies. The first criterion for a retrieved study to be used in the meta-analysis was that an effect size for a gender difference could be calculated or estimated from the reported results (including levels of significance) for at least one relevant finding. Second, the two studies of behavioral (state or situational) anxiety were deleted to form a more homogeneous trait anxiety category.

Third, studies were not used in the meta-analysis when the inclusion of a study (or kinds of studies) by Maccoby and Jacklin in a particular category was questionable in terms of current practices. For example, the Machiavellianism scale (Christie & Geis, 1970) is not a generally accepted measure of assertiveness, and sex differences in Machiavellianism were excluded.

Finally, some of the reviewed studies had examined gender differences for two relevant personality dimensions. In a few cases, one of the two findings had been inadvertently omitted from review by Maccoby and Jacklin (1974). Such omitted results were, however, included in the meta-analysis when the other inclusion criteria were met.

Studies used in the meta-analysis. Maccoby and Jacklin (1974) tabled 30 studies of sex differences in self-esteem, and 23 of those studies were used in the meta-analysis.[2] Also used were

[2] Two studies (Amatora, 1955; Shrader & Leventhal, 1968) were excluded because self-esteem was assessed only by observers familiar with subjects (parents or teachers), and such ratings may have been contaminated by gender stereotypes (Ashmore, Del Boca, & Wohlers, 1986; Feingold, 1993b; Ruble & Ruble, 1982). Five other studies (Bortner & Hultsch, 1972; Goldrich, 1967; Lepper, 1973; Nawas, 1971; Schaie & Strother, 1968) that measured constructs (e.g., ego complexity, optimism, and happiness) not generally recognized as being synonymous with self-esteem were also deleted.

the studies by Feather (1969) and L'Abate (1960), each of which had examined self-esteem but was cited by Maccoby and Jacklin only for another gender difference. The 25 studies used 34 independent samples (N = 6,256; 50% male).

Maccoby and Jacklin (1974) reported sex differences in internal locus of control for 14 studies, 1 (Solomon, Houlihan, & Parelius, 1969) of which was excluded because it used the same data as an earlier study (Crandall, Katkovsky, & Crandall, 1965). Effect sizes were extracted from 21 independent samples from the remaining 13 studies (N = 2,234; 52% male).

Maccoby and Jacklin (1974) reported examinations of sex differences in anxiety from 20 studies, of which 17 were included in the meta-analysis.[3] Effect sizes were calculated from the 28 independent samples (N = 5,789; 52% male).

Maccoby and Jacklin (1974) reviewed 26 studies of gender differences in assertiveness, of which 17 were used in the meta-analysis.[4] In addition, although Maccoby and Jacklin did not include the studies by Schaie and Strother (1968) and Silverman, Shulman, and Wiesenthal (1970) in their assertiveness category, both studies had examined assertiveness and were included in the meta-analysis. Therefore, 19 studies that examined 22 independent samples (N = 5,015; 50% male) were used in the meta-analysis.

Calculation of effect sizes. When means and standard deviations of personality measures were reported separately by sex, the effect size (d) was obtained by dividing the difference between the male and female means by the pooled within-sex standard deviation (Cohen, 1977), with positive effect-size values indicating that males scored higher than females. Otherwise, the effect size was derived from the results of significance tests (t or F ratios) by use of the conversion formulas provided by Rosenthal (1991). Gender effects that were reported only as nonsignificant were assigned values of .00 for effect size (Rosenthal, 1991). Because sample sizes were sometimes very small, correction procedures described in Hedges and Olkin (1985) were routinely applied to effect sizes to yield unbiased estimates of their population values. (When sample sizes are not very small, corrected and uncorrected effect sizes are essentially identical.)

Coding of effect sizes. All effect sizes were coded for (a) mean age of sample and (b) operationalization (personality scale vs. behavioral measure). The mean age was used to dichotomize studies into two subgroups: studies of children (mean ages = 2–12 years) and studies of adolescents–adults (mean age = 13 years or older). The studies in the latter age subcategory were comparable in age to those used by Hall (1984) in her meta-analysis.

Meta-analysis of effect sizes. The meta-analysis used the methods developed by Hedges (e.g., Hedges & Olkin, 1985). First, the weighted mean effect size and its 95% confidence interval (CI) were calculated for each effect category. (A mean effect size is statistically significant when the CI does not include .00.) Next, the homogeneity of within-category effect sizes was examined for each category. Finally, because Maccoby and Jacklin (1974) reviewed studies that were heterogeneous in both operationalization and age, moderator variable analysis was conducted to examine whether effect sizes varied significantly with these study characteristics. Whereas some researchers assessed traits by personality scales (of main interest in this article), others measured behaviors (or self-reports of behaviors) reflective of traits (e.g., displays of assertiveness in small groups). Therefore, studies were dichotomized by operationalization (i.e., scales vs. behaviors), weighted mean effect sizes were calculated separately for each operationalization, and the significance of the differences between effect sizes ascribable to operationalization was assessed for all traits except anxiety (which was always measured by scales). Moderation of effect sizes by age (2–12 years vs. 13 years or older) was examined for all traits. These moderator variable analyses consisted of the partitioning of chi-square values reflecting heterogeneity among effect sizes into between-groups and within-groups sources (Hedges & Becker, 1986; Hedges & Olkin, 1985).

Replication of Hall's (1984) Meta-Analysis

Retrieval of studies. A manual search was conducted of the *Journal of Personality and Social Psychology,* the *Journal of Personality,* the *Journal of Personality Assessment,* and *Sex Roles* to obtain studies published from 1984 through 1992 that examined sex differences (or contained information, such as means and standard deviations of scales by sex, that afforded extraction of sex differences) in self-esteem, internal locus of control, anxiety, or assertiveness.

Criteria for selection of studies. To be consistent with Hall's (1984) inclusion criteria, only studies of clinically normal adolescents or adults were used in the meta-analysis. Hall developed an anxiety category that subsumed sex differences in measures of both general anxiety and social anxiety (also known as *shyness*) but—unlike Maccoby and Jacklin (1974)—excluded measures called neuroticism or depression, and the replication meta-analysis used the same categorization criteria. Following Hall, the assertiveness category subsumed sex differences in measures of assertiveness, dominance, and social poise. Sex differences on measures of self-esteem or self-concept constituted the self-esteem category. Only sex differences on measures explicitly labeled locus of control were used in the locus of control category.

Calculation of effect sizes. Hall (1984) expressed effect sizes in the r metric (the point-biserial correlation between traits and gender). The replication study, however, used the more common d metric, which—for relatively small effects—is about twice as large as the algebraically equivalent r (Rosenthal, 1991). Effect

[3] Although Maccoby and Jacklin (1974) included a study of death anxiety (Templer, Ruff, & Franks, 1971) in their general anxiety category, it was not used in the meta-analysis. As noted earlier, the two studies of state anxiety (Benton, Gelber, Kelley, & Liebling, 1969; MacDonald, 1970) were also excluded.

[4] Two studies (Anderson, 1939; Gellert, 1962) were not used in the meta-analysis because effect sizes could not be computed for them. The four studies that examined Machiavellianism (Braginsky, 1970; Christie, 1970a, 1970b; Nachamie, 1969) were also excluded, as was Omark, Omark, and Edelman's (1973) study examining "toughness" (because psychological rather than physical assertiveness was of interest). Finally, studies by Omark and Edelman (1973) and Bee (1967) were excluded because the behaviors measured (e.g., speaking first in a problem-solving interaction) could not be unambiguously classified as indicants of assertive behavior (as opposed to, for example, self-confident behavior).

sizes were calculated by the same methods used in the meta-analysis of the studies reviewed by Maccoby and Jacklin (1974).

Meta-analysis of effect sizes. Hall (1984) calculated unweighted mean effect sizes for each trait, and the same procedure was used in the replication.

Comparisons Across the Three Meta-Analyses

A comparison of findings from (a) the meta-analysis of the subgroup of the studies of adolescents and adults reviewed by Maccoby and Jacklin (1974), (b) Hall's (1984) meta-analysis, and (c) the current replication of Hall's meta-analysis—each of which reviewed studies from different periods—was conducted to examine temporal trends in personality gender differences in the literature. However, because Hall used the r metric instead of the d metric for effect sizes, the mean correlations she obtained were first transformed to algebraically equivalent ds by the conversion formula provided in Rosenthal (1991):

$$d = \frac{2r}{\sqrt{1 - r^2}}.$$

Interpretation of Effect Sizes

According to Cohen (1977), effect sizes of .20, .50, and .80 indicate small, medium, and large effects, respectively, and his criteria were used to assess the magnitude of gender differences. Effect sizes of .15–.19 were interpreted as very small, and effect sizes below .15 were interpreted as negligible (and not practically different from zero, regardless of statistical significance).

Results

Meta-Analysis of Studies Reviewed by Maccoby and Jacklin (1974)

Description of database. Sixty-eight studies that yielded findings from 105 independent samples (N = 17,729) were used in the meta-analysis. The numbers of effect sizes (samples) ranged from 21 to 34 across effect categories (traits), and the pooled sample sizes ranged from 2,234 to 6,256 across categories. Because only 4 studies (each of which had used a single sample) yielded findings that were included in more than one effect category, effect sizes were essentially independent across and within effect categories.

Table 1 lists the included studies of sex differences for the four traits and the study characteristics (age and method) used in the moderator variable analyses. (Because only findings from self-report anxiety measures were used in the meta-analysis, the name of the anxiety measure is listed for method in that category.)

Self-esteem. The meta-analysis found essentially no overall sex difference in self-esteem (median d = .00, weighted mean d = −.05, k = 34, N = 6,256). However, homogeneity of effect sizes was rejected, $\chi^2(33) = 65.88$, $p < .001$, and age of subjects was statistically associated with variation in effect sizes, $\chi^2(1) = 12.82$, $p < .001$. Although female children had higher self-esteem than male children (weighted mean d = −.11, CI = −.05 to −.17, k = 22, N = 4,544), male adolescents and adults had higher self-esteem than female adults and adolescents (weighted mean d = .10, CI = .00 to .19, k = 12, N = 1,712). However, because the absolute values of these effect sizes were both very small, there was essentially no gender difference for self-esteem in either age group. Effect sizes did not vary significantly with operationalization of self-esteem (weighted mean ds = .05 for behavioral measures and −.06 for personality scales), $\chi^2(1) = 1.82$.

Internal locus of control. The meta-analysis of all effect sizes found no overall gender difference in internal locus of control (median d = .00, weighted mean d = .01, k = 21, N = 2,229). However, homogeneity of effect sizes was rejected, $\chi^2(20) = 93.41$, $p < .001$, and much of the heterogeneity was accounted for by operationalization, $\chi^2(1) = 8.48$, $p < .01$. When behavioral measures were used, males were found to be more internally controlled than females (weighted mean d = .25, CI = .07 to .44, k = 5, N = 466), whereas there was no gender difference when internal control was measured by personality scales (weighted mean d = −.05, k = 16, N = 1,763). However, although the effect sizes in the behavioral subcategory were homogeneous, $\chi^2(4) = 6.85$, *ns*, there was significant heterogeneity among effect sizes in the personality scales subcategory, $\chi^2(15) = 78.08$, $p < .001$. Most of the studies in the latter subcategory measured internal locus of control with the Intellectual Achievement Responsibility scale (IAR; Crandall, Katkovsky, & Crandall, 1965) (of which only the gender differences in total scores were used in the meta-analysis), and females scored as more internally controlled than males on the IAR (weighted mean d = −.28, CI = −.16 to −.40, k = 11, N = 1,108). By comparison, males scored as more internally controlled than females on other internal locus of control scales (weighted mean d = .34, CI = .18 to .50, k = 5, N = 655). Effect sizes did not vary significantly with age of subjects, $\chi^2(1) = 2.76$.

Anxiety. The meta-analysis found that females scored higher than males, to a small degree, on measures of anxiety (median d = −.30, weighted mean d = −.29, CI = −.24 to −.34, k = 28, N = 5,789). Homogeneity of effect sizes was not rejected, $\chi^2(27) = 33.54$, *ns*, indicating that variation in effect sizes across studies could be attributable entirely to sampling error. The effect size was about the same for children (weighted mean d = −.24, CI = −.15 to −.33, k = 22, N = 1,901) as it was for adolescents and adults (weighted mean d = −.31, CI = −.25 to −.38, k = 6, N = 3,888), $\chi^2(1) = 1.73$, *ns*.

Assertiveness. The meta-analysis found males to be more assertive than females when effect sizes were weighted (mean d = .38, CI = .32 to .44, k = .22, N = 5,049). However, the corresponding median effect size of .00 indicated no gender difference in assertiveness. The large discrepancy between the weighted mean effect size and the median effect size was attributable to the inclusion of two studies (Baltes & Nesselroade, 1972; Strongman & Champness, 1968)—one with a large sample size—that yielded effect sizes that were outliers (see Table 1).[5] When the two outliers were deleted, the median effect size remained .00, but the weighted mean effect size was reduced to

[5] An outlier was defined as an effect size that diverged by more than two standard deviations from the unweighted mean effect size.

