British Journal of Psychology (2005), 96, 505–524
*© 2005 The British Psychological Society*
The British Psychological Society
www.bpsjournals.co.uk

# Sex differences in means and variability on the progressive matrices in university students: A meta-analysis

Paul Irwing[1]* and Richard Lynn[2]
[1]University of Manchester, UK
[2]University of Ulster, UK

A meta-analysis is presented of 22 studies of sex differences in university students of means and variances on the Progressive Matrices. The results disconfirm the frequent assertion that there is no sex difference in the mean but that males have greater variability. To the contrary, the results showed that males obtained a higher mean than females by between *.22d* and *.33d*, the equivalent of 3.3 and 5.0 IQ conventional points, respectively. In the 8 studies of the SPM for which standard deviations were available, females showed significantly greater variability ($F_{(882,656)} = 1.20$, $p < .02$), whilst in the 10 studies of the APM there was no significant difference in variability ($F_{(3344,5660)} = 1.00$, $p > .05$).

It has frequently been asserted that there is no sex difference in average general intelligence but that the variance is greater in males. In this paper we examine these two propositions by a meta-analysis of studies of sex differences on the Progressive Matrices among university students. We find that both are incorrect.

The assertion that there is no sex difference in average general intelligence has been made repeatedly since the early decades of the twentieth century. One of the first to adopt this position was Terman (1916, pp. 69–70) who wrote of the American standardization sample of the Stanford–Binet test on approximately 1,000 4- to 16-year-olds that girls obtained a slightly higher average IQ than boys but 'the superiority of girls over boys is so slight . . . that for practical purposes it would seem negligible'. In the next decade Spearman (1923) asserted that there is no sex difference in *g*. Cattell (1971, p. 131) concluded that, 'it is now demonstrated by countless and large samples that on the two main general cognitive abilities – fluid and crystallized intelligence – men and women, boys and girls, show no significant differences'. Brody (1992, p. 323) contended that, 'gender differences in general intelligence are

*Correspondence should be addressed to Paul Irwing, Manchester, Business School East, The University of Manchester, Booth Street West, Manchester, M15 6PB (e-mail: paul.irwing@mbs.ac.uk).*

DOI:10.1348/000712605X53542

small and virtually non-existent'. Jensen (1998, p. 531) calculated sex differences in *g* on five samples and concluded that, 'no evidence was found for sex differences in the mean level of *g*'. Similarly: 'there is no sex difference in general intelligence worth speaking of' (Mackintosh, 1996, p. 567); 'the overall pattern suggests that there are no sex differences, or only a very small advantage of boys and men, in average IQ scores' (Geary, 1998, p. 310); 'most investigators concur on the conclusion that the sexes manifest comparable means on general intelligence' (Lubinski, 2000, p. 416); 'sex differences have not been found in general intelligence' (Halpern, 2000, p. 218); 'we can conclude that there is no sex difference in general intelligence' (Colom, Juan-Espinosa, Abad, & Garcia, 2000, p. 66); 'there are no meaningful sex differences in general intelligence' (Lippa, 2002); 'there are negligible differences in general intelligence' (Jorm, Anstey, Christensen, & Rodgers, 2004, p. 7); and 'the evidence that there is no sex difference in general ability is overwhelming' (Anderson, 2004, p. 829).

The question of whether there is a sex difference in average general intelligence raises the problem of how general intelligence should be defined. There have been three principal answers to this question. First, general intelligence can be defined as the IQ obtained on omnibus intelligence tests such as the Wechslers. The IQ obtained from these is the average of the scores on a number of different abilities including verbal comprehension and reasoning, immediate memory, visualization, and spatial and perceptual abilities. This is the definition normally used by educational, clinical, and occupational psychologists. When this definition is adopted, it has been asserted by Halpern (2000) and reaffirmed by Anderson (2004, p. 829) that 'the overall score does not show a sex difference' Halpern (2000, p. 90). Second, general intelligence can be defined as reasoning ability or fluid intelligence. This definition has been adopted by Mackintosh (1996, p. 564; Mackintosh, 1998a) who concludes that there is no sex difference in reasoning ability. Third, general intelligence can be defined as the *g* obtained as the general factor derived by factor analysis from a number of tests. This definition was initially proposed by Spearman (1923, 1946) and was first adopted to analyse whether there is a sex difference in *g* by Jensen and Reynolds (1983). They analysed the American standardization of the WISC-R on 6- to 16-year-olds and found that this showed boys to have a higher *g* by $d = .16$ (standard deviation units), equivalent to 2.4 IQ points (this advantage is highly statistically significant). In a second study of this issue using a different method for measuring *g*, Jensen (1998, p. 539) analysed five data sets and obtained rather varied results, in three of which males obtained a higher *g* than females by 2.83, 0.18, and 5.49 IQ points, while in two of which females obtained a higher *g* than males by 7.91 and 0.03 IQ points. Jensen handled these discrepancies by averaging the five results to give a negligible male advantage of .11 IQ points, from which he concluded that there is no sex difference in *g*. This conclusion has been endorsed by Colom and his colleagues in Spain (Colom, Garcia, Juan-Espinoza, & Abad, 2002, Colom *et al.*, 2000).