Table 1
Effect Sizes From Older Studies of Gender Differences in Self-Esteem, Internal Locus of Control, Anxiety, and Assertiveness

Study	Male *n*	Female *n*	*M* age (years)	Method	*d*
		Self-esteem			
Baumrind & Black (1967)	52	51	3.9	Behavior	.00
Bledsoe (1961)	101	96	10.5	Scale	−.28
Bledsoe (1967)	65	60	9.0	Scale	−.59
Bledsoe (1967)	76	70	11.0	Scale	−.62
Carlson (1965)	33	16	14.0	Scale	.00
Carpenter & Busse (1969)	10	10	6.9	Scale	.20
Carpenter & Busse (1969)	10	10	7.0	Scale	1.01
Carpenter & Busse (1969)	10	10	11.4	Scale	.73
Carpenter & Busse (1969)	10	10	11.5	Scale	.19
Coopersmith (1959)	44	43	11.0	Scale	−.14
Coopersmith (1967)	874	874	10.5	Scale	−.16
Feather (1969)	89	78	19.5	Scale	.30
Goldschmid (1968)	38	43	7.3	Scale	.00
Harris & Braun (1971)	30	30	7.5	Scale	.00
Herbert, Gelfand, & Hartmann (1969)	20	20	9.0	Scale	.41
Jacobson, Berger, & Millham (1969)	121	155	19.5	Behavior	.00
Kaplan (1973)	226	274	NA[a]	Scale	.12
Klaus & Gray (1968)	40	40	7.8	Scale	.00
Koenig (1966)	20	20	19.5	Behavior	.00
L'Abate (1960)	49	47	6.0	Scale	.23
Lekarczyk & Hill (1969)	63	51	10.8	Scale	.00
Long, Henderson, & Ziller (1967)	26	26	6.0	Scale	.05
Long et al. (1967)	26	26	7.0	Scale	−.24
Long et al. (1967)	26	26	8.0	Scale	−.62
Long et al. (1967)	26	26	9.0	Scale	.05
Long et al. (1967)	26	26	10.0	Scale	−.19
Long et al. (1967)	26	26	11.0	Scale	1.01
Nisbett & Gordon (1967)	76	76	19.5	Scale	.00
Sarason & Koenig (1965)	24	24	19.5	Behavior	−.59
Sarason & Winkel (1966)	24	24	19.5	Behavior	.29
Shipman (1971)	685	686	3.5	Scale	.00
Silverman, Shulman, & Wiesenthal (1970)	56	42	19.5	Scale	.00
Skolnick (1971)	57	57	19.5	Scale	.00
Zander, Fuller, & Armstrong (1972)	44	46	19.5	Behavior	.51
		Internal locus of control			
Benton, Gelber, Kelley, & Liebling (1969)	40	40	19.5	Behavior	.45
Brannigan & Tolor (1971)	205	128	19.5	Scale	.58
Buck & Austrin (1971)	25	25	13.0	Scale[b]	−.60
Buck & Austrin (1971)	25	25	13.0	Scale[b]	.36
Crandall, Katkovsky, & Crandall (1965)	44	58	8.0	Scale[b]	−.02
Crandall et al. (1965)	59	44	9.0	Scale[b]	.02
Crandall et al. (1965)	52	47	10.0	Scale[b]	−.08
Crandall et al. (1965)	93	73	11.0	Scale[b]	−.53
Crandall et al. (1965)	68	93	13.0	Scale[b]	−.34
Crandall et al. (1965)	90	93	15.0	Scale[b]	−.29
Crandall et al. (1965)	52	57	17.0	Scale[b]	−.88
Crandall & Lacey (1972)	28	22	9.6	Scale[b]	−.20
Dweck & Reppucci (1973)	20	20	10.0	Scale[b]	.00
Feather (1969)	89	78	19.5	Behavior	.51
Levy et al. (1972)	55	55	19.5	Scale	.00
MacMillan & Keogh (1971a)	30	30	11.8	Behavior	.00
MacMillan & Keogh (1971b)	60	60	8.0	Behavior	.00
Pallak, Brock, & Kiesler (1967)	19	20	19.5	Behavior	.00
Walls & Cox (1971)	20	20	8.5	Scale	.00
Walls & Cox (1971)	20	20	8.5	Scale	1.06
Zytkoskee, Strickland, & Watson (1971)	71	61	15.5	Scale	.00

Table 1 (*continued*)

Study	Male *n*	Female *n*	*M* age (years)	Method	*d*
		Anxiety			
Baltes & Nesselroade (1972)	625	624	14.0	HSPQ Tension	−.30
Barton (1971)	32	32	10.0	STAI	−.68
Cowen & Danset (1962)	66	66	9.0	CMAS	−.57
Cowen, Zax, Klein, Izzo, & Trost (1965)	83	86	9.0	CMAS	−.46
Goldschmid (1968)	38	43	7.3	CMAS	.00
Grams, Hafner, & Quast (1965)	47	63	10.0	GASC	−.32
Hafner & Kaplan (1959)	55	53	11.0	CMAS	−.10
Hanna, Storm, & Caird (1965)	1,154	804	18.5	MPI Neuroticism	−.30
Holloway (1958)	64	57	8.0	CMAS	−.06
Iwawaki, Sumida, Okuno, & Cowen (1967)	71	84	9.0	CMAS	−.04
Kidd & Cherymisin (1965)	50	50	19.0	TMAS	.00
L'Abate (1960)	49	47	11.0	CMAS	−.15
Lott & Lott (1968)	14	15	9.0	CMAS	−.18
Lott & Lott (1968)	43	48	9.0	CMAS	.20
Lott & Lott (1968)	15	10	10.0	CMAS	−.46
Lott & Lott (1968)	42	46	10.0	MCAS	−.04
Mendelsohn & Griswold (1967)	33	46	19.5	MMPI Anxiety	−.64
Mendelsohn & Griswold (1967)	42	60	19.5	MMPI Anxiety	−.35
Palermo (1959)	23	24	9.0	CMAS	−.06
Palermo (1959)	79	70	9.0	CMAS	−.11
Palermo (1959)	21	27	10.0	CMAS	−.19
Palermo (1959)	65	54	10.0	CMAS	−.41
Palermo (1959)	17	24	11.0	CMAS	−.67
Palermo (1959)	63	63	11.0	CMAS	−.20
Penney (1965)	17	17	9.0	CMAS	−.52
Penney (1965)	17	17	10.0	CMAS	−.46
Penney (1965)	17	17	11.0	CMAS	−.91
Vassiliou, Georgas, & Vassiliou (1967)	188	212	NA[a]	TMAS	−.43
		Assertiveness			
Anderson (1937)	47	47	4.0	Behavior	−.49
Arkoff, Meredith, & Iwahara (1962)	116	136	19.1	Scale	.44
Baltes & Nesselroade (1972)	625	624	14.0	Scale	1.52
Baumrind & Black (1967)	52	51	3.5	Behavior	.00
Denmark & Diggory (1966)	194	114	19.5	Behavior	.00
Emmerich (1971)	208	207	4.5	Behavior	.00
Emmerich (1971)	298	298	4.5	Behavior	.00
Feshback (1969)	65	61	6.0	Behavior	.00
Gardiner (1968)	46	153	19.5	Scale	.24
Harrison, Rawls, & Rawls (1971)	349	345	8.5	Behavior	.00
Markel, Prebor, & Brandt (1972)	36	36	19.5	Behavior	.66
Parten (1933)	17	17	3.0	Behavior	.00
Schaie & Strother (1968)	25	25	76.0	Scale	.75
Sharma (1969)	165	128	19.0	Scale	.06
Silverman et al. (1970)	56	42	19.5	Scale	−.11
Strongman & Champness (1968)	5	5	19.5	Behavior	−1.16
Sutton-Smith & Savasta (1972)	8	9	3.5	Behavior	.00
Szal (1972)	30	39	4.5	Behavior	.00
Whiting & Edwards (1973)	34	33	4.5	Behavior	.50
Whiting & Edwards (1973)	33	34	9.0	Behavior	.00
Zander & Van Egmond (1958)	38	42	8.5	Behavior	.48
Zander & Van Egmond (1958)	58	73	8.5	Behavior	.39

Note. NA = not available; HSPQ = High School Personality Questionnaire; STAI = State-Trait Anxiety Scale–Trait Anxiety; CMAS = Children's Manifest Anxiety Scale; GASC = General Anxiety Scale for Children; MPI = Maudsley Personality Inventory, MMPI = Minnesota Multiphasic Personality Inventory; TMAS = Taylor Manifest Anxiety Scale. Positive effect-size values indicate that males were higher on the trait than females. Effect sizes reported as .00 were known only to be nonsignificant.

[a] Adult sample. [b] Intellectual Achievement Responsibility scale (total score).

Table 2
Recent Studies of Gender Differences in Self-Esteem, Internal Locus of Control, Anxiety, and Assertiveness

Study	Male *n*	Female *n*	Sample	*M* age (years)	Method	*d*
			Self-esteem			
Carlson & Baxter (1984)	49	23	Irish homosexuals	24.5	RSE	.00[a]
Rice, Yoder, Adams, Priest, & Prince (1984)	1,024	86	Cadets	NA	TSCS	.06
Feather (1985)	83	114	College, Australia	23.0	SES	.31
Zeldow, Clark, & Daugherty (1985)	80	23	Medical students	23.6	RSE	−.10
Zuckerman (1985)	127	804	College	NA	RSE/TSBI	.10
de Jong-Gierveld (1987)	277	277	Adults, Holland	NA[b]	SES	.20
Marsh, Antill, & Cunningham (1987)	104	133	College, Australia	NA	JF + Berg	.22
Orlofsky & O'Heron (1987)	200	211	College	NA	MSCS + TSBI	.21
Payne (1987)	92	92	College	NA	JPI/RSE	.00
Watson, Taylor, & Morris (1987)	83	120	College	20.4	CSE/RSE	.20
Reynolds (1988)	247	342	College	NA	RSE	.00[a]
Rowlison & Felner (1988)	300	382	7th–12th graders	NA	SAI	.12
Lau (1989)	95	96	11th graders, Hong Kong	16.0	RSE + SCAS	.19
Ryff (1989)	79	54	Students	19.5	RSE	.11
Ryff (1989)	64	44	Adults	49.8	RSE	.20
Ryff (1989)	48	32	Adults	75.0	RSE	.34
Pennebaker, Colder, & Sharp (1987)	63	67	College	NA	RSE	.30
Allgood-Merten & Stockard (1991)	387	412	9th–12th graders	15.5	RSE	.43
Allgood-Merten & Stockard (1991)	20	32	12th graders	17.0	RSE	.25
Heatherton & Polivy (1991)	144	284	College, Canada	20.3	SSES	.18
Heatherton & Polivy (1991)	30	72	College, Canada	22.0	SSES	.17
Heatherton & Polivy (1991)	29	99	College, Canada	NA	SSES	−.27
Marsh & Byrne (1991)	350	548	College, Canada	NA	SDQ	.10
Marsh & Byrne (1991)	948	910	7th–11th graders, Australia	NA	SDQ	.12
Brenner & Cunningham (1992)	59	60	College and models	20.6	RSE	.57
Lortier-Lussier, Simond, Rinfret, & De Konick (1992)	32	32	Parents, Canada	34.4	RSE	.00
Stein, Newcomb, & Bentler (1992)	192	200	Adolescent	18.0	BPI + KSDS	.23
			Internal locus of control			
Rice et al., (1984)	1,024	86	Cadets	NA	Rotter	.04
Zeldow et al. (1985)	80	35	Medical students	23.6	Rotter	.21
Reynolds (1988)	247	342	College	NA	Rotter	.00[a]
Ryff (1989)	79	54	Students	19.5	Ad hoc	.17
Ryff (1989)	64	44	Adults	49.9	Ad hoc	.27
Ryff (1989)	48	32	Adults	75.0	Ad hoc	.40
Santelli, Bernstein, Zborowski, & Bernstein (1990)	25	36	Adults	31.0	Rotter	−.06
Santelli et al. (1990)	103	100	College, India	25.5	Rotter	−.32
Santelli et al. (1990)	45	116	College and faculty, Thailand	23.0	Rotter	.02
			Anxiety			
Jones, Briggs, & Smith (1986)	76	176	College	NA	SRS[c]	.05
de Jong-Gierveld (1987)	227	227	Adults	NA[b]	Ad hoc[c]	.16
McCann, Woolfolk, Lehrer, & Schwarcz (1987)	97	111	College	19.3	STAI	−.31
Payne (1987)	92	92	College	NA	FNE/SAD[c]	.02
Ben-Zur & Zeidner (1988)	151	223	College, Israel	23.8	STPI	−.64
Ingram, Cruet, Johnson, & Wisnicki (1988)	36	36	College	NA	SCS-SA[c]	.30
Rowlison & Felner (1988)	300	382	7th–12th graders	NA	RCMAS	−.54
Bruch, Gorsky, Collins, & Berger (1989)	42	42	College	19.4	CBS[c]	.18
Nystedt & Smari (1989)	175	223	High school, Sweden	18.0	SCS-SA[c]	−.33
Nystedt & Smari (1989)	39	200	College, Sweden	NA	SCS-SA[c]	−.13
Raskin & Novacek (1989)	28	29	College	21.0	MMPI-Anxiety	.05
Raskin & Novacek (1989)	87	86	College	20.0	MMPI-Anxiety	.03
O'Heron & Orlofsky (1990)	94	141	College	NA	STAI	.02
Endler, Parker, Bagby, & Cox (1991)	703	1,293	College, Canada	21.6	EMAS	−.35
Bernstein & Carmel (1991)	58	35	Medical school, Israel	NA	STAI	−.48
Hendryx, Haviland, & Shaw (1991)	70	40	Medical school	24.1	STAI	.00[a]
Kashani (1991)	75	75	High school	15.0	Ad hoc	−.48
Endler, Cox, Parker, & Bagby (1992)	221	384	College, Canada	20.5	EMAS/STAI	−.19

Table 2 (*continued*)

Study	Male *n*	Female *n*	Sample	*M* age (years)	Method	*d*
			Assertiveness			
Rice et al. (1984)	1,024	86	Cadets	NA	Ratings of leadership	.26
Vollmer (1984)	55	75	College, Norway	NA	SR Dominance	.00
Biaggio, Mohan, & Baldwin (1985)	41	50	College and adults	NA	16PF-Dominance	.90
Robinson & Follingstad (1985)	143	177	Singles	NA	RAS	.17
Robinson & Follingstad (1985)	60	63	Married	NA	RAS	.19
Watson et al. (1987)	83	120	College	24.0	NPI-Leadership	.23
Santelli et al. (1990)	25	36	Adults	31.0	RAS	−.18
Santelli et al. (1990)	103	100	College, India	25.5	RAS	.35
Santelli et al. (1990)	45	116	College and faculty, Thailand	23.0	RAS	.26
Lorr, Youmiss, & Stefic (1991)	544	335	High school	NA	SRS-Assertiveness	.04
Sawrie, Watson, & Biderman (1991)	174	195	College	20.9	CSES	−.06
Gurman & Long (1992)	21	39	College	21.2	SR-Leadership	−.02
Lortier-Lussier et al. (1992)	32	32	Parents	NA	PRF-Dominance	.16
Piedmont, McCrae, & Costa (1992)	50	114	College	18.3	EPPS-Dominance	.13
Piedmont et al. (1992)	57	109	College	18.3	EPPS-Dominance	.13

Note. RSE = Rosenberg Self-Esteem Scale; NA = not available; TSCS = Tennessee Self-Concept Scale; SES = Self-Esteem Scale; TSBI = Texas Social Behavior Inventory; JF + Berg = Janis-Field Self-Esteem and Berger Self-Esteem; MSCS = Mong Self-Concept Scale; JPI = Jackson Personality Inventory—Self-Esteem; CSE = Coopersmith Self-esteem scale; SAI = Self-Appraisal Inventory; SCAS = Self-Concept of Ability Scale; SSES = State Self-Esteem Scale; SDQ = Self-Description Questionnaire; BPI = Bentler Psychological Inventory; KSDS = Kaplan Self-Derogation Scale; Rotter = Rotter Internal-External Locus of Control Scale; SRS = Social Reticence Scale; STAI = Strait-Trait Anxiety Inventory–Trait Anxiety; FNE = Fear of Negative Evaluation Scale; SAD = Social Avoidance and Distress scale; STPI = State-Trait Personality Inventory–Trait Anxiety; SCS-SA = Self-Consciousness Scale—Social Anxiety; RCMAS = Revised Children's Manifest Anxiety Scale; CBS = Cheek–Buss Shyness scale; MMPI = Minnesota Multiphasic Personality Inventory; EMAS = Endler Multidimensional Anxiety Scale; SR-Dominance = self-ratings of dominance; 16PF = Sixteen Personality Factor Questionnaire; RAS = Rathus Assertiveness Scale; NPI = Narcissistic Personality Inventory; SRS = Social Relations Survey; CSES = College Self-Expression Scale; SR-Leadership = self-ratings of leadership; PRF = Personality Research Form; EPPS = Edwards Personal Preference Schedule. When two measures of a trait were used, the tabled effect size is either the average of the two separately computed effect sizes (when the two measures were treated as separate variables by investigators) or the effect size on a composite score combining the two measures (when investigators used such composite scores to represent the trait in analyses).
[a] Known only to be nonsignificant. [b] 25–75 years. [c] Social anxiety measure.

.08 (CI = .02 to .15), with both average effect sizes indicating no notable gender difference in assertiveness (although the mean effect size remained statistically significant). Moreover, the elimination of the outliers reduced the chi-square value for (lack of) homogeneity from 438.86 ($df = 21, p < .001$) to 39.63 ($df = 19, p < .01$).

The moderator variable analysis (conducted with the outliers removed) indicated that effect sizes varied significantly with both operationalization, $\chi^2(1) = 5.56, p < .05$, and age of subjects, $\chi^2(1) = 5.62, p < .05$. Although the sexes did not differ on the behavioral measures of assertiveness (weighted mean d = .04, k = 15, N = 2,864), males scored as more assertive than females on personality scales of assertiveness (weighted mean d = .23, CI = .09 to .36, k = 5, N = 926). And although there was no sex difference in assertiveness among children (weighted mean d = .03, k = 13, N = 2,484), male adolescents and adults were more assertive than female adolescents and adults (weighted mean d = .20, CI = .08 to .31, k = 7, N = 1,306). However, because all studies conducted with children used behavioral measures of assertiveness and most of the studies of adolescents and adults used personality scales, operationalization and age were largely confounded. Thus, it can be concluded only that (a) there was no gender difference in assertive behaviors among children and (b) male adolescents and adults scored higher than female adolescents and adults on personality scales of assertiveness (the effect size for greater male assertiveness, however, was small).

Replication of Hall's (1984) Meta-Analysis

The search retrieved 42 studies that were suitable for inclusion in the replication meta-analysis; these studies yielded 69 effect sizes from 54 independent samples (N = 18,730; 46% male). The number of effect sizes (samples) ranged from 9 to 27 across categories (traits), and the pooled sample sizes ranged from 2,942 to 10,755. The studies (and their effect sizes) are listed in Table 2.

The (unweighted) mean effect size was .16 for the sex difference in self-esteem (k = 27, N = 10,755; 48% male), indicating that males scored higher on measures of self-esteem, but to a very small degree. Effect sizes were also averaged by nation, yielding the following means: .19 for the United States (k = 16), .04 for Canada (k = 5), and .17 for other countries (k = 6). The mean effect size was .08 for internal locus of control (k = 9, N = 2,560; 67% male), indicating no notable gender difference on the trait.

The mean effect size for the gender difference in anxiety was −.15 (k = 18, N = 6,366; 40% male), indicating that females

Table 3
Temporal Trends in Gender Difference Effect Sizes for Adolescents and Adults

Source	Study years	Mean effect size (*d*)			
		Self-esteem	Internal locus of control	Anxiety	Assertiveness
Maccoby & Jacklin (1974)	1958–1974[a]	.10	.07	−.31	.20[b]
Hall (1984)	1975–1983	.12	.24	−.32	.12
Replication of Hall	1984–1992	.16	.08	−.15	.17

Note. A positive effect-size value indicates that males were higher on the trait than were females, and a negative effect-size value indicates that females were higher on the trait than were males.
[a] A few earlier studies were included. [b] Outliers deleted.

were slightly higher in anxiety than males. However, it was noted that effect sizes varied markedly by both type of anxiety measure (social vs. general) and nation. The mean effect size for the sex difference in general anxiety was −.26 (k = 11), and the mean effect size for the sex difference in social anxiety was .04 (k = 7). Thus, females scored higher than males in general anxiety to a small degree, but the sexes did not differ at all in social anxiety. When the effect sizes were averaged by nation, the means were −.04 for the United States (k = 12) and −.35 for other countries (k = 6), which suggests that males were higher than females in anxiety only outside of the United States. However, the overall mean effect size for the United States was misleading because of effect-size variations in U.S. samples with type of anxiety. Within the subset of U.S. studies, the mean effect sizes were −.18 for general anxiety (k = 7) and .14 for social anxiety (k = 5). Thus, in the United States, females were higher than males in general anxiety, but males were very slightly higher than females in social anxiety.

The mean effect size for assertiveness was .17 (k = 15, N = 4,104; 60% male), indicating a very small male advantage. Moreover, the mean effect size from the three samples obtained from outside of the United States was .20, which did not differ appreciably from the mean effect size of .16 in the subset of U.S. studies (k = 12).[6]

Comparisons of Gender Differences Across Meta-Analyses

Table 3 reports the mean effect sizes for sex differences in self-esteem, internal locus of control, anxiety, and assertiveness from three meta-analyses: the meta-analysis of findings from studies of adolescents through adults reviewed by Maccoby and Jacklin (1974), Hall's (1984) meta-analysis, and the current replication of Hall's meta-analysis. Although the effect sizes were sometimes minuscule (e.g., below .10), all three meta-analyses found that males, in comparison with females, were (a) higher in self-esteem, (b) more assertive, (c) more internally controlled, and (d) less anxious.

The effect sizes were most consistent across meta-analyses for self-esteem (ds = .10 to .16) and assertiveness (ds = .12 to .20). The effect size for greater female anxiety was about the same in the Maccoby and Jacklin (1974) and Hall (1984) data sets but was much smaller in the current replication of Hall's meta-analysis. However, this apparent decrease in effect size was attributable to the inclusion of findings of gender differences in social anxiety on which males and females did not generally differ, and which were more common in recent studies than in older studies reviewed by Hall (1984). Thus, if only the results for general anxiety are compared across meta-analyses, the gender differences in anxiety were also relatively constant (ds = −.26 to −.32) and were consistently larger than the gender differences in self-esteem, internal locus of control, and assertiveness.

There was a marked inconsistency across meta-analyses only for internal locus of control. Both of the meta-analyses conducted in this study found essentially no gender difference in internal control (ds = .07 to .08), whereas Hall's (1984) meta-analysis found males to be more internally controlled than females (d = .24).

Discussion

The meta-analysis of studies reviewed qualitatively by Maccoby and Jacklin (1974) generally confirmed Maccoby and Jacklin's conclusions about sex differences regarding internal locus of control, self-esteem, anxiety, and assertiveness: Males were significantly more assertive and significantly less anxious than females, and there was no appreciable overall sex difference in self-esteem or locus of control.

The meta-analysis of the sex differences in locus of control found previously unobserved variation in effect sizes associated with operationalization: Females were found to be more internally controlled than males when locus of control was mea-

[6] Because reviewers suggested that the findings from the meta-analyses may have been biased by the inclusion of estimated values for unknown effect sizes, that possibility was examined in the replication of Hall's (1984) meta-analysis. Of the 69 effect sizes included in the replication, only 5—all of which had values of .00 for effects known only to be nonsignificant—were not exact (see Table 2). With assumed values of .00 deleted, the mean effect sizes were .18, .09, −.16, and .17 for self-esteem, internal locus of control, anxiety, and assertiveness, respectively. These values differed by no more than .02 from the corresponding findings from analyses that used known and estimated effect sizes (see Table 3).

sured by the total score on the IAR scale, but males were found to be more internally controlled on other locus of control scales and when internal locus of control was measured behaviorally. In addition, although Maccoby and Jacklin concluded only that males were more assertive than females, the meta-analysis of the studies they reviewed found that the magnitude of the overall sex difference in assertiveness was very small (less than one tenth of a standard deviation after deletion of outliers), whereas the sex difference in (trait) anxiety was noteworthy (nearly one third of a standard deviation). Moreover, although the magnitude of the sex difference in anxiety was relatively constant across studies, the effect sizes for assertiveness were very heterogeneous, and much of that heterogeneity was ascribable to ages of subjects, operationalization of assertiveness, or both. There was no sex difference in assertiveness among children (for whom assertiveness was measured behaviorally); however, a small but notable greater male assertiveness was found among adolescents and adults (for whom assertiveness was usually measured by personality scales).

In addition, the mean effect sizes from the Maccoby and Jacklin data sets were generally not much different from the corresponding mean effect sizes in Hall's (1984) meta-analysis or from the replication of Hall's meta-analysis. Indeed, the mean effect sizes for contemporary (1984–1992) studies never differed by more than .06 from the corresponding mean effect sizes in the older studies (published mainly between 1958 and 1974) reviewed by Maccoby and Jacklin (when anxiety was defined consistently as general anxiety). Most important, the current meta-analysis of recent studies revealed that (a) the sexes did not differ in social anxiety and (b) the effect sizes for gender differences in personality found in studies conducted in the United States were about the same as those found in samples drawn from outside of the United States and Canada (although the number of international samples was limited).

Finally, although the meta-analysis of Maccoby and Jacklin's (1974) studies—the only data sets that included studies of children—compared gender differences in children with those in adolescents and adults, the developmental effects were usually confounded with method effects because the studies of children usually used behavioral measures of traits, whereas the studies of adolescents and adults usually used personality scales. Anxiety, however, was always measured by personality scales in the studies used in the meta-analysis, and the magnitude of the gender difference in anxiety did not vary significantly with age.

Study 2: Gender Differences on Personality Inventory Norms

Method

Selection of Personality Inventories

The three major approaches used to develop personality inventories are (a) empirical criterion keying, (b) factor-analytic strategies, and (c) the theory-guided rational method (Anastasi, 1986; Megaree, 1972). The meta-analysis included gender differences on selected scales from the most widely used commercially published personality inventories (mainly as identified by Anastasi, 1986) that exemplify each approach.

The empirically keyed inventories included the original and revised editions of both the Minnesota Multiphasic Personality Inventory (MMPI/MMPI-2) and the California Psychological Inventory (CPI/CPI-R), and the MMPI-Adolescent (MMPI-A). The factor-based inventories included the Guilford-Zimmerman Temperament Survey (GZTS), the Cattell (1973) inventories—two editions of the High School Personality Questionnaire (HSPQ), three of four editions of the Sixteen Personality Factor Questionnaire (16PF),[7] and the Institute for Personality and Ability Testing (IPAT) Anxiety Scale Questionnaire (IASQ)—the NEO Personality Inventory (NEO-PI/NEO-PI-R), the Gordon Personal Profile (GPP)/Gordon Personal Inventory (GPI), the Comrey Personality Scales (CPS), and two editions of the Eysenck personality inventories (Eysenck & Eysenck, 1976), the Maudsley Personality Inventory (MPI) and the Revised Eysenck Personality Questionnaire (EPQ-R).[8] Two inventories based on the rational method, both operationalizations of personality constructs posited by Murray (e.g., Murray et al., 1938), the Edwards Personal Preference Schedule (EPPS), and the Personality Research Form (PRF) were also included in the meta-analysis. (For all test references, see Appendix C.)

Categorization of Scales by Five-Factor Facet Model

With the exception of the IASQ (which measures only anxiety), the selected personality inventories are batteries that measure multiple traits. The scales from inventories used in the meta-analysis of gender differences were organized by the 30 hierarchically arranged traits (called *facets*) in Costa and McCrae's (1992) five-factor model, which are operationalized by their revised NEO-PI (NEO-PI-R), and which yield five higher order factors—Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness.

Comparisons of the descriptions of the scales contained in the selected inventories with the descriptions of the Costa–McCrae (1992) facets indicated that nine facets were most often measured by scales contained in the selected personality inventories: anxiety, impulsiveness, gregariousness, assertiveness, activity, ideas, trust, tender-mindedness,[9] and order. Classification of tests by trait consisted of grouping together scales with synonymous or nearly synonymous names (see Table 4 for the classification of personality scales by facet).

Classification of Scales by Trait Names

Scales labeled anxiety, neuroticism, and emotional stability constituted the anxiety category. Because scales named emo-

[7] The first edition of the 16PF reported norms only for combined sexes.

[8] Separate-sex norms for the United States were not reported in the test manuals for the other two Eysenck inventories (the Eysenck Personality Inventory and the Eysenck Personality Questionnaire).

[9] The construct labeled nurturance, empathy, or tender-mindedness on most of the selected personality inventories actually overlaps with two Costa and McCrae facets: tender-mindedness and altruism. The tender-mindedness label was used to name the effect category because tender-mindedness is the more common name for the construct. However, the practical effects of this decision were inconsequential because the sex difference on NEO-PI-R Altruism was virtually identical to the sex difference on NEO-PI-R Tender-mindedness.

Table 4
Classification of Personality Inventory Scales by Facets (Traits) in the Costa—McCrae Five-Factor Model

Facet	Scale
	Neuroticism
Anxiety	CPS Emotional Stability[a]
	GPP Emotional Stability[a]
	GZTS Emotional Stability[a]
	HSPQ Emotional Stability[a]
	IASQ
	MMPI/MMPI-2/MMPI-A Anxiety (Welsh)
	MPI/EPQ-R Neuroticism
	NEO-PI/NEO-PI-R Anxiety
	16PF Emotional Stability[a]
Impulsiveness	CPI/CPI-R Self-Control[a]
	GPI Cautiousness[a]
	GZTS Restraint[a]
	NEO-PI/NEO-PI-R Impulsiveness
	PRF Impulsiveness
	Extraversion
Gregariousness	CPI/CPI-R Sociability
	CPS Extraversion
	EPPS Affiliation
	GPP Sociability
	GZTS Sociability
	MMPI/MMPI-2/MMPI-A Social Introversion[a]
	MPI/EPQ-R Extraversion
	NEO-PI/NEO-PI-R Gregariousness
	PRF Affiliation
Assertiveness	CPI/CPI-R Dominance
	EPPS Dominance
	GPP Ascendancy
	GZTS Ascendancy
	HSPQ Dominance
	NEO-PI/NEO-PI-R Assertiveness
	PRF Dominance
	16PF Dominance
Activity	CPS Activity
	EPPS Endurance
	GPI Vigor
	GZTS Activity
	NEO-PI/NEO-PI-R Activity
	PRF Endurance
	Openness
Ideas	EPPS Intraception
	GPI Original Thinking
	GZTS Reflection
	NEO-PI/NEO-PI-R Ideas
	PRF Understanding
	Agreeableness
Trust	CPI/CPI-R Tolerance
	CPS Trust
	GPI Personal Relations
	GZTS Personal Relations
	NEO-PI-R Trust[b]
Tender-mindedness	CPI-R Empathy[c]
	CPS Empathy
	EPPS Nurturance
	HSPQ Tender-Mindedness
	NEO-PI-R Tender-Mindedness[b]
	PRF Nurturance
	16PF Tender-Mindedness

Table 4 (*continued*)

Facet	Scale
	Conscientiousness
Order	CPS Orderliness
	EPPS Order
	NEO-PI-R Order[b]
	PRF Order

Note. CPS = Comrey Personality Scales; GPP = Gordon Personal Profile; GZTS = Guilford-Zimmerman Temperament Survey; HSPQ = High School Personality Questionnaire; IASQ = Institute for Personality and Ability Testing Anxiety Scale Questionnaire; MMPI = Minnesota Multiphasic Personality Inventory; MMPI-A = MMPI Adolescent; MPI = Maudsley Personality Inventory; EPQ-R = Revised Eysenck Personality Questionnaire; NEO-PI = NEO Personality Inventory; 16PF = Sixteen Personality Factor Questionnaire; CPI = California Psychological Inventory; GPI = Gordon Personal Inventory; PRF = Personality Research Form; EPPS = Edwards Personal Preference Schedule.
[a] Reverse scored. [b] Norms for this scale not available for NEO-PI. [c] Norms for this scale not available for CPI.

tional stability scales were keyed so that lower scores indicated greater anxiety, effect sizes for sex differences on stability scales were reversed in sign (because the degree to which males are, for example, more emotionally stable than females is equivalent to the degree to which females are more anxious than males).