Thus there has evolved a widely held consensus that there is no sex difference in general intelligence, whether this is defined as the IQ from an omnibus intelligence test, as reasoning ability, or as Spearman's *g*. However, this consensus has not been wholly unanimous. Dissent from the consensus that there is no sex difference in general intelligence has come from Lynn (1994, 1998, 1999). The starting-point of his dissenting position was the discovery by Ankney (1992) and Rushton (1992), using measures of external cranial capacity, that men have larger average brain tissue volume than women, even when allowance is made for the larger male body size (see also, Gur *et al.*, 1999).

That men have a larger cerebrum than women by about 8–10% is now well established by studies using the more precise procedure of magnetic resonance imaging (Filipek, Richelme, Kennedy, & Caviness, 1994; Nopoulos, Flaum, O'Leary, & Andreasen, 2000; Passe *et al.*, 1997; Rabinowicz, Dean, Petetot, & de Courtney-Myers, 1999; Witelson, Glezer, & Kigar, 1995). The further link in the argument is that brain volume is positively associated with intelligence at a correlation of .40. This has been shown by Vernon, Wickett, Bazana, and Stelmack (2000, p. 248) in a summary of 14 studies in which magnetic resonance imaging was used to estimate brain size, as compared with previous studies based on measures of external cranial capacity (Jensen & Sinha, 1993). It appears to follow logically that males should have higher average intelligence than females attributable to their greater average brain tissue volume. It has been argued by Lynn (1994, 1998, 1999) that this is the case from the age of 16 onwards and that the higher male IQ is present whether general intelligence is defined as reasoning ability or the full scale IQ of the Wechsler tests, and that the male advantage reaches between 3.8 and 5 IQ points among adults. Lynn maintains that before the age of 16, males and females have approximately the same IQs because of the earlier maturation of girls. Lynn attributes the consensus that there is no sex difference in general intelligence to a failure to note this age effect. Further evidence supporting the theory that males obtain higher IQs from the age of 16 years has been provided by Lynn, Allik, and Must (2000), Lynn, Allik, Pullmann, and Laidra (2002), Lynn, Allik, and Irwing (2004), Lynn and Irwing (2004) and Colom and Lynn (2004). The only independent support for Lynn's thesis has come from Nyborg (2003, p. 209), who has reported on the basis of research carried out in Denmark that there is no significant sex difference in *g* in children, but among adults, men have a significant advantage of 5.55 IQ points. Most authorities, however, have rejected this thesis, including Mackintosh (1996, p. 564, Mackintosh (1998a, 1998b), Jensen (1998, p. 539), Halpern (2000), Anderson (2004, p. 829) and Colom and his colleagues (Colom *et al.*, 2002, 2000).

There is a large amount of research literature on sex differences in intelligence including books by Caplan, Crawford, Hyde, and Richardson (1997), Kimura (1999) and Halpern (2000). This makes integrating all the studies and attempting to find a solution to the problem of sex differences an immense task. We believe that the way to handle this is to examine systematically the studies of sex differences in intelligence in each of the three ways defined above (i.e. as the IQ in omnibus tests, in reasoning, and in *g*). To make a start on this research programme we have undertaken the task of examining whether there are sex differences in non-verbal reasoning assessed by Raven's Progressive Matrices. The Progressive Matrices is a particularly useful test on which to examine sex differences in intelligence defined as reasoning ability, because it is one of the leading and most frequently used tests of this ability. The test has been described as 'the paradigm test of non-verbal, abstract reasoning ability' (Mackintosh, 1996 p. 564). It is also widely regarded as the best or one of the best tests of Spearman's *g*, the general factor underlying all cognitive abilities. This was asserted by Spearman himself (1946) and confirmed in an early study by Rimoldi (1948).

By the early 1980s Court (1983, p. 54) was able to write that it is 'recognised as perhaps the best measure of *g*'; and some years later Jensen (1998, p. 541) wrote, that 'the Raven tests, compared with many others, have the highest *g* loading'. In a recent factor analysis of the Standard Progressive Matrices administered to 2,735 12- to 18-year-olds in Estonia, Lynn *et al.* (2004) showed the presence of three primary factors, and a higher order factor, which was interpreted as g, and correlated at .99 with total scores, providing further support that scores on the Progressive matrices can be identified with *g*. Sex differences on the Progressive Matrices should, therefore, reveal whether there is a difference in *g* as well as whether there is a sex difference in reasoning ability.

The Progressive Matrices test consists of a series of designs that form progressions. The problem is to understand the principle governing the progression and then to extrapolate this to identify the next design from a choice of six or eight alternatives. The first version of the test was the Standard Progressive Matrices constructed in the late 1930s as a test of non-verbal reasoning ability and of Spearman's *g* (Raven, 1939) for the ages of 6 years to adulthood. Subsequently, the Coloured Progressive Matrices was constructed as an easier version of the test designed for children aged 5 through 12; and later the Advanced Progressive Matrices was constructed as a harder version of the test designed for older adolescents and adults with higher ability.

The issue of whether there are any sex differences on the Progressive Matrices has frequently been discussed and it has been virtually universally concluded that there is no difference in the mean scores obtained by males and females. This has been one of the major foundations for the conclusion that there is no sex difference in reasoning ability or in *g*, of which the Progressive Matrices is widely regarded as an excellent measure. The first statement that there is no sex difference on the test came from Raven himself, who constructed the test and wrote that in the standardization sample 'there was no sex difference, either in the mean scores or the variance of scores, between boys and girls up to the age of 14 years. There were insufficient data to investigate sex differences in ability above the age of 14' (Raven, 1939, p. 30).

The conclusion that there is no sex difference on the Progressive Matrices has been endorsed by numerous scholars. Eysenck (1981, p. 41) stated that the tests 'give equal scores to boys and girls, men and women'. Court (1983) reviewed 118 studies of sex differences on the Progressive Matrices and concluded that most showed no difference in mean scores, although some showed higher mean scores for males and others found higher mean scores for females. From this he concluded that 'there is no consistent difference in favour of either sex over all populations tested. . .the most common finding is of no sex difference. Reports which suggest otherwise can be shown to have elements of bias in sampling' (p. 62) and that 'the accumulated evidence at all ability levels indicates that a biological sex difference cannot be demonstrated for performance on the Raven's Progressive Matrices' (p. 68). This conclusion has been accepted by Jensen (1998) and by Mackintosh (1996, 1998a, 1998b). Thus, Jensen (1998, p. 541) writes, 'there is no consistent difference on the Raven's Standard Progressive Matrices (for adults) or on the Coloured Progressive Matrices (for children)'. Mackintosh (1996, p. 564) writes, 'large scale studies of Raven's tests have yielded all possible outcomes, male superiority, female superiority and no difference', from which he concluded that 'there appears to be no difference in general intelligence' (Mackintosh (1998a, p. 189). More recently, Anderson (2004, p. 829) has reaffirmed that 'there is no sex difference in general ability. . . whether this is defined as an IQ score calculated from an omnibus test of intellectual abilities such as the various Wechsler tests, or whether it is defined as a score on a single test of general intelligence, such as the Raven's Matrices'.

While Jensen and Mackintosh have both relied on Court's (1983) conclusion, based on his review of 118 studies, that there are no sex differences on the Progressive Matrices, Court's review cannot be accepted as a satisfactory basis for the conclusion that no sex differences exist. The review has at least three deficiencies. First, it is more than 20 years old and a number of important studies of this issue have appeared subsequently and need to be considered. Second, the review lists studies of a variety of convenience samples (e.g. psychiatric patients, deaf children, retarded children, shop assistants, clerical workers, college students, primary school children, secondary school children, Native Americans, and Inuit) many of which cannot be regarded as representative of males and females. Third, Court does not provide information on the sample sizes for approximately half of the studies he lists, and where information on sample sizes is given, the numbers are generally too small to establish whether there is a significant sex difference. To detect a statistically significant difference of 3 to 5 IQ points requires a sample size of around 500. Court's review gives only one study of adults with an adequate sample size of this number or more. This is Heron and Chown's (1967) study of a sample of 600 adults in which men obtained a higher mean score than women of approximately 2.8 IQ points. For all these reasons, Court's review cannot be accepted as an adequate basis for the contention that there is no sex difference on the Progressive Matrices.

During the last quarter century it has become widely accepted that the best way to resolve issues on which there are a large number of studies is to carry out a meta-analysis. The essentials of this technique are to collect all the studies on the issue, convert the results to a common metric and average them to give an overall result. This technique was first proposed by Glass (1976) and by the end of the 1980s had become accepted as a useful method for synthesizing the results of many different studies. Thus, an epidemiologist has described meta-analysis as 'a boon for policy makers who find themselves faced with a mountain of conflicting studies' (Mann, 1990, p. 476). By the late 1990s, over 100 meta-analyses a year were being reported in *Psychological Abstracts.* Perhaps surprisingly, there have been no meta-analyses of sex differences in reasoning ability (apart from our own presented in Lynn & Irwing, 2004), although there have been meta-analyses of sex differences in several cognitive abilities including verbal abilities (Hyde & Linn,1988), spatial abilities (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995) and mathematical abilities (Hyde, Fennema, & Lamon, 1990). To tackle the question of whether there is a sex difference in reasoning abilities, we have carried out a meta-analysis of sex differences on the Progressive Matrices in general population samples. The conclusion of this study was that there is no sex difference among children between the ages of 6 and 15, but that males begin to secure higher average means from the age of 16 and have an advantage of 5 IQ points from the age of 20 (Lynn & Irwing, 2004). In view of the repeated assertion that there is no sex difference on the Progressive Matrices by such authorities as Eysenck (1981), Court (1983), Jensen (1998), Mackintosh (1996, 1998a, 1998b), and Anderson (2004), we anticipate that our conclusion that among adults, men have a 5 IQ point advantage on this test, is likely to be greeted with scepticism. Consequently, one of the objectives of the present paper is to present the results of a further meta-analysis of studies of sex differences on the Progressive Matrices, this time not on general population samples, but on university students.