In addition to measures labeled impulsiveness, the impulsiveness category subsumed scales labeled self-control, restraint, and cautiousness, which were keyed so that greater impulsivity is indicated by lower scores. For these "reverse-scored" impulsivity measures, effect sizes were reversed in sign.

Scales labeled gregariousness, extraversion, sociability, and affiliation constituted the gregariousness effect category. Assertiveness-classified scales included measures of assertiveness, dominance, and ascendancy. Measures labeled activity, vigor, and endurance were assigned to the activity trait category. The ideas category subsumed tests having labels referring to cognition (e.g., original thinking, reflection, understanding, and intraception); this category is said to identify people who are curious, analytic, introspective, and meditative.

Scales labeled trust, tolerance, or personal relations were assigned to the trust category. Scales in the tender-mindedness category were labeled tender-mindedness, nurturance, and empathy.

Interrater Reliability of Classification of Scales

The names of the 113 scales contained in the selected inventories—excluding the NEO-PI/NEO-PI-R (which provided the taxonomy) and the HSPQ (whose scales share names with the related 16PF)—were first transcribed onto index cards. The trait names that were chosen to define each trait category were then recorded on separate cards to form effect category piles. Finally, two raters independently sorted each of the 113 scale names either into one of the nine trait piles or into an "other" category (for scales with names not contained in the category piles). The interrater agreement was essentially perfect, raters disagreeing over the classification of only two scales. One disagreement was attributable to carelessness. In the other case, there was a genuine disagreement as to whether a scale belonged in a particular category (and it was not used in that category).

Description of Norms

Norms for independent samples of males and females were obtained, as available, for three normative groups—high school students, college students, and general adults[10]—for each inventory. In addition, for six of the inventories (CPI, HSPQ, MMPI, MPI, PRF, and 16PF), new forms have been developed since their initial publication, and the norms for both the original forms and the norms for subsequent revisions were available and examined for those inventories. Norms for nations other than the United States were rarely reported in test manuals, and sex differences from them were not used in the main meta-analysis. However, because extensive international norms were obtainable for the PRF (Research Psychologists Press, 1993), a separate cross-cultural meta-analysis of gender differences on relevant PRF scales was conducted.

Calculation of Effect Sizes

Effect sizes (*ds*) were calculated for gender differences in both U.S. test norms and international PRF test norms by subtracting female means from corresponding male means and dividing the raw score differences by the pooled within-sex standard deviations. Thus, positive effect-size values indicated that males scored higher on the trait than did females, and negative effect-size values indicated that females scored higher than males. (However, as noted, the signs of effect sizes were reversed for scales that were keyed so that lower scores indicated higher levels of the trait.)

Data Analysis

Gender differences in U.S. norms. Although the meta-analytic methods of Hedges and Olkin (1985) are the most frequently used in contemporary meta-analyses (and were used in Study 1's meta-analysis of the Maccoby–Jacklin, 1974, data-

[10] Many test manuals contain separate norms for general (unselected) adults and for adults in specific occupations. Only sex differences in norms for general adults were used in the meta-analysis.

base), the earlier methods of Glass (e.g., those of Glass, 1976, which were used by Hall, 1984, and in the replication of Hall's analysis) are arguably superior in the rare cases (e.g., standardized test norms for the United States) in which all studies have very large sample sizes and effect sizes for individual studies can be viewed as parameters rather than statistics. Glass's method involves calculating the unweighted mean effect size to summarize the findings and the standard deviation of the raw effect sizes to examine homogeneity. By comparison, Hedges's method of meta-analysis involves examining homogeneity by using a chi-square test of significance, which is of dubious value when all sample sizes are large. Moreover, chi-squares cannot be compared across different trait categories to assess cross-category variations in homogeneity because of cross-category variations in both number of findings and pooled sample sizes.

Thus, effect sizes were grouped by trait, and the unweighted mean effect size, the standard deviations of effect sizes, and the median effect size (less affected by outliers than the mean effect size) were first calculated for each trait by averaging effect sizes over normative years and normative groups. Then, unweighted mean effect sizes (and standard deviations of effect sizes) were calculated by averaging effect sizes over normative groups (for inventories having multiple norms) to examine the constancy of sex differences across inventories. Next, effects of year on sex differences were examined by comparing unweighted mean effect sizes from normative data collected in 1940–1967 with the corresponding unweighted mean effect sizes from data collected in 1968–1992.[11] The reason that 1967 was chosen as the year for dichotomization was that the earlier norms for the six restandardized measures were all collected before 1968 and the later standardizations were conducted in 1968 or later. Thus, the use of 1967 as the year for dichotomization reduced the confounding of inventories with normative years in this moderator variable analysis. Separate analyses were also conducted that fully controlled for differences in both inventories and normative years by comparing effect sizes from earlier standardizations with those from later standardizations in the subset of inventories restandardized with the same normative groups. Finally, effect sizes were also averaged as a function of normative group by computing unweighted mean effect sizes for high school students, college students, and general adults. In addition, to fully control for confounding of normative categories with both tests and years, separate moderator variable analyses were conducted that compared the mean effect sizes of (a) high school students and college students, (b) college students and general adults, and (c) high school students and general adults. Each analysis used only inventories normed for both groups in the comparison.

Cross-cultural gender differences on the PRF. Because the sample sizes for the international norms of the PRF were small, the effect sizes (corrected for bias) were examined with the same meta-analytic procedures of Hedges and Olkin (1985) used in Study 1's meta-analysis of the Maccoby and Jacklin (1974) database. First, the weighted mean and median effect size were calculated for each of the seven trait categories of the PRF that measure Costa–McCrae facets (see Table 4). Second, the 95% CI was calculated for each weighted mean effect size. Third, the chi-square value for total variation among effect sizes in each category was partitioned into two sources: variation among countries (cross-national variations among effect sizes) and variation within countries (examining variations in gender differences across PRF forms). Finally, the weighted mean effect size was computed by trait for each country (i.e., collapsed over form).

Results

Description of Databases

Gender differences were calculated from U.S. norms for 13 personality inventories (36 independent normative groups; N = 105,742). One hundred forty-seven effect sizes (7 to 25 per trait) were used in the meta-analysis. The international normative data on the PRF (collected in 1985–1992) consisted of descriptive statistics that yielded 77 effect sizes from 11 independent samples of high school or college students in six nations (N = 1,050; 40% male).

Gender Differences in U.S. Test Norms

Overall gender differences. Table 5 lists the effect sizes by test, standardization year, and norm group, and the last three rows report the main results from the meta-analysis: unweighted mean effect sizes, standard deviations of effect sizes, and median effect sizes. Because of the essential equivalence between corresponding mean and median effect sizes, only the mean effect sizes (and associated standard deviations) are noted in the text. Moreover, because the sample sizes of the normative groups were consistently large (more than 500 examinees per group), sampling errors were trivial. Thus, the effect sizes reported in Table 5 were viewed as parameters rather than statistics, obviating the need for significance testing.

In terms of Cohen's (1977) criteria for interpreting effect magnitude, there were no appreciable (i.e., absolute values of ds > .19) gender differences for five traits: impulsiveness, gregariousness, activity, ideas, and order. However, the effect size of −.15 for gregariousness, although very small, indicated greater female gregariousness. Females scored higher than males, to a small degree (ds = −.25 to −.28), on scales of anxiety and trust. However, females scored much higher than males on tender-mindedness (d = −.97). Males scored higher than females, to a medium degree (d = .50), only on assertiveness.

The standard deviations of effect sizes were not constant across traits, indicating that the heterogeneity among effect sizes was greater for some traits than for others. The effect sizes were most consistent over scales, years, and normative groups for anxiety, activity, trust, and order (standard deviations of ds = .09 to .15). There was somewhat greater heterogeneity for impulsiveness, gregariousness, assertiveness, and ideas (standard deviations of ds = .19 to .36), and atypically large heterogeneity of effect sizes in the tender-mindedness category (standard deviation of d = .51). However, half of the 18 effect sizes for

[11] When normative data were collected over a period of years, the mean of the 2 years constituting the data collection range was used to categorize norms by year.

Table 5
Effect Sizes of Gender Differences in Personality From Standardized Personality Test Norms

	Norms			Costa–McCrae factors and traits								
				Neuroticism		Extraversion			Openness	Agreeableness		Conscientiousness
Scale	Year(s)	Group	*N*	Anxiety	Impulsiveness	Gregariousness	Assertiveness	Activity	Ideas	Trust	Tender-mindedness	Order
CPI	1956	High school	7,628		.28	.02	−.08			−.17		
CPI	1956	College	3,253		.43	−.12	−.03			−.38		
CPI-R	1987[a]	High school	8,655		.39	−.11	.06			−.34	−.16	
CPI-R	1987[a]	College	7,362		.31	−.23	.18			−.44	−.18	
CPS	1970	NA	727	−.22		.10		.17		−.42	−.43	−.34
EPPS	1959	College	1,509			−.57	.69	.01	−.24		−.52	.00
EPPS	1959	General adults	8,963			−.76	.85	.10	−.26		−.60	−.19
GPI	1956–1963	High school	1,833		.03			.03	.14	−.26		
GPI	1956–1978	College	3,109		.03			.12	.19	−.15		
GPP	1952–1963	High school	6,276	−.39		−.34	.00					
GPP	1952–1978	College	5,847	−.30		−.28	.04					
GZTS	1952–1955	High school	1,070	−.28	−.37	−.43	.32	.21	−.29	−.16		
GZTS	1955–1971	College	5,801	−.22	.64	−.12	.43	.09	.12	−.04		
GZTS	1955	General adults	1,380	−.13	.04	−.11	.45	.04	.04	−.25		
HSPQ	1958–1962	High school	7,867	−.33			.65				−1.57	
HSPQ-AB	1968	High school	4,951	−.43			1.12				−1.77	
IASQ	1970	High school	525	−.27								
IASQ	1970	College	1,392	−.20								
IASQ	1970	General adults	935	−.26								
MMPI	1940	General adults	724	−.32		.54						
MMPI-A	1992	High school	1,620	−.31		.12						
MMPI-2	1989	General adults	2,600	−.22		.24						
MPI	1959	College	1,064	−.14		−.12						
EPQ-R[b]	1992	Mixed	890	−.57		−.11						
NEO-PI	1989	College	526	−.59	−.20	−.09	.24	.01	.20			
NEO-PI-R	1985	General adults	1,000	−.41	−.23	−.21	.19	−.11	.23	−.19	−.32	−.05
PRF Form AA	1967	College	2,031		−.15	−.36	.55	.15	−.10		−.87	.03
PRF Form E	1974	Junior high/ high school	1,877		−.07	−.49	.49	.10	.18		−1.22	−.26
PRF Form E	1974	College	2,765		−.28	−.16	.59	.23	.16		−.56	−.07
16PF Form A	1958	College	604	−.28			.49				−.97	
16PF Forms A and B	1962	High school	1,193	−.24			.92				−1.39	
16PF Forms A and b	1962	College	1,412	−.18			1.08				−1.13	
16PF Forms A and B	1962	General adults	2,011	.04			1.09				−1.37	
16PF Forms A and B	1968	High school	2,699	−.18			.70				−1.52	
16PF Forms A and B	1968	College	1,710	−.02			.74				−1.54	
16PF Forms A and B	1968	General adults	1,933	−.48			.78				−1.37	
k				25	14	22	25	13	12	11	18	7
Median *d*				−.27	−.04	−.14	.49	.09	.13	−.25	−1.07	−.07
Unweighted mean *d*				−.28	.06	−.15	.50	.09	.03	−.25	−.97	−.13
Standard deviation of *d*				.15	.29	.28	.36	.09	.19	.12	.51	.13

Note. CPI = California Psychological Inventory; CPI-R = revised CPI; CPS = Comrey Personality Scales; NA = not available; EPPS = Edwards Personal Preference Schedule; GPI = Gordon Personal Inventory; GPP = Gordon Personal Profile; GZTS = Guilford-Zimmerman Temperament Survey; HSPQ = High School Personality Questionnaire; HSPQ-AB = HSPQ Forms A and B combined; IASQ = Institute for Personality and Ability Testing Anxiety Scale Questionnaire; MMPI = Minnesota Multiphasic Personality Inventory; MMPI-A = MMPI-Adolescent; MPI = Maudsley Personality Inventory; EPQ-R = Revised Eysenck Personality Questionnaire; NEO-PI = NEO Personality Inventory; NEO-PI-R = revised NEO-PI; PRF = Personality Research Form; 16PF = Sixteen Personality Factor Questionnaire; *k* = number of effect sizes. Positive effect sizes indicate that males scored higher on the trait than did females.

[a] Year norms were published (collected earlier over an unspecified period of years). [b] Revision of the MPI.

tender-mindedness sex differences involved Cattell's scales (HSPQ and 16PF), and the mean of those 9 effect sizes was −1.40. By comparison, the mean effect size for the other 9 tender-mindedness scales was −.54, with a much smaller associated standard deviation of .32. Thus, with the outliers from Cattell's scales deleted, the gender difference in tender-mindedness was of medium size, and there was typical heterogeneity among effect sizes.

Moderation of gender differences by inventory. The top section of Table 6 reports mean effect sizes by inventory (i.e., collapsed over standardizations and normative groups). Females scored higher (*d*s of −.19 or greater) than males on all anxiety scales, although the effect size was small on all scales except the NEO-PI/NEO-PI-R (on which it was medium). On impulsiveness, males scored higher than females on the CPI/CPI-R (d = .35) and lower than females on the NEO-PI/NEO-PI-R and PRF (*d*s = −.17 to −.22). There were essentially no gender differences in impulsiveness on the GPI and GZTS (*d*s = .03 to .10).

Females scored somewhat higher than males on the gregariousness scales of the GPP, GZTS, and PRF (*d*s = −.22 to −.34) and, to a larger degree, on the gregariousness scale of the EPPS (d = −.66). Gender differences in gregariousness were practically nonexistent on the CPI/CPI-R, CPS, MPI/EPQ-R, and NEO-PI/NEO-PI-R (*d*s = −.15 to .10). Males scored higher than females in gregariousness only on the MMPI-A/MMPI/MMPI-2 (d = .30).

Males scored higher than females on all assertiveness scales except those of the CPI/CPI-R and GPP, on which there were essentially no gender differences. The effect sizes for the greater male assertiveness on the other scales varied; they were small on the GZTS and NEO-PI/NEO-PI-R (.22 to .40), medium on the PRF (.54), and large on the EPPS, HSPQ, and 16PF (.77 to .88).