The second objective of the paper is to present the results of a meta-analysis of studies on sex differences in the variance of reasoning ability assessed by the Progressive Matrices. It has frequently been asserted that while there is no difference between males and females in average intelligence, the variance is greater in males than in females. The effect of this is that there are more males with very high and very low IQs. This contention was advanced a century ago by Havelock Ellis (1904, p. 425), who wrote that, 'It is undoubtedly true that the greater variational tendency in the male is a psychic as well as a physical fact'. In the second half of the twentieth century this opinion received many endorsements. For instance, Penrose (1963, p. 186) opined that, 'the larger range of variability in males than in females for general intelligence is an outstanding phenomenon'. In similar vein: 'while men and women average pretty much the same IQ score, men have always shown more variability in intelligence. In other words, there are more males than females with very high IQs and very low IQs' (Eysenck, 1981, p. 42); 'the consistent story has been that men and women have nearly identical IQs but that men have a broader distribution. . .the larger variation among men means that there are more men than women at either extreme of the IQ distribution' (Herrnstein & Murray, 1994, p. 275); 'males are more variable than females' (Lehrke, 1997, p. 140); 'males' scores are more variable on most tests than are those of females' (Jensen, 1998, p. 537); 'the general pattern suggests that there is greater variability in general intelligence within groups of boys and men than within groups of girls and women' (Geary, 1998, p. 315); and 'there is some evidence for slightly greater male variability' (Lubinski, 2000, p. 416). This position has been confirmed by the largest data set on sex differences in the variability of intelligence. In the 1932 Scottish survey of 86,520 11-year-olds there was no significant difference between boys and girls in the means but boys had a significantly larger standard deviation of 14.9 compared with 14.1 for girls as reported by Deary, Thorpe, Wilson, Starr, and Whalley (2003). The excess of boys was present at both extremes of the distribution. In the 50–59 IQ band 58.6% of the population were boys and in the 130–139 IQ band 57.7% of the population were boys.

The theory that the variability of intelligence is greater among males is not only a matter of academic interest but, in addition, has social and political implications. As Eysenck, Herrnstein, and Murray and several others have pointed out, a greater variability of intelligence among males than among females means that there are greater proportions of males at both the low and high tails of the intelligence distribution. The difference in these proportions can be considerable. For instance, it has been reported by Benbow (1988) that among the top 3% of American junior high school students in the age range of approximately 12–15 years scoring above a score of 700 for college entrance on the maths section of the Scholastic Aptitude Test, boys outnumber girls by 12.9 to 1. The existence of a greater proportion of males at the high tail of the intelligence distribution would do something to at least partially explain both the greater proportion of men in positions that require high intelligence and, also, the greater numbers of male geniuses. This theory evidently challenges the thesis of the 'glass ceiling' according to which the under-representation of women in senior positions in the professions and industry, and among Nobel Prize winners and recipients of similar honours, is caused by men discriminating against women. As Lehrke (1997, p. 149) puts it: 'the excess of men in positions requiring the very highest levels of intellectual or technical capacity would be more a consequence of genetic factors than of male chauvinism'. Nevertheless, the magnitudes of the suggested sex differences with regard to either mean levels or variability would not be sufficient to explain the sex biases currently observed in the labour market (Gottfredson, 1997; Lynn & Irwing, 2004; Powell, 1999).

However, the thesis of the greater variability of intelligence among males has not been universally accepted. In the most recent review of some of the research on this issue, Mackintosh (1998a 1998b, p. 188) concludes that there is no strong evidence for greater male variability in non-verbal reasoning or verbal ability and that 'the one area where there is consistent and reliable evidence that males are more variable than females and where, as a consequence, there are more males than females with particularly high scores, is that of numerical reasoning and mathematics'.

We have two objectives in this paper. First, to determine whether our meta-analysis of sex differences on the Progressive Matrices among general population samples, that showed a 5 IQ point male advantage among adults, can be replicated in a meta-analysis of sex differences on the Progressive Matrices among university students. Second, to present the first meta-analysis of studies of the frequent assertion that the variance of non-verbal reasoning ability measured by the Progressive Matrices is greater in males than in females.

## Method

The meta-analyst has to address three problems. These have been memorably identified by Sharpe (1997) as the 'Apples and Oranges', 'File Drawer', and 'Garbage in–Garbage out' problems. The 'Apples and Oranges' problem is that different phenomena are sometimes aggregated and averaged, where disaggregation may show different effects for different phenomena. The best way of dealing with this problem is to carry out meta-analyses, in the first instance, on narrowly-defined phenomena and populations and then attempt to integrate these into broader categories. In the present meta-analysis, this problem has been dealt with by confining the analysis to studies using the Progressive Matrices on university students.

The 'File Drawer' problem is that studies producing significant effects tend to be published, while those producing non-significant effects tend not to be published and remain unknown in the file drawer. This is a serious problem for some meta-analyses such as those on effects of treatments of illness and whether various methods of psychotherapy have any beneficial effect, in which studies finding positive effects are more likely to be published, while studies showing no effects are more likely to remain unpublished. It is considered that this should not be a problem for our present inquiry because we have not found any studies that have been carried out with the primary objective of ascertaining whether there are sex differences in the mean or variance on the Progressive Matrices. Data on sex differences on the Progressive Matrices are available because they have been reported in a number of studies as by-products of studies primarily concerned with other phenomena.

The 'Garbage in–Garbage out' problem concerns what to do with poor quality studies. Meta-analyses that include many poor quality studies have been criticized by Feinstein (1995) as 'statistical alchemy' which attempt to turn a lot of poor quality dross studies into good quality gold. Poor quality studies are liable to obscure relationships that exist and can be detected by good quality studies. Meta-analysts differ in the extent to which they judge studies to be of such poor quality that they should be excluded from the analysis. Some meta-analysts are 'inclusionist' while others are 'exclusionist', in the terminology suggested by Kraemer, Gardner, Brooks, and Yesavage (1998). The problem of what should be considered 'garbage' generally concerns samples that are likely to be unrepresentative of males and females, such as those of shop assistants, clerical workers, psychiatric patients in the military and the mentally retarded listed in Court's (1983) review of sex differences on the Progressive Matrices. As our meta-analysis is confined to samples of university students, this problem is not regarded as serious, except possibly in cases where the students come from a particular faculty in which the sex difference might be different from that in representative samples of all students. Our meta-analysis is 'inclusionist' in the sense that it includes all the studies of sex differences on the Progressive Matrices among students that we have been able to locate.

The next problem in meta-analysis is to obtain all the studies of the issue. This is a difficult problem and one that it is rarely and probably never possible to solve completely. Meta-analysts attempt to find all the relevant studies of the phenomena being considered by examining previous reviews and searching sources like *Psychological Abstracts* and other abstracts. But these do not identify all the relevant studies, a number of which provide data incidental to the main purpose of the study and which are not mentioned in the abstract or among the key words. Hence the presence of these data cannot be identified from abstracts or key word information. A number of studies containing the relevant data can only be found by searching a large number of publications. It is virtually impossible to identify all relevant studies. It is considered that, although this is a problem for this and for many other meta-analyses, it is not a serious problem for our present study because the results are sufficiently clear-cut that they are unlikely to be seriously overturned by further studies that have not been identified. If this should prove incorrect, no doubt other researchers will produce these unidentified studies and integrate them into the meta-analysis. For the present meta-analysis, the studies were obtained from Court's (1980, 1983) bibliography and review of studies of the sex difference on the Progressive Matrices, from the series of manuals on the Progressive Matrices published by Raven and his colleagues (e.g. Raven, 1981; Raven, Raven, & Court, 1998; Raven, Court, & Raven, 1996), from *Psychological Abstracts* from 1937 (the year the Progressive Matrices was first published). In addition to consulting these bibliographies, we made computerized database searches of PsycINFO, ERIC, Web of Science, Dissertation Abstracts, the British Index to Theses and Cambridge Scientific Abstracts for the years covered up to and including 2003. Finally, we contacted active researchers in the field and made a number of serendipitous discoveries in the course of researching this issue. Our review of the literature covers the years 1939–2002.

The method of analysis follows procedures developed by Hunter and Schmidt (1990). This consists of adopting Cohen's $d$ (the difference between the male and female means divided by the within group standard deviation) as the measure of effect size (Cohen, 1977). In the majority of studies, means and standard deviations were reported, which allowed direct computation of $d$. However, in the case of Backhoff-Escudero (1996) there were no statistics provided from which an estimate of $d$ could be derived. In this instance, $d$ was calculated by dividing the mean difference by the averaged standard deviations of the remaining studies.

### *Meta-analysis of effect sizes*

For each analysis, the mean of effect sizes was calculated, weighted by sample size (see Table 2). The estimates were then corrected for measurement error. The weighted artifact distributions used in this calculation were derived from those reliability studies reported in Court and Raven (1995) for the Standard and Advanced Progressive Matrices, with a sample $N \geq 300$. In order to detect the presence of moderator variables, tests of homogeneity of effect sizes were conducted using Hunter and Schmidt's 75% rule. Each corrected mean $d$-score was fitted with a confidence interval, using the appropriate formula for homogeneity or heterogeneity of effect sizes, whichever was indicated.

## Results

Table 1 gives data for 22 samples of university students. The table gives the authors and locations of the studies; the sizes and means of the male and female samples; the standard deviations of the males and females for 19 of the studies (*SD*s were not given in the other three); the male–female difference in *d*-scores (both corrected and uncorrected for measurement error) with positive signs denoting higher means by males and negative signs higher means by females; and whether the Standard or Advanced form of the test was used. It will be noted that in all the samples, males obtained a higher mean than females, with the single exception of the Van Dam (1976) study in Belgium. This is based on a sample of science students. Females are typically uncommon among science students and are not representative of all female students, and this may be the explanation for this anomalous result.