The effect sizes for gender differences in activity were consistently trivial (−.05 to .17). Females scored higher than males on all measures of trust (*d*s = −.15 to −.42), although the effect size on the GZTS (−.15) was trivial.

Females scored higher than males on all scales of tender-mindedness. The effect sizes were very small on the CPI-R (−.17), small on the NEO-PI-R (−.32), medium or near medium on the CPS and EPPS (−.43 to −.56), and large on the HSPQ, PRF, and 16PF (−.88 to −1.67). Although there were essentially no gender differences in order on the EPPS, PRF, and NEO-PI-R (*d*s = −.05 to −.10), males were lower than females in order on the CPS (d = −.34).

Moderation of gender differences by year. The middle section of Table 6 reports mean effect sizes by trait and year category (i.e., collapsed over scales and normative groups) using all inventories. The absolute values of the mean effect sizes ranged from .05 to 1.05 (M = .28) across the nine traits for older normative data (1940–1967) and from .01 to .91 (M = .29) for recent normative data (1968–1992). Thus, averaged over normative groups, scales, and traits, the magnitude of the differences between the sexes was, on average, about the same in both time periods.

However, some gender differences may have increased while others decreased over the same period. That is, there could have been a three-way Gender × Year Level × Trait interaction on personality scores but no two-way Gender × Year Level interactions, resulting in negligibly different effect sizes between year levels when absolute values of mean effect sizes were averaged over traits by year level. Thus, year-related variations in gender differences were also examined by determining the absolute values of the differences between mean effect sizes from early and recent norms by trait. These absolute-value differences ranged from .01 to .24 across traits (M = .12), indicating few year-related variations in effect sizes.

The mean effect sizes were also calculated by normative year range for the six inventories that had been standardized with the same normative groups in each year category (i.e., CPI/CPI-R, HSPQ, MMPI/MMPI-2, MPI/EPQ-R, PRF, and 16PF third and fourth editions). Therefore, unlike the findings from the moderator variable analysis that used all tests and all norms, these year-related differences were not partially confounded with normative groups and scale differences. However, because it was believed that at least two effect sizes were needed in each year category (from at least two different scales) to afford a meaningful comparison of effect sizes over years, only five trait categories—anxiety, impulsiveness, gregariousness, assertiveness, and tender-mindedness—were examined for gender differences in this controlled analysis.[12]

The absolute values of effect sizes in the controlled analysis ranged from .01 to 1.27 (M = .45) across traits in the earlier norms and from .03 to 1.35 (M = .49) in recent norms. Thus, averaged over traits, normative groups, and scales, the magnitude of gender differences was essentially the same in both normative year ranges, corroborating the results from the analysis of normative data from all inventories.

The absolute values of the differences between mean effect sizes from older and newer data sources in this subgroup of renormed inventories ranged from .00 to .12 (M = .05) across the five traits. Thus, there was no appreciable moderation of effect size by year for any trait.

Moderation of gender differences by normative group. The bottom section of Table 6 reports the mean effect sizes as a function of normative group (i.e., collapsed over years and scales) from the analysis that used data from all normative groups (except for the CPS and EPQ-R, each of which used a single standardization sample that was either undefined by age or contained examinees heterogeneous in age). The absolute values of effect sizes ranged from .01 to 1.18 (M = .31) across traits for high school students,[13] from .01 to .82 (M = .25) for college students, and from .00 to .92 (M = .26) for general adults. Thus, averaged over traits, scales, and years, the magnitude of gender differences did not vary notably across normative groups.

The findings in each column in Table 6 can be conceptualized as reflecting three pairwise comparisons among effect sizes

[12] For the samples examined in 1940–1967 and 1968–1992, the effect sizes, respectively, were −.20 and −.32 for anxiety (*k*s = 6), .19 and .14 for impulsiveness (*k*s = 3), −.01 and −.03 for gregariousness (*k*s = 5), .60 and .60 for assertiveness (*k*s = 7), and −1.27 and −1.35 for tender-mindedness (*k*s = 5).

[13] Sex differences on the PRF that were based on norms that mixed junior high school and high school students were included in the high school category in this analysis.

Table 6
Effect Sizes of Gender Differences in Personality From Standardized Personality Inventories by Inventory, Year of Standardization, and Normative Group

			Costa–McCrae factors and traits								
			Neuroticism		Extraversion			Openness	Agreeableness		Conscientiousness
Category	*k*	*N*	Anxiety	Implusiveness	Gregariousness	Assertiveness	Activity	Ideas	Trust	Tender-mindedness	Order
By inventory											
CPI/CPI-R	4	26,898		.35	−.11	.03			−.33	−.17[a]	
CPS	1	727	−.22		.10		.17		−.42	−.43	−.34
EPPS	2	10,472			−.66	.77	.06	−.25		−.56	−.10
GPI	2	4,942		.03			.08	.16	−.20		
GPP	2	12,123	−.34		−.31	.02					
GZTS	3	8,251	−.21	.10	−.22	.40	.11	−.04	−.15		
HSPQ	2	12,818	−.38			.88				−1.67	
ISAQ	3	2,852	−.24								
MMPI[b]	3	4,388	−.28		.30						
MPI/EPQ-R	2	1,958	−.36		.00						
NEO-PI/NEO-PI-R	2	1,526	−.50	−.22	−.15	.22	−.05	.22	−.19[c]	−.32[c]	−.05[c]
PRF	3	6,673		−.17	−.34	.54	.16	.08		−.88	−.10
16PF	7	11,562	−.19			.83				−1.19	
No. of inventories			9	5	9	8	6	5	5	7	4
Median *d*			−.28	.03	.15	.47	.10	.08	−.20	−.56	−.10
Unweighted mean *d*			−.30	.02	.15	.46	.09	.03	−.26	−.75	−.15
By year of standardization											
1940–1967	19	63,575	−.23	.12	−.22	.50	.09	−.05	−.20	−1.05	−.05
k_T			12	8	12	15	8	8	7	8	3
1968–1992	17	42,167	−.32	−.01	−.07	.51	.08	.19	−.35	−.91	−.18
k_T			13	6	10	10	5	4	4	10	4
By normative group											
High school	12	46,194	−.30	.05	−.20	.46	.11	.01	−.23	−1.18	−.26
k_T			8	5	6	9	3	3	4	6	1
College	14	38,385	−.24	.11	−.23	.45	.10	.06	−.25	−.82	−.01
k_T			8	7	9	11	6	6	4	7	3
General adults	8	19,546	−.25	−.10	−.06	.67	.01	.00	−.22	−.92	−.12
k_T			7	2	5	5	3	3	2	4	2

Note. CPI = California Psychological Inventory; CPI-R = revised CPI; CPS = Comrey Personality Scales, EPPS = Edwards Personal Preference Schedule; GPI = Gordon Personality Inventory; GPP = Gordon Personal Profile; GZTS = Guilford-Zimmerman Temperament Survey; HSPQ = High School Personality Questionnaire; ISAQ = Institute for Personality and Ability Testing Anxiety Scale Questionnaire; MMPI = Minnesota Multiphasic Personality Inventory; MPI = Maudsley Personality Inventory; EPQ-R = Revised Eysenck Personality Questionnaire; NEO-PI-R = revised NEO Personality Inventory; PRF = Personality Research Form; 16PF = Sixteen Personality Factor Questionnaire; *k* = number of samples, k_T = number of effect sizes per trait. Positive effect-size values indicate that males scored higher than females on the trait.
[a] $k = 2$, $N = 16{,}017$. [b] Includes the MMPI-2 and MMPI Adolescent. [c] $k = 1$, $N = 1{,}000$.

(high school vs. college, college vs. general adults, and high school vs. general adults). Therefore, the three possible age comparisons among effect sizes were conducted for each trait by calculating absolute values of the differences between all three possible pairs of effect sizes. The absolute values of the differences between effect sizes for high school students and the corresponding effect sizes for college students ranged from .01 to .36 across traits; the absolute values of the differences between effect sizes for college students and the corresponding effect sizes for general adults ranged from .01 to .22; and the absolute values of the differences between effect sizes for high school students and the corresponding effect sizes for general adults ranged from .01 to .26. The means of these absolute values (i.e., collapsed over traits) for high school students, college students, and general adults were .09, .11, and .12, respectively. Therefore, variations in effect sizes among the different normative groups were consistently trivial.

Unfortunately, scales and years were confounded with normative group in the preceding analyses. Thus, three corresponding but controlled analyses were conducted that used only test standardizations normed for both of the two relevant normative groups in each pairwise comparison. More specifically, the norms from the CPI/CPI-R, GPI, GPP, IASQ, PRF Form E, and 16PF (third and fourth editions) were used in the analysis that compared gender differences in high school students with those of college students; the norms from the EPPS, IASQ, NEO-PI/NEO-PI-R, and 16PF (third and fourth editions) were used in the analysis that compared gender differences in college students with those for adults; and the norms from the GZTS, IASQ, MMPI-2/MMPI-A, and 16PF (third and fourth editions) were used in the analysis that compared gender differences in high school students with those for adults. Because there had to be at least two effect sizes per normative group (from at least two different scales) if mean effect sizes were to be compared across groups, there was an insufficient number of effect sizes to afford comparisons for all nine traits. However, norms for both high school and college students had been obtained in enough standardizations to afford comparisons of gender differences between the two kinds of students for all traits except order. Differences in effect sizes between college students and general adults could be examined only for anxiety, gregariousness, assertiveness, and activity, and differences in effect sizes between high school students and general adults could be examined only for anxiety, gregariousness, assertiveness, and tender-mindedness.[14]

In the high school versus college comparison, the absolute values of the effect sizes ranged from .06 to 1.02 across eight traits in high school norms and from .12 to .95 in college norms, with .31 as the mean of the absolute values for both groups. Thus, averaged over scales, years, and traits, the size of the average male–female difference in personality traits was about the same in high school as it was in college (in the subset of inventories normed for both kinds of students). The absolute values of the differences between high school effect sizes and corresponding college effect sizes ranged from .02 to .12 (M = .07) for the eight traits, indicating that there was no notable moderation of effect size by normative group for any trait.

In the college versus general adult comparisons, the absolute values of the effect sizes ranged from .01 to .69 across the four traits examined for the college group and from .00 to .73 for the general adult group. The means of these absolute values were .32 and .37, respectively. Thus, averaged over traits, years, and scales (for standardizations that obtained norms for both college and adult samples), gender differences were essentially the same for college students as for general adults. The absolute values of the differences between college effect sizes and the corresponding general adult effect sizes were all trivial (.01 to .15, M = .06).

In the high school versus general adult comparison, the absolute values of the gender differences ranged from .16 to 1.02 across four traits for high school students and from .06 to 1.12 for general adults. The means of these absolute values were virtually identical: .52 for high school students and .54 for adults. Thus, averaged over traits, years, and scales (for standardizations that obtained norms for both high school students and general adults), the magnitude of gender differences did not vary appreciably by normative group. The absolute values of the differences between high school effect sizes and the corresponding general adult effect sizes ranged from .05 to .22 (M = .12) across the four traits, indicating no notable effect-size variation across the two normative groups for any trait.

Thus, the results of pairwise comparisons among effect sizes by traits all showed no notable variations of effect sizes across normative groups. These findings corroborated those from the moderator variable analyses that used all effect sizes but in which effects of normative groups on effect sizes were confounded with scale and year effects.

Meta-Analysis of Gender Differences in Non-U.S. Norms

The effect sizes for gender differences in the seven PRF traits that measure Costa and McCrae (1992) facets are reported in Table 7 by nation and PRF form, and the results from the meta-analysis are reported in Table 8.

Because the weighted mean effect sizes were always very similar to the corresponding median effect sizes, only the latter are noted in the text. The mean effect sizes were statistically significant for four traits. Males scored significantly higher than females on the PRF measure of assertiveness, and females scored significantly higher on the PRF measures of impulsivity,

[14] For the samples of high school and college students in the standardizations that obtained norms for both groups, the effect sizes, respectively, were −.27 and −.18 for anxiety (ks = 4), .16 and .12 for impulsiveness (ks = 4), −.23 and −.14 for gregariousness (ks = 4), .32 and .40 for assertiveness (ks = 5), .06 and .18 for activity (ks = 2), −.26 and −.32 for trust (ks = 3), and −1.02 and −.95 for tender-mindedness (ks = 3). For the samples of high school students and college students in the standardizations that obtained norms for both groups, the effect sizes, respectively, were −.25 and −.28 for anxiety (ks = 4), −.33 and −.48 for gregariousness (ks = 2), .69 and .73 for assertiveness (ks = 4), and .01 and .00 for activity (ks = 2). For the samples of high school students and general adults in the standardizations that obtained norms for both groups, the effect sizes, respectively, were −.26 and −.21 for anxiety (ks = 5), −.16 and −.06 for gregariousness (ks = 2), .65 and .77 for assertiveness (ks = 3), and −1.02 and −1.12 for tender-mindedness (ks = 3).

Table 7
Effect Sizes for Gender Differences in Personality on the Personality Research Form (PRF) for Costa–McCrae Facets in Non-U.S. Student Samples by PRF Form

				Costa–McCrae factors and traits						
				Neuroticism	Extraversion			Openness	Agreeableness	Conscientiousness
Country	Form	Male *N*	Female *N*	Impulsiveness	Gregariousness	Assertiveness	Activity	Ideas	Tender-mindedness	Order
Canada	E	100	120	−.41	−.16	.61	.21	.23	−.48	.02
Canada	N	13	17	.07	−.40	.76	1.16	−.14	−1.12	−.16
Canada	N	11	19	−.24	.11	−.01	.14	−.57	−.60	−1.42
China	B	61	43	−.10	.09	.16	−.20	.03	.03	−.22
Finland	B	24	68	−.14	.24	−.25	.07	−.28	−.08	−.05
Finland	E	16	72	−.44	−.69	−.06	.76	.34	−.67	−.33
Germany	B	36	54	.08	.03	.06	.13	−.43	−.25	−.35
Germany	E	41	61	−.13	−.20	.31	.11	−.09	−.44	−.11
Poland	B	36	64	.00	−.30	.32	−.25	−.25	−.49	−.37
Poland	E	49	45	−.09	.09	.55	.07	−.13	−.29	−.20
Russia	B	28	72	−.04	−.22	.32	−.55	.14	−.47	−.46

Note. The names of the PRF scales measuring the Costa–McCrae traits (facets) of impulsiveness, gregariousness, assertiveness, activity, ideas, tender-mindedness, and order are Impulsiveness, Affiliation, Dominance, Endurance, Understanding, Nurturance, and Order, respectively. Norms from which effect sizes were calculated were for college students, except for Canadian Form E.

tender-mindedness, and order. Significant cross-national variations in effect sizes were observed only for assertiveness and activity. Although there was cross-national variation for assertiveness, the effect sizes were uniformly positive and of small (or approaching small) or medium value. Thus, the Country × Gender interaction was clearly ordinal, with the homogeneity test merely indicating that the male advantage in assertiveness was significantly larger in some countries than in others. Moreover, a comparison of these gender differences with those in the 1974 U.S. college norms for the PRF (the most relevant comparison group; see Table 6) indicated few cross-national differences in findings. In the U.S. PRF norms, there was a notable sex difference in gregariousness but not in order. Finally, because only one of seven homogeneity tests for variations in effect sizes over PRF forms (within-country variations) was significant (at about the chance level), there was consistency of effect sizes over PRF forms.