The corrected mean effect size was calculated for seven subsamples. Since the Hunter and Schmidt approach to meta-analysis appears controversial, Table 2 provides uncorrected mean *d* scores, in addition to median, and mean corrected *d* scores together with their associated confidence intervals. The number of studies, the total sample size, and percentage of variance explained by sampling and measurement error are also included. For the total sample, the percentage variance in *d* scores explained by artifacts is only 28.6%, indicating a strong role for moderator variables.

We explored two possible explanations for the heterogeneity in effect sizes. One such explanation may reside in the use of two different tests: the Standard Progressive Matrices (SPM) and Advanced Progressive Matrices (APM). These are intended for quite different populations and, if the tests were chosen appropriately, this in itself would lead to different estimates of the sex difference. Secondly, it would seem that selectivity in student populations operates somewhat differently for men and women. For example, for the UK in 1980, only 37% of first degrees were obtained by women. From the perspective of the current study, the major point is that estimates of the sex difference in scores on the Progressive Matrices based on the 1980 population of undergraduates would probably have produced an underestimate, since it may be inferred that the female population was more highly selected, whereas by 2001 the reverse situation obtained. Since our data includes separate estimates of the variance in scores on the Progressive Matrices, we were able to code a dummy variable with '1' corresponding to greater variability in the male population and '2' representing greater variability amongst females, the expectation being that this variable would moderate the *d* score estimate of the magnitude of the sex difference in scores on the Progressive Matrices.

An analysis to test for these possibilities was carried out using weighted least squares regression analysis as implemented in Stata. This procedure produces correct estimates of sums of squares, obviating the need for the corrections prescribed by Hedges and Becker (1986), which apply to standard weighted regression. Weighted mean corrected estimates of *d* formed the dependent variable, and the two dummy variables for type of test and relative selectivity in male and female populations constituted the independent variables, which were entered simultaneously. The analysis revealed that the two dummy variables combined explained 57.4% of variance (using the adjusted $R^2$) in the weighted corrected mean *d* coefficients. The respective beta coefficients were $-0.22$ ($t(1, 16) = -4.3$, $p = .001$) for the effect of type of test and $-.15$ ($t(1, 16) = -2.17$, $p = .047$) for the effects of relative selectivity in male versus female populations, indicating that both variables explained significant amounts of variance in *d* scores.

**Table 1.** Sex differences on the standard and advanced progressive matrices among university students

| Reference | Location | N | | M | | SD | | *d* | δ | Age | Test |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Males | Females | Males | Females | Males | Females | | | | |
| Abdel-Khalek, 1988 | Egypt | 205 | 247 | 44.20 | 40.80 | 7.80 | 8.40 | .42 | .48 | 23–24 | S |
| Mohan, 1972 | India | 145 | 165 | 46.48 | 43.88 | 7.32 | 7.70 | .35 | .39 | 18–25 | S |
| Mohan & Kumar, 1979 | India | 200 | 200 | 45.54 | 44.99 | 7.00 | 7.02 | .08 | .09 | 19–25 | S |
| Silverman *et al.*, 2000 | Canada | 46 | 65 | 16.57 | 14.77 | 3.73 | 3.87 | .47 | .54 | 21 | S |
| Rushton & Skuy, 2000 | S.Africa-Black | 49 | 124 | 46.13 | 42.15 | 7.19 | 9.11 | .46 | .53 | 17–23 | S |
| Rushton & Skuy, 2000 | S.Africa-White | 55 | 81 | 54.44 | 53.33 | 4.57 | 3.76 | .27 | .31 | 17–23 | S |
| Lovaglia *et al.*, 1998 | USA | 62 | 62 | 55.26 | 53.77 | 3.02 | 3.60 | .46 | .51 | – | S |
| Crucian & Berenbaum, 1998 | USA | 86 | 132 | 22.5 | 21.6 | 3.8 | 3.69 | .24 | .27 | 18–46 | S* |
| Grieve & Viljoen, 2000 | S.Africa | 16 | 14 | 38.06 | 36.57 | 6.06 | 9.69 | .19 | .22 | | S |
| Backhoff–Escudero, 1996 | Mexico | 4,878 | 4,170 | 46.68 | 46.28 | – | – | .06 | .07 | 18–22 | S |
| Pitariu, 1986 | Romania | 785 | 531 | 21.89 | 21.28 | 5.02 | 4.63 | .10 | .11 | – | A |
| Paul, 1985 | USA | 110 | 190 | 28.40 | 26.23 | 6.44 | 6.15 | .43 | .49 | – | A |
| Bors & Stokes, 1998 | Canada | 180 | 326 | 23.00 | 21.68 | 4.85 | 5.11 | .24 | .27 | 17–30 | A |
| Florquin, 1964 | Belgium | 214 | 64 | 43.53 | 41.27 | – | – | .37 | .42 | – | A |
| Van Dam, 1976 | Belgium | 517 | 327 | 34.83 | 35.36 | 6.21 | 5.60 | −.11 | −.12 | – | A |
| Kanekar, 1977 | USA | 71 | 101 | 26.08 | 25.97 | 5.04 | 4.96 | .02 | .02 | – | A |
| Colom & Garcia-Lopez, 2002 | Spain | 303 | 301 | 23.90 | 22.40 | 4.8 | 5.3 | .30 | .34 | 18 | A |
| Lynn & Irwing 2004 | USA | 1,807 | 415 | 25.78 | 24.22 | 4.80 | 5.30 | .29 | .33 | – | A |
| Yates & Forbes, 1967 | Australia | 565 | 180 | 23.67 | 22.75 | 5.29 | 5.57 | .19 | .21 | – | A |
| Abad, Colom, Rebollo, & Escorial (2004) | Spain | 1,069 | 901 | 24.19 | 22.73 | 5.37 | 5.47 | .27 | .31 | 17–30 | A |
| Colom, Escorial, & Rebello, 2004 | Spain | 120 | 119 | 24.57 | 23.32 | 4.13 | 4.52 | .29 | .33 | 18–24 | A |
| Geary, Saults, Liu, & Hoard, 2000[a] | USA | 113 | 123 | 114.6 | 113.40 | 11.6 | 12.0 | .12 | .13 | – | – |

*Note*. S = Standard Progressive Matrices, S* = Standard Progressive Matrices, short version, A = Advanced Progressive Matrices, *d* = Cohen's measure of effect size, δ = Cohen's *d* corrected for measurement error.
[a] Scores expressed in terms of IQ.

Since the Hunter and Schmidt procedures for the 18 studies included in the regression analysis indicated that 40.3% of the variance in *d* scores was explained by artifacts, and 57.4% is explained by the two moderator variables, it appears that type of test and relative selectivity represent the sole moderator variables.

The existence of two moderator variables provided the rationale for carrying out separate analyses for those studies that employed the SPM and APM, and for those studies in which male variance exceeded that of females, together with the converse case. The need for separate analyses, which included and excluded the study of Backhoff-Escudero (1996) were indicated, since this Mexican study was both large ($N = 9,048$) and an outlier at 3.22 standard deviations below the weighted mean *d*. As would be anticipated, excluding the Mexican study from the analysis increases the magnitude of estimates of the weighted mean corrected *d*, particularly for those studies which employed the SPM, in which the increase is from .11*d* to .33*d*, although the effect on median corrected *d*s is markedly less (see Table 2). Excluding Mexico, studies that employed the SPM provide higher estimates of *d* than are derived from studies using the APM. This would be expected if the SPM were employed as intended for samples that more closely correspond to general population samples, since range restriction, which operates in the more selected populations for which the APM is appropriate, would attenuate the magnitude of the sex difference. Also as expected, those studies in which the sample is more selective with respect to males provide higher estimates of *d* (.28) than samples that are more selective with respect to females ($d = .11$).

A notable feature of the summary estimates of *d* is that they all show higher average scores for males than females on the Progressive Matrices (see Table 2). Higher average scores for males are present even when the samples are biased in favour of female superiority, by being more selective for women, and this is particularly true for median estimates. In all cases, except for that in which samples were more selective for women, the 95% confidence interval did not include zero and, therefore, the male superiority was significant at the .05 level. The question arises as to which of the estimates of *d* is most accurate. There is an argument that, since the estimates based on the SPM best approximate general population samples, provided the outlier of Backhoff-Escudero (1996) is eliminated, that this provides the best estimate of the weighted corrected mean *d*. However, since we have established that the relative magnitude of variance in the male and female populations is an even more important moderator of *d*, we need to examine whether these variances differ overall. Ferguson (1989) suggests an *F*-test to determine whether a difference in variance between two independent samples is significant. For those studies which employed the APM, the *F*-test, $F(3344, 5660) = 1.00$, $p > .05$, was not significant, so we may conclude that relative selectivity of males versus females has no effect on this estimate. However, for the SPM, studies showed a significant net greater variability in the female samples, $F(882, 656) = 1.20$, $p < .02$, which suggests that the estimate of the weighted corrected mean *d* in these samples is biased upwards. Since the estimate of *d* based on the studies using the APM is unaffected by differential variance in the male and female samples, but is almost certainly subject to attenuation due to range restriction, we may conclude that the estimate of the weighted corrected mean *d* of .22 provided by these studies represents a lower bound. Equally, the estimate of .33 *d* based on the studies using the SPM is probably an upper bound estimate, though this too has probably been subject to range restriction. To obtain an optimal estimate of *d*, correcting the estimate based on the APM for range restriction might appear an obvious strategy. Unfortunately, because the APM is unsuitable for general population samples, there

**Table 2.** Summary statistics for sex differences on the SPM and APM

| Studies included | *k* | *N* | *d* | % variance explained by artifacts | δ | 95% confidence interval for δ | Median δ |
|---|---|---|---|---|---|---|---|
| 1. All | 22 | 20,432 | .14 | 28.6 | .15 | .11–.27 | .31 |
| 2. All minus Mexico | 21 | 11,386 | .21 | 43.2 | .23 | .18–.28 | .31 |
| 3. All using SPM | 10 | 11,002 | .10 | 33.2 | .11 | .05–.17 | .35 |
| 4. All using SPM minus Mexico | 9 | 1,954 | .31 | 99.6 | .33 | .33–.34 | .39 |
| 5. All using APM | 11 | 9,196 | .20 | 32.3 | .22 | .15–.28 | .31 |
| 6. Selective for males | 13 | 7,483 | .28 | 119.8 | .31 | .30–.31 | .34 |
| 7. Selective for females | 6 | 3,539 | .10 | 41.4 | .10 | −.003–.21 | .24 |

is no basis on which to assess the extent to which this estimate is subject to range restriction.