Discussion

The meta-analysis of sex differences on standardized tests of personality indicated that males and females differed on five of the nine Costa–McCrae traits (facets) frequently measured by personality inventories. Males generally scored higher than females on scales of assertiveness, and females scored higher than males on scales of anxiety, gregariousness, trust, and tender-mindedness. Most important, the effect sizes generally did not vary appreciably across years of norms, ages of examinees, educational levels of examinees, or nations.

However, sex differences were not constant across different measures of the same traits. For example, although males usually scored higher than females in assertiveness, there were two scales of assertiveness that did not differentiate between the sexes.

Variations in sex differences across scales of the same traits were not surprising. Different measures of the same trait do not assess exactly the same latent dimension because the correlations among such tests are invariably well below their respective reliability coefficients. The reliable variance in each test can be partitioned into two sources: common factor variance (variance shared by all tests of a trait) and unique variance (variance unique to each measure in its category). Moreover, the proportion of each test's reliable variance that is unique varies by test. The sex difference in the common factor variance component of each test is constant across tests, but the sex difference in the unique variance may vary markedly across tests. The sex difference on a given test is a function of both the sex difference in the common factor variance and the sex difference in the unique variance. Thus, sex differences would be expected to vary across different measures of the same trait unless the sex difference in each test's unique variance is constant across tests. Tests that would yield the most discrepant sex differences are those containing a unique variance component that differentiates between the sexes differently than does the unique variance component on most other tests in the same category, particularly when the unique variance component accounts for a large amount of the test's reliable variance (i.e., when the test is not

Table 8
Cross-National Meta-Analysis of Effect Sizes for Gender Differences in Personality for Students (1985–1992) on the Personality Research Form (PRF) for Costa–McCrae Facets

				Costa–McCrae factors and traits						
				Neuroticism	Extraversion			Openness	Agreeableness	Conscientiousness
Country	*k*	Male *N*	Female *N*	Impulsiveness	Gregariousness	Assertiveness	Activity	Ideas	Tender-mindedness	Order
Canada	3	124	156	−.34	−.16	.56	.29	.11	−.55	−.11
China	1	61	43	−.10	.09	.16	.20	.03	.03	−.22
Finland	2	40	140	−.27	−.15	.17	.36	.02	−.33	−.17
Germany	2	77	115	−.03	−.09	.19	.12	.25	−.35	−.22
Poland	2	85	109	−.05	−.10	.43	−.09	−.19	−.39	−.29
Russia	1	28	72	−.04	−.22	.32	−.55	.14	−.47	−.46
Median *d*				−.07	−.12	.26	.18	.07	−.37	−.22
Weighted mean *d*				−.16	−.11	.30	.07	−.04	−.38	−.22
LCI for mean *d*				−.29	−.24	.18	−.06	−.17	−.51	−.35
UCI for mean *d*				−.03	.02	.43	.20	.09	−.25	−.09
Among-countries homogeneity χ^2 ($df = 5$)				4.31	1.47	13.07*	16.42**	5.36	6.54	2.09
Within-countries homogeneity χ^2 ($df = 5$)				2.83	9.66	4.26	9.97	8.69	5.78	12.18*

Note. The names of the PRF scales measuring the Costa–McCrae traits (facets) of impulsiveness, gregariousness, assertiveness, activity, ideas, tender-mindedness, and order are Impulsiveness, Affiliation, Dominance, Endurance, Understanding, Nurturance, and Order, respectively. *k* = number of effect sizes; LCI = lower limit of 95% confidence interval; UCI = upper limit of 95% confidence interval.
* $p < .05$. ** $p < .01$.

very highly correlated with other tests in its category after corrections for attenuation).

Variations in sex differences across scales may also be a function of the multidimensional nature of the traits. If, for example, all of the items from the various gregariousness scales were extracted, administered together as a single scale, and factor analyzed, factors representing different subfacets of gregariousness would probably emerge. Although scores from each of the inventory scales of gregariousness may be reproducible from the set of factors, the weight given to each factor in a regression equation would vary by test. Thus, if sex differences vary across gregariousness subfacets, they would also vary across gregariousness tests, because different gregariousness tests would tap different subfacets to different degrees.

This could explain, for example, why the MMPI and (to a trivial degree) CPS norms showed greater male gregariousness, whereas females were either equally gregarious or more gregarious than males on all other scales. The MMPI and CPS gregariousness scales contain many items concerning social anxiety (on which the replication meta-analysis in Study 1 found males in the United States to score slightly higher than females), whereas the items in the other scales were generally less clinical, focusing mainly on need for affiliation. Differences in test content could also account for variations in gender differences across different anxiety tests. Research has shown that sex differences on some components tapped by anxiety measures, mainly those relating to subjective well-being and happiness, are not the same as sex differences on other components (females report the same level of happiness as, or a greater level of happiness than, men; Fujita, Diener, & Sandvik, 1991; Myers, 1992; Wood, Rhodes, & Whelan, 1989). Thus, the sex difference found for anxiety will vary as a function of the amount of variance in the anxiety measure tapping subjective well-being.

General Discussion

Overview of Findings

Study 1 examined gender differences in assertiveness, internal locus of control, self-esteem, and anxiety in the sets of studies reviewed by Maccoby and Jacklin (1974) and Hall (1984), as well as from more recent studies (the replication of Hall's meta-analysis). Males were found to be more assertive and less anxious than females. However, the greater male assertiveness was established mainly on self-report personality scales completed by adolescents and adults, and the greater female anxiety was found for measures of general but not social anxiety. There was no consistent overall sex difference in locus of control, although the meta-analysis of the Maccoby–Jacklin data indicated that sex differences in locus of control varied with its operationalization. In addition, males were found to have higher self-esteem than females, but the effect size was very small. Most important, gender differences were found to have remained relatively constant across generations (roughly from the late 1950s to the early 1990s) and to be generally invariant across nations.

Study 2 examined gender differences in nine of Costa and McCrae's (1992) traits (facets) in the norms for scales measuring them from widely used personality inventories. The results showed that (a) males scored higher than females on scales of assertiveness; (b) females scored notably higher than males on scales of anxiety, trust, and—especially—tendermindedness (e.g., nurturance); (c) females were slightly higher than males on extraversion; and (d) there were essentially no overall gender differences on scales of impulsiveness, activity, ideas, and order (although gender differences were sometimes found on specific operationalizations of these traits). Gender differences were generally invariant across ages, educational levels, or nations.

Comparisons Between Meta-Analyses of Studies and Meta-Analyses of Test Norms

The traits examined for gender differences in Study 1 overlapped only partially with the traits examined in Study 2. Only anxiety and assertiveness were examined in both studies. The weighted mean effect size for the gender difference in (general) anxiety was found to be about −.30 in both studies. Thus, although females were higher in trait anxiety than males, the effect size was small in both data sources.

Although Study 1 and Study 2 both found male adolescents and adults to be more assertive than females, the mean effect sizes were much smaller in Study 1. There are two possible explanations for this finding. First, samples used to standardize tests may differ systematically from samples used in general psychological research. For example, introductory psychology students are overrepresented in the latter. Gender differences in assertiveness may be moderated by sample composition. In addition, results found in the literature may not accurately reflect findings obtained in psychological research because of selectivity among scientists as to which findings to report. For example, findings that males are more assertive than females, if deemed "politically incorrect," may more often go unreported than findings of no gender differences in assertiveness, thus resulting in an artifactually deflated effect size in the results from a meta-analysis of studies in the literature. Because (a) the samples obtained by major test publishers to standardize tests are probably more representative than the samples used in psychological research by individual researchers and (b) gender differences in test norms are completely uncompromised by "publication biases," the larger effect size for the gender difference in assertiveness found in Study 2 is probably the more accurate estimate of the effect size in the general population. Moreover, an important implication of the discrepancy between findings for assertiveness from the two studies is that mean effect sizes for other gender differences in personality found in meta-analyses of the literature may be also be biased.

Theoretical Issues

The findings from the meta-analyses are consistent with the theory that males and females are differentiated along the agentic versus communal continuum described by Bakan (1966), which posits that males are higher than females on agentic (sometimes known as instrumental) traits and that females are higher than males on communal (sometimes known as expressive) traits. The personality dimensions that most strongly differentiated between the sexes were assertiveness and tender-

mindedness, which are nearly pure measures of agency and communality, respectively. Moreover, the finding of a large sex difference in tender-mindedness is consistent with the large effect size (weighted mean $d = .91$, unweighted mean $d = .99$) found in a previous meta-analysis of sex differences in personality scales of empathy (Eisenberg & Lennon, 1983).

Magnitude of Gender Differences in Personality

In addition to the use of Cohen's (1977) criteria, the magnitude of effect sizes of sex differences in personality can be assessed through comparisons with sex differences on other individual-difference variables: cognitive, affective, and physical. In the cognitive realm, Feingold's (1988a) analysis of sex differences among high school seniors in the 1980 norms for eight abilities measured by the Differential Aptitude Tests (DAT; Bennett, Seashore, & Wesman, 1982) found male students to score higher than female students on mechanical reasoning ($d = .89$) and spatial visualization ($d = .22$), whereas female students scored higher than male students on measures of spelling ($d = -.51$), language use ($d = -.40$), and perceptual speed ($d = -.32$). No sex difference was found on verbal reasoning, figural reasoning, or arithmetic ($ds = -.01$ to $.02$). The mean of the absolute values of these contemporary cognitive gender differences on the DAT was .29. Meta-analytic research has also shown females to be appreciably better than males at decoding nonverbal cues ($d = -.42$; Hall, 1984).

Noteworthy sex differences also have been found for affective dimensions. A meta-analysis by Feingold (1992b) of U.S. and Canadian studies of gender differences in mate selection preferences (essentially one-item self-report personality scales expressing the value people accord to different characteristics when they evaluate prospective mates or romantic partners) found that (a) males rate partners' physical attractiveness to be appreciably more important to them than do females ($d = .54$), (b) females rate partners' socioeconomic status and ambitiousness to be much more important to them than do males ($ds = -.67$ to $-.69$), (c) females rate character (e.g., honesty and sincerity) and intelligence to be somewhat more important than do males ($ds = -.30$ to $-.35$), and (d) there are no notable gender differences in value accorded to sense of humor or colloquially defined "personality" ($ds = -.08$ to $-.14$). The mean of the absolute values of the seven effect sizes for gender differences in mate selection preferences was .40.

Oliver and Hyde (1993) recently examined sex differences in sexual attitudes and behaviors through meta-analysis. Most of their mean effect sizes were .30 or larger, and two sex differences (masturbation incidence and casual intercourse) were very large ($ds = .81$ to $.96$).

The largest gender differences of all have been found for physical abilities characteristics, with males greater than females on all examined measures. A review of meta-analytic findings by Linn and Hyde (1989) found mean effect sizes ranging from .09 to 2.60. Averaged over six dimensions, the mean effect size for the male advantage was 1.35.

Most important, meta-analyses of sex differences in the literature and from standardized personality inventory norms have both suggested that sex differences in personality have remained constant over the past generation and are generally about the same in samples obtained outside of the United States and Canada as in data sources obtained from the United States. The constancy of personality sex differences across generations, and cultures found in both studies reported here mirrors findings from research on gender differences in mate selection preferences (Buss, 1989; Feingold, 1992b). By comparison, the recent findings of cognitive gender differences are much smaller than those found in the past, at least for adolescents (Feingold, 1988a, 1993a; Linn, 1992; Wilder & Powell, 1989), and cognitive gender differences have also been found to exhibit marked variations across cultures (Born, Bleichrodt, & Van Der Flier, 1987; Feingold, 1994).

Directions for Future Research

Future work is needed that focuses on the causes of the variations in sex differences across different personality scales of the same constructs. For example, the hypothesis that Costa–McCrae facets can be divided into even more homogeneous subfacets across which gender differences vary, needs to be tested. Additional cross-cultural studies of sex differences in personality test norms from multiple inventories are also needed to examine the generalizability of the findings from the meta-analysis of international norms for the PRF. Because findings of sex differences in research conducted outside of the United States are readily available in the literature (as was suggested by the replication of Hall's meta-analysis), a comprehensive cross-cultural meta-analysis using such findings is also needed, although the limited cross-cultural analyses of sex differences in recent studies noted in Study 1 were consistent with Study 2's finding that effect sizes do not typically vary across nations.

In addition, because Study 1 tapped only a small sample of all available U.S. studies containing findings of sex differences in personality traits, there clearly is a need for additional meta-analyses of sex differences in the United States, particularly meta-analyses that examine gender differences in traits not examined in Study 1. Such meta-analyses could also examine possible moderation of gender differences by race and residence (e.g., urban vs. rural), which could not be examined in Study 2 because personality inventory norms had not been broken down by ethnicity or geography. Additional meta-analyses of sex differences in personality trait norms would also be valuable, especially considering that Study 2 examined only selected traits from a small sample of available archival test norms.

Finally, Feingold (1992c, in press) examined gender differences in cognitive abilities on standardized test norms and found that the implicit assumption of homogeneity of variance across sex was often rejected. Thus, similar work is needed to determine whether males and females vary equally on standardized tests of personality by using the procedures previously applied in the cognitive domain (see Feingold, 1992a, 1992c, 1993c; in press; Hedges & Friedman, 1993). Moreover, if there is heterogeneity of variance between the sexes for personality traits on which there are also sex differences in means, the joint effects of sex differences in central tendency and those in variability must be examined together to comprehend the differ-

ences between male and female distributions of personality trait scores.

References

Amatora, S. M. (1955). Comparisons in personality self-evaluation. *Journal of Social Psychology, 42,* 315–321.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences. *Psychological Bulletin, 111,* 256–274.

Anastasi, A. (1986). *Psychological testing* (6th ed.). New York: Macmillan.

Anderson, H. H. (1939). Domination and social integration in the behavior of kindergarten children in an experimental play situation. *Journal of Experimental Education, 8,* 123–131.

Ashmore, R. D., Del Boca, F. K., & Wohlers, A. J. (1986). Gender stereotypes. In R. D. Ashmore & F. K. Del Boca (Eds.), *The social psychology of female-male relations* (pp. 69–119). San Diego, CA: Academic Press.

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Chicago: Rand McNally.

Baltes, P. B., & Nesselroade, J. R. (1972). Cultural change and adolescent personal development: An application of longitudinal sequences. *Developmental Psychology, 7,* 244–256.

Bee, H. L. (1967). Parent-child interaction and distractibility in 9-year-old children. *Merrill-Palmer Quarterly, 13,* 175–190.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1982). *Differential aptitude tests Form V and W: Administrator's handbook.* New York: Psychological Corporation.

Benton, A. A., Gelber, E. R., Kelley, H. H., & Liebling, B. A. (1969). Reaction to various degrees of deceit in a mixed-motive relationship. *Journal of Personality and Social Psychology, 12,* 170–180.

Born, M. Ph., Bleichrodt, N., & Van Der Flier, H. (1987). Cross-cultural comparison of sex-related differences on intelligence tests: A meta-analysis. *Journal of Cross-Cultural Psychology, 18,* 283–314.

Bortner, R. W., & Hultsch, D. F. (1972). Personal time perspective in adulthood. *Developmental Psychology, 7,* 98–103.

Braginsky, D. D. (1970). Machiavellianism and manipulative interpersonal behavior in children. *Journal of Experimental Social Psychology, 6,* 77–99.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco: Jossey-Bass.

Christie, R. C. (1970a). Scale construction. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 10–34). San Diego, CA: Academic Press.

Christie, R. C. (1970b). Social correlates of Machiavellianism. In R. Christie & F. L. Geis (Eds.), *Studies in Machiavellianism* (pp. 314–338). San Diego, CA: Academic Press.

Christie, R., & Geis, F. L. (Eds.). (1970). *Studies in Machiavellianism.* San Diego, CA: Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). San Diego, CA: Academic Press.

Cohn, L. D. (1991). Sex differences in the course of personality development: A meta-analysis. *Psychological Bulletin, 109,* 252–266.

Cooley, C. H. (1900). *Human nature in the social order.* New York: Scribner.

Cooper, H. M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology, 37,* 131–146.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Crandall, V. C., Katkovsky, W., & Crandall, V. J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. *Child Development, 36,* 91–109.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35,* 867–881.

Deaux, K., & Kite, M. (1993). Gender stereotypes. In F. L. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 107–139). Westport, CT: Greenwood Press.

Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review, 94,* 369–389.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Erlbaum.

Eagly, A. H. (in press). The science and politics of comparing women and men. *American Psychologist.*

Eagly, A. H., & Carli, L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin, 90,* 1–20.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 283–308.

Eagly, A. H., & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic review of the literature. *Psychological Bulletin, 100,* 309–330.

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin, 17,* 306–315.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94,* 100–131.

Eysenck, H. J. (1992). Four ways five factors are *not* basic. *Personality and Individual Differences, 13,* 667–673.

Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality.* London: Hodder & Stoughton.

Fausto-Sterling, A. (1985). *Myths of gender: Biological theories about men and women.* New York: Basic Books.

Feather, N. T. (1969). Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology, 13,* 129–144.

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

Feingold, A. (1988b). Matching for attractiveness in romantic partners and same-sex friends: A meta-analysis and theoretical critique. *Psychological Bulletin, 104,* 226–235.

Feingold, A. (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59,* 981–993.

Feingold, A. (1991). Sex differences in the effects of similarity and physical attractiveness on opposite-sex attraction. *Basic and Applied Social Psychology, 12,* 357–367.

Feingold, A. (1992a). Cumulation of variance ratios. *Review of Educational Research, 62,* 433–434.

Feingold, A. (1992b). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological Bulletin, 112,* 125–139.

Feingold, A. (1992c). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304–341.

Feingold, A. (1992d). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84.

Feingold, A. (1993a). Cognitive gender differences: A developmental perspective. *Sex Roles, 29,* 91–112.

Feingold, A. (1993b). *Gender stereotyping: An examination using the bogus stranger paradigm.* Unpublished manuscript, Yale University, New Haven, CT.

Feingold, A. (1993c). Joint effects of gender differences in central tendency and gender differences in variability. *Review of Educational Research, 63,* 106–109.

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles, 30,* 81–92.

Feingold, A. (in press). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist.*

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case of emotional intensity. *Journal of Personality and Social Psychology, 61,* 427–434.

Gellert, E. (1962). The effect of changes in group composition on the dominant behavior of young children. *British Journal of Social and Clinical Psychology, 1,* 168–181.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5,* 3–8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Goldrich, J. M. (1967). A study in time orientation: The relation between memory for past experience and orientation to the future. *Journal of Personality and Social Psychology, 6,* 216–222.

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85,* 845–875.

Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Baltimore: Johns Hopkins University Press.

Hall, J. A., & Briton, N. J. (1993). Gender, nonverbal behavior, and expectations. In P. D. Blanck (Ed.), *Interpersonal expectations: Theory, research, and applications* (pp. 276–295). Cambridge, England: Cambridge University Press.

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

Hedges, L. V., & Friedman, L. (1993). Gender differences in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63,* 94–105.

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

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage.

Hyde, J. S. (1984). How large are gender differences in aggression? A developmental meta-analysis. *Developmental Psychology, 20,* 722–736.

Hyde, J. S., & Linn, M. C. (Eds.). (1986). *The psychology of gender: Advances through meta-analysis.* Baltimore: Johns Hopkins University Press.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93,* 429–445.

L'Abate, L. (1960). Personality correlates of manifest anxiety in children. *Journal of Consulting Psychology, 24,* 342–348.

Lepper, M. R. (1973). Dissonance, self-perception, and honesty. *Journal of Personality and Social Psychology, 25,* 65–74.

Linn, M. C. (1992). Gender differences in educational achievement. In J. Pfleiderer (Ed.), *Sex equity in educational opportunity, achievement, and testing* (pp. 11–50). Princeton, NJ: Educational Testing Service.

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

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

MacDonald, A. P., Jr. (1970). Anxiety, affiliation, and social isolation. *Developmental Psychology, 3,* 242–254.

Marsh, H. W. (1989). Sex differences in the development of verbal and mathematical constructs: The High School and Beyond Study. *American Educational Research Journal, 26,* 191–225.

McArthur, L. Z. (1982). Judging a book by its cover: A cognitive analysis of the relationship between physical appearance and stereotyping. In A. H. Hastorf & A. M. Isen (Eds.), *Cognitive social psychology* (pp. 149–211). Amsterdam: Elsevier/North-Holland.

Mead, G. H. (1934). *Mind, self, and society.* Chicago: University of Chicago Press.

Megaree, E. I. (1972). *The California Psychological Inventory handbook.* San Francisco: Jossey-Bass.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual Review of Psychology, 37,* 233–256.

Mullen, B. (1989). *Advanced BASIC meta-analysis.* Hillsdale, NJ: Erlbaum.

Murray, H. A., et al. (1938). *Explorations in personality.* New York: Oxford University Press.

Myers, D. G. (1992). *The pursuit of happiness: Who is happy and why.* New York: William Morrow.

Nachamie, S. (1969). *Machiavellianism in children: The children's Mach scale and the bluffing game.* Unpublished doctoral dissertation, Columbia University, New York.

Nawas, M. M. (1971). Change in efficiency of ego functioning and complexity from adolescence to young adulthood. *Developmental Psychology, 4,* 412–415.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression. *Psychological Bulletin, 101,* 259–282.

Oliver, M. B., & Hyde, J. S. (1993). Gender differences in sexuality: A meta-analysis. *Psychological Bulletin, 114,* 29–51.

Omark, D. R., & Edelman, M. (1973). *Peer group social interactions from an evolutionary perspective.* Paper presented at the meeting of the Society for Research in Child Development, Philadelphia.

Omark, D. R., Omark, M., & Edelman, M. (1973). *Dominance hierarchies in young children.* Paper presented at the International Congress of Anthropological and Ethological Sciences.

Perris, C. (1966). A study of bipolar (manic-depressive) and unipolar recurrent depressive psychosis. *Acta Psychiatrica Scandinavica* (Suppl. 194), 1–89.

Research Psychologists Press. (1993). [International norms for the PRF]. Unpublished raw data.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Newbury Park, CA: Sage.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart & Winston.

Ruble, D. N., & Ruble, T. L. (1982). Sex stereotypes. In A. G. Miller (Ed.), *In the eye of the beholder: Contemporary research in stereotyping* (pp. 188–252). New York: Praeger.

Schaie, K. W., & Strother, C. R. (1968). Cognitive and personality variables in college graduates of advanced age. In G. A. Talland (Ed.), *Human aging and behavior* (pp. 281–308). San Diego, CA: Academic Press.

Shields, S. A. (1975). Functionalism, Darwinism, and the psychology of women. *American Psychologist, 30,* 739–754.

Shrader, W. K., & Leventhal, T. (1968). Birth order of children and parental report of problems. *Child Development, 39,* 1165–1175.

Silverman, I., Shulman, A. D., & Wiesenthal, D. L. (1970). Effects of deceiving and debriefing psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology, 14,* 203–212.

Solomon, D., Houlihan, K. A., & Parelius, R. J. (1969). Intellectual

achievement responsibility in Negro and White children. *Psychological Reports, 24,* 479–483.

Strongman, K. T., & Champness, B. G. (1968). Dominance hierarchies and conflict in eye contact. *Acta Psychologica, 28,* 376–386.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception. *Psychological Review, 91,* 457–477.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66,* 21–36.

Templer, D. I., Ruff, C. F., & Franks, C. M. (1971). Death anxiety: Age, sex, and parental resemblance in diverse populations. *Developmental Psychology, 4,* 108.

Wilder, G. Z., & Powell, K. (1989). *Sex differences in test performance: A review of the literature* (College Board Report No. 89-3). New York: College Entrance Examination Board.

Winokur, G., & Tanna, V. L. (1969). Possible role of X-linked dominant factor in manic-depressive disease. *Diseases of the Nervous System, 30,* 89–93.

Wood, W., Rhodes, N., & Whelan, M. (1989). Sex differences in positive well-being: A consideration of emotional style and marital status. *Psychological Bulletin, 106,* 249–264.

Zuckerman, M. (1991). *Psychobiology of personality.* Cambridge, England: Cambridge University Press.

Appendix A

Studies Used in the Meta-Analysis of Maccoby and Jacklin's (1974) Database (Study 1)

Anderson, H. H. (1937). Domination and integration in the social behavior of young children in an experimental play situation. *Genetic Psychology Monographs, 19,* 341–408.

Arkoff, A., Meredith, G., & Iwahara, S. (1962). Dominance-patterning in motherland-Japanese, Japanese-American, and Caucasian-American students. *Journal of Social Psychology, 58,* 61–66.

Baltes, P. B., & Nesselroade, J. R. (1972). Cultural change and adolescent personality development: An application of longitudinal sequences. *Developmental Psychology, 7,* 244–256.

Barton, K. (1971). Block manipulation by children as a function of social reinforcement, anxiety, arousal, and ability pattern. *Child Development, 38,* 817–826.

Baumrind, D., & Black, A. E. (1967). Socialization practices associated with dimensions of competence in preschool boys and girls. *Child Development, 38,* 291–327.

Benton, A. A., Gelber, E. R., Kelley, H. H., & Liebling, B. A. (1969). Reaction to various degrees of deceit in a mixed-motive relationship. *Journal of Personality and Social Psychology, 12,* 170–180.

Bledsoe, J. C. (1961). Sex differences in mental health analysis scores of elementary pupils. *Journal of Consulting Psychology, 25,* 364–365.

Bledsoe, J. C. (1967). Self-concepts of children and their intelligence, achievement, interests, and anxiety. *Childhood Education, 43,* 436–438.

Brannigan, G. G., & Tolor, A. (1971). Sex differences in adaptive styles. *Journal of Genetic Psychology, 119,* 143–149.

Buck, M. R., & Austrin, H. R. (1971). Factors related to school achievement in an economically disadvantaged group. *Child Development, 42,* 1813–1826.

Carlson, R. (1965). Stability and change in the adolescent's self-image. *Child Development, 36,* 659–666.

Carpenter, T. R., & Busse, T. V. (1969). Development of self concept in Negro and White welfare children. *Child Development, 40,* 935–939.

Coopersmith, S. (1959). A method of determining types of self-esteem. *Journal of Abnormal and Social Psychology, 59,* 87–94.

Coopersmith, S. (1967). *The antecedents of self-esteem.* Palo Alto, CA: Consulting Psychologists Press.

Cowen, E. L., & Danset, A. (1962). Étude comparée des réponses d'ecoliers français et américains à une échelle d'anxiété. *Revue de Psychologies Applique, 12,* 263–274.

Cowen, E. L., Zax, M., Klein, R., Izzo, L. D., & Trost, M. A. (1965). The relation of anxiety in school children to school record, achievement, and behavioral measures. *Child Development, 36,* 685–695.

Crandall, V. C., Katkovsky, W., & Crandall, V. J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. *Child Development, 36,* 91–109.

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.

Denmark, F. L., & Diggory, J. C. (1966). Sex differences in attitudes toward leader's display of authoritarian behavior. *Psychological Reports, 18,* 863–872.

Dweck, C. S., & Reppucci, N. D. (1973). Learned hopelessness and reinforcement responsibility in children. *Journal of Personality and Social Psychology, 25,* 109–116.

Emmerich, W. (1971). *Structure and development of personal-social behaviors in preschool settings* (Research report, Head Start Longitudinal Study). Princeton, NJ: Educational Testing Service.

Feather, N. T. (1969). Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology, 13,* 129–144.

Feshback, N. D. (1969). Sex differences in children's modes of aggressive responses toward outsiders. *Merrill-Palmer Quarterly, 15,* 249–278.

Gardiner, H. W. (1968). Dominance-deference patterning in Thai students. *Journal of Social Psychology, 76,* 281–282.

Goldschmid, M. L. (1968). The relation of conservation to emotional and environmental aspects of development. *Child Development, 39,* 579–589.

Grams, A., Hafner, A. J., & Quast, W. (1965). Children's anxiety: Self-estimates, parent reports, and teacher ratings. *Merrill-Palmer Quarterly, 11,* 261–266.

Hafner, A. J., & Kaplan, A. M. (1959). Children's manifest anxiety and intelligence. *Child Development, 30,* 269–271.

Hanna, F., Storm, T., & Caird, W. K. (1965). Sex differences and relationships among neuroticism, extraversion, and expressed fears. *Perceptual and Motor Skills, 20,* 1214–1216.

Harris, S., & Braun, J. R. (1971). Self-esteem and racial preference in Black children. *Proceedings of the 79th Annual Convention of the American Psychological Association, 6,* 259–260.

Harrison, C. W., Rawls, J. R., & Rawls, D. J. (1971). Differences between leaders and nonleaders in six to eleven-year-old children. *Journal of Social Psychology, 84,* 269–272.

Herbert, E. W., Gelfand, D. M., & Hartmann, D. P. (1969). Imitation

and self-esteem as determinants of self-critical behavior. *Child Development, 40,* 421–430.

Holloway, H. D. (1958). Reliability of the Children's Manifest Anxiety Scale at the rural third grade level. *Journal of Educational Psychology, 4,* 193–196.

Iwawaki, S., Sumida, K., Okuno, S., & Cowen, E. L. (1967). Manifest anxiety in Japanese, French, and United States children. *Child Development, 38,* 713–722.

Jacobson, L. I., Berger, S. E., & Millham, J. (1969). Self-esteem, sex differences, and the tendency to cheat. *Proceedings of the 77th Annual Convention of the American Psychological Association, 4,* 353–354.

Kaplan, H. B. (1973). Self-derogation and social position: Interaction effects of sex, race, education, and age. *Social Psychiatry, 8,* 92–99.

Kidd, A. H., & Cherymisin, D. G. (1965). Figural reversal as related to specific personality variables. *Perceptual and Motor Skills, 20,* 1175–1176.

Klaus, R. A., & Gray, S. W. (1968). The early training project for disadvantaged children: A report after five years. *Monographs of the Society for Research in Child Development, 33*(4, Serial No. 120).

Koenig, K. P. (1966). Verbal behavior and personality change. *Journal of Personality and Social Psychology, 3,* 223–227.

L'Abate, L. (1960). Personality correlates of manifest anxiety in children. *Journal of Consulting Psychology, 24,* 342–348.

Lekarczyk, D. T., & Hill, K. T. (1969). Self-esteem, test anxiety, stress, and verbal learning. *Developmental Psychology, 1,* 147–154.

Levy, P., Lundgren, D., Ansel, M., Fell, D., Fink, B., & McGrath, J. E. (1972). Bystander effect in a demand-without-threat situation. *Journal of Personality and Social Psychology, 24,* 166–171.

Long, B. H., Henderson, E. H., & Ziller, R. C. (1967). Developmental changes in self-concept during middle childhood. *Merrill-Palmer Quarterly, 13,* 201–215.

Lott, B. E., & Lott, A. J. (1968). The relation of manifest anxiety in children to learning task performance and other variables. *Child Development, 39,* 207–220.

MacMillan, D. L., & Keogh, B. K. (1971a). Effect of instructional set on twelve-year-old children's perception of interruptions. *Developmental Psychology, 4,* 106.

MacMillan, D. L., & Keogh, B. K. (1971b). Normal and retarded children's expectancy for failure. *Developmental Psychology, 4,* 343–348.

Markel, N. N., Prebor, L. D., & Brandt, J. F. (1972). Biosocial factors in dyadic communication: Sex and speaking intensity. *Journal of Personality and Social Psychology, 23,* 11–13.

Mendelsohn, G. A., & Griswold, B. B. (1967). Anxiety and repression as predictors of the use of incidental cues in problem solving. *Journal of Personality and Social Psychology, 6,* 353–359.

Nisbett, R. E., & Gordon, A. (1967). Self-esteem and susceptibility to social influence. *Journal of Personality and Social Psychology, 5,* 268–276.

Palermo, D. S. (1959). Racial comparions and additional normative data on the Children's Manifest Anxiety Scale. *Child Development, 30,* 53–57.

Pallak, M. S., Brock, T. C., & Kiesler, C. A. (1967). Dissonance arousal and task performance in an incidental verbal learning paradigm. *Journal of Personality and Social Psychology, 7,* 11–20.

Parten, M. B. (1933). Leadership among preschool children. *Journal of Abnormal and Social Psychology, 27,* 430–440.

Penney, R. K. (1965). Reactive curiosity and manifest anxiety in children. *Child Development, 36,* 697–702.

Sarason, I. G., & Koenig, K. P. (1965). Relationships of test anxiety and hostility to description of self and parents. *Journal of Personality and Social Psychology, 2,* 617–621.

Sarason, I. G., & Winkel, G. H. (1966). Individual differences among subjects and experimenters and subjects' self-descriptions. *Journal of Personality and Social Psychology, 3,* 448–457.

Schaie, K. W., & Strother, C. R. (1968). Cognitive and personality variables in college graduates of advanced age. In G. A. Talland (Ed.), *Human aging and behavior* (pp. 281–308). San Diego, CA: Academic Press.

Sharma, K. L. (1969). Dominance-deference: A cross-cultural study. *Journal of Social Psychology, 70,* 265–266.

Shipman, V. C. (1971). *Disadvantaged children and their first school experience* (Research report, Head Start Longitudinal Study). Princeton, NJ: Educational Testing Service.

Silverman, I., Shulman, A. D., & Wiesenthal, D. L. (1970). Effects of deceiving and debriefing psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology, 14,* 203–212.

Skolnick, P. (1971). Reactions to personal evaluations: A failure to replicate. *Journal of Personality and Social Psychology, 18,* 62–67.

Strongman, K. T., & Champness, B. G. (1968). Dominance hierarchies and conflict in eye contact. *Acta Psychologica, 28,* 376–386.

Sutton-Smith, B., & Savasta, M. (1972, April). *Sex differences in play and power.* Paper presented at the annual meeting of the Eastern Psychological Association, Boston.

Szal, J. A. (1972). *Sex differences in the cooperative and competitive behaviors of nursery school children.* Unpublished master's thesis, Stanford University, Stanford, CA.

Vassiliou, V., Georgas, J. G., & Vassiliou, G. (1967). Variations in manifest anxiety due to sex, age, and education. *Journal of Personality and Social Psychology, 6,* 194–197.

Walls, R. T., & Cox, J. (1971). Disadvantaged and non-disadvantaged children's expectancy in skill and chance outcomes. *Developmental Psychology, 4,* 299.

Whiting, B., & Edwards, C. P. (1973). A cross-cultural analysis of sex differences in the behavior of children aged three through 11. *Journal of Social Psychology, 91,* 171–188.

Zander, A., Fuller, R., & Armstrong, W. (1972). Attributed pride or shame in group and self. *Journal of Personality and Social Psychology, 23,* 346–352.

Zander, A., & Van Egmond, E. (1958). Relationship of intelligence and social power to interpersonal behavior of children. *Journal of Educational Psychology, 49,* 257–268.

Zytkoskee, A., Strickland, B. R., & Watson, J. (1971). Delay of gratification and internal vs. external control among adolescents of low socioeconomic status. *Developmental Psychology, 4,* 93–98.

Appendix B

Studies Used in the Replication of Hall's (1984) Meta-Analysis (Study 1)

Allgood-Merten, B., & Stockard, J. (1991). Sex role identity and self-esteem: A comparison of children and adolescents. *Sex Roles, 25,* 129–139.

Ben-Zur, H., & Zeidner, M. (1988). Sex differences in anxiety, curiosity, and anger: A cross-cultural study. *Sex Roles, 19,* 335–347.

Bernstein, J., & Carmel, S. (1991). Gender differences over time in medical school stressors, anxiety, and the sense of coherence. *Sex Roles, 24,* 335–344.

Biaggio, M. K., Mohan, P. J., & Baldwin, C. (1985). Relationships among attitudes toward children, women's liberation, and personality characteristics. *Sex Roles, 12,* 47–62.

Brenner, J. B., & Cunningham, J. G. (1992). Gender differences in eating attitudes, body concept, and self-esteem among models. *Sex Roles, 27,* 413–437.

Bruch, M. A., Gorsky, J. M., Collins, T. M., & Berger, P. A. (1989). Shyness and sociability reexamined: A multicomponent study. *Journal of Personality and Social Psychology, 57,* 904–915.

Carlson, H. M., & Baxter, L. A. (1984). Androgyny, depression, and self-esteem in Irish homosexual and heterosexual males and females. *Sex Roles, 10,* 457–467.

de Jong-Gierveld, J. (1987). Developing and testing a model of loneliness. *Journal of Personality and Social Psychology, 53,* 119–128.

Endler, N. S., Cox, B. J., Parker, J. D. A., & Bagby, R. M. (1992). Self-reports of depression and state-trait anxiety: Evidence of differential assessment. *Journal of Personality and Social Psychology, 63,* 832–838.

Endler, N. S., Parker, J. D. A., Bagby, R. M., & Cox, B. J. (1991). Multidimensionality of state and trait anxiety: Factor structure of the Endler Multidimensional Anxiety Scales. *Journal of Personality and Social Psychology, 60,* 919–926.

Feather, N. T. (1985). Masculinity, femininity, self-esteem, and subclinical depression. *Sex Roles, 12,* 491–500.

Gurman, E. B., & Long, K. (1992). Gender orientation and emergent leader behavior. *Sex Roles, 27,* 391–400.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60,* 895–910.

Hendryx, M. S., Haviland, M. G., & Shaw, D. G. (1991). Dimensions of alexithymia and their relationships to anxiety and depression. *Journal of Personality Assessment, 56,* 227–237.

Ingram, R. E., Cruet, D., Johnson, B. R., & Wisnicki, K. S. (1988). Self-focused attention, gender, gender role, and vulnerability to negative affect. *Journal of Personality and Social Psychology, 55,* 967–978.

Jones, W. H., Briggs, S. R., & Smith, T. G. (1986). Shyness: Conceptualization and measurement. *Journal of Personality and Social Psychology, 51,* 629–639.

Lau, S. (1989). Sex role orientation and domains of self-esteem. *Sex Roles, 21,* 415–422.

Lorr, M., Youniss, R. P., & Stefic, E. C. (1991). Inventory of social skills. *Journal of Personality Assessment, 57,* 506–520.

Lortie-Lussier, M., Simond, S., Rinfret, N., & De Konick, J. (1992). Beyond sex differences: Family and occupational roles' impact on women's and men's dreams. *Sex Roles, 26,* 79–96.

Marsh, H. W., Antill, J. K., & Cunningham, J. D. (1987). Masculinity, femininity, and androgyny: Relations to self-esteem and social desirability. *Journal of Personality, 55,* 625–663.

Marsh, H. W., & Byrne, B. M. (1991). Differentiated additive androgyny model: Relations between masculinity, femininity, and multiple dimensions of self-concept. *Journal of Personality and Social Psychology, 61,* 811–828.

McCann, B. S., Woolfolk, R. L., Lehrer, P. M., & Schwarcz, L. (1987). Gender differences in the relationship between hostility and the Type A behavior pattern. *Journal of Personality Assessment, 51,* 355–366.

Nystedt, L., & Smari, J. (1989). Assessment of the Fenigstein, Scheier, and Buss Self-Consciousness Scale: A Swedish translation. *Journal of Personality Assessment, 53,* 342–352.

O'Heron, C. A., & Orlofsky, J. L. (1990). Stereotypic and nonstereotypic sex role trait and behavior orientations, gender identity, and psychological adjustment. *Journal of Personality and Social Psychology, 58,* 134–143.

Orlofsky, J. L., & O'Heron, C. A. (1987). Stereotypic and nonstereotypic sex role trait and behavior orientations: Implications for personal adjustment. *Journal of Personality and Social Psychology, 52,* 1034–1042.

Payne, F. D. (1987). "Masculinity," "femininity," and the complex construct of adjustment. *Sex Roles, 17,* 359–374.

Pennebaker, J. W., Colder, M., & Sharp, L. K. (1990). Accelerating the coping process. *Journal of Personality and Social Psychology, 58,* 528–537.

Piedmont, R. L., McCrae, R. R., & Costa, P. T., Jr. (1992). An assessment of the Edwards Personal Preference Schedule from the perspective of the five-factor model. *Journal of Personality Assessment, 58,* 67–78.

Raskin, R., & Novacek, J. (1989). An MMPI description of the narcissistic personality. *Journal of Personality Assessment, 53,* 66–80.

Reynolds, W. M. (1988). Measurement of academic self-concept in college students. *Journal of Personality Assessment, 52,* 223–240.

Rice, R. W., Yoder, J. D., Adams, J., Priest, R. F., & Prince, H. T. II. (1984). Leadership ratings for male and female military cadets. *Sex Roles, 10,* 885–901.

Robinson, E. A., & Follingstad, D. R. (1985). Development and validation of a behavioral sex-role inventory. *Sex Roles, 13,* 691–713.

Rowlison, R. T., & Felner, R. D. (1988). Major life events, hassles, and adaptation in adolescence: Confounding of conceptualization and measurement of life stress and adjustment revisited. *Journal of Personality and Social Psychology, 55,* 432–444.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57,* 1069–1081.

Santelli, J., Bernstein, D. M., Zborowski, L., & Bernstein, J. M. (1990). Pursuing and distancing, and related traits: A cross-cultural assessment. *Journal of Personality Assessment, 55,* 663–672.

Sawrie, S. M., Watson, P. J., & Biderman, M. D. (1991). Aggression, sex role measures, and Kohut's psychology of the self. *Sex Roles, 25,* 141–161.

Shepperd, J. A., & Kashani, J. H. (1991). The relationship of hardiness, gender, and stress to health outcomes in adolescents. *Journal of Personality, 59,* 747–768.

Stein, J. A., Newcomb, M. D., & Bentler, P. M. (1992). The effect of agency and communality on self-esteem: Gender differences in longitudinal data. *Sex Roles, 26,* 465–483.

Vollmer, F. (1984). Sex differences in personality and expectancy. *Sex Roles, 11,* 1121–1139.

(Appendix continues on next page)

Watson, P. J., Taylor, D., & Morris, R. J. (1987). Narcissism, sex roles, and self-functioning. *Sex Roles, 16,* 335–350.

Zeldow, P. B., Clark, D., & Daugherty, S. R. (1985). Masculinity, femininity, Type A behavior, and psychosocial adjustment in medical students. *Journal of Personality and Social Psychology, 48,* 481–492.

Zuckerman, D. M. (1985). Confidence and aspirations: Self-esteem and self-concepts as predictors of students' life goals. *Journal of Personality, 53,* 543–560.

Appendix C
Sources of U.S. Test Norms Used in the Meta-Analysis (Study 2)

Archer, R. P. (1992). *MMPI-A: Assessing adolescent psychopathology.* Hillsdale, NJ: Erlbaum.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco: Jossey-Bass.

Cattell, R. B., Cattell, M. D., & Johns, E. (1984). *Manual and norms for the High School Personality Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B., & Eber, H. W. (1962). *Supplement of norms for Forms A and B of the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B., & Scheier, I. H. (1963). *Handbook for the IPAT Anxiety Scale Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Comrey, A. L. (1970). *EDITS manual for the Comrey Personality Scales.* San Diego, CA: EDITS.

Costa, P. T., Jr., & McCrae, R. R. (1989). *NEO PI/FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1975). *An MMPI handbook. Vol. 2. Research applications.* Minneapolis: University of Minnesota Press.

Edwards, A. L. (1959). *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation.

Eysenck, H. J. (1962). *Manual for the Maudsley Personality Inventory.* San Diego, CA: EDITS.

Eysenck, H. J., & Eysenck, S. B. G. (in press). *Manual for the Eysenck Personality Questionnaire-Revised.* San Diego, CA: EDITS.

Gordon, L. V. (1978). *Manual for the Gordon Personal Profile.* New York: Psychological Corporation.

Gough, H. G. (1957). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1987). *California Psychological Inventory: Administrator's guide.* Palo Alto, CA: Consulting Psychologists Press.

Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual.* Needham Heights, MA: Allyn & Bacon.

Guilford, J. S., Zimmerman, W. S., & Guilford, J. P. (1976). *The Guilford-Zimmerman Temperament Survey handbook.* San Diego, CA: EDITS.

Institute for Personality and Ability Testing. (1958). *Tabular supplement to the 16 Personality Factor Questionnaire handbook.* Champaign, IL: Author.

Institute for Personality and Ability Testing. (1970). *Tabular supplement No. 1: Norms for the 16 PF, Forms A and B (1967–68 edition).* Champaign, IL: Author.

Jackson, D. N. (1974). *Personality Research Form manual.* Port Huron, MI: Research Psychologists Press.

Received May 12, 1993
Revision received May 9, 1994
Accepted May 10, 1994 ■