Although we have bypassed the issue, the question of whether overall variability is greater in the male or female samples is of interest in itself since, as noted above, it is frequently asserted that males show greater variability in *g*. As indicated by the *F*-tests, in student samples there is either no difference in variability of scores between males and females, or in the case of studies based on the SPM, women show greater variance than men.

## Discussion

There are five points of interest in the results. First, the present meta-analysis of sex differences on the Progressive Matrices among university students showing that men obtain significantly higher means than females confirms the results of our meta-analysis of sex differences on this test among general population samples (Lynn & Irwing, 2004). The magnitude of the male advantage found in the present study lies between 3.3 and 5 IQ points, depending on various assumptions. Arguably the best estimate of the advantage of men to be derived from the present study is .31*d*, based on all the studies and shown in the first row of Table 2. This is the equivalent of 4.6 IQ points and is closely similar to the 5 IQ points found in the meta-analyses of general population samples previously reported. We suggest that the 5 IQ point male advantage among adults based on general population samples is to be preferred as the best estimate of the advantage of men on the Progressive Matrices currently available. The 4.6 IQ advantage among student samples is somewhat less persuasive because of sampling issues, but should be regarded as strong corroboration of the results obtained on general population samples.

These results are clearly contrary to the assertions of a number of authorities including Eysenck (1981), Court (1983), Mackintosh (1996, 1998a, 1998b) and Anderson (2004, p. 829). These authorities have asserted that there is no difference between the means obtained by men and women on the Progressive Matrices. Thus, the tests 'give equal scores to boys and girls, men and women' (Eysenck, 1981, p. 41); 'there appears to be no difference in general intelligence' (Mackintosh, 1998a, 1998b, p. 189); and 'the evidence that there is no sex difference in general ability is overwhelming' (Anderson, 2004, p. 829). Mackintosh in his extensive writings on this question has sometimes been more cautious, e.g. 'If I was thus overconfident in my assertion that there was *no* sex difference. . . if general intelligence is defined as Cattell's Gf, best measured by tests such as Raven's Matrices. . . then the sex difference in general intelligence among young adults today . . .is trivially small, surely no more than 1–2 IQ points either way' 1998b, p. 538). Contrary to these assertions, our meta-analyses show that the sex difference on the Progressive Matrices is neither non-existent nor 'trivially small' and certainly not '1–2 IQ points either way', that is, in favour of men or women. Our results showing a 4.6 to 5 IQ point advantage for men is testimony to the value of meta-analysis as compared with impressions gained from two or three studies.

Second, the Progressive Matrices is widely regarded as one of the best tests of Spearman's *g*, the general factor underlying all cognitive abilities. For instance, Court (1983, p. 54) has written that it is 'recognised as perhaps the best measure of *g*' and Jensen (1998, p. 541) that 'the Raven tests, compared with many others, have the highest *g* loading'. Mackintosh (1996, p. 564) has written that 'general intelligence' can be equated with abstract reasoning ability and 'fluid intelligence' (Gf), and that the Progressive Matrices is 'the paradigm test of non-verbal, abstract reasoning ability'. These writers have argued that there is no sex difference on the Progressive Matrices and, therefore, that there is no sex difference on general intelligence or *g*. Now that we have established that men obtain higher means than women on the Progressive Matrices, it follows that men have higher general intelligence or *g*.

Third, the finding that males have a higher mean reasoning ability than females raises the question of how this can be explained. It has been proposed that it can be accounted for in terms of the larger average brain volume (Lynn, 1994; Nyborg, 2003). The argument is that men, on average, have larger brain volume than women, even when this is controlled for body size, as shown independently by Ankney (1992) and Rushton (1992). Brain volume (measured by magnetic resonance imaging) is positively associated with intelligence at a correlation of .40, as shown in the meta-analysis of Vernon *et al.* (2000, p. 248). Intelligence is conceptualized as reasoning ability or fluid intelligence and operationally defined as IQs obtained on the Progressive Matrices, as proposed by Mackintosh (1996) and Jensen (1998). Ankney expressed the male-female difference in brain size in standard deviation units as $0.78d$. Hence, the larger average brain size of men may theoretically give men an advantage in intelligence arising from a larger average brain size of 0.78 multiplied by 0.40, giving a theoretical male advantage of $.31d = 4.7$ IQ points. This is a close fit to the sex difference obtained empirically in our previous meta-analysis of the sex difference of 5 IQ points on the Progressive Matrices in general population samples, and of 4.6 IQ points on the Progressive Matrices, in the present meta-analysis of the sex difference in college student samples.

This theory has not been universally accepted. Mackintosh (1996), Jensen (1998), and Anderson (2004) dispute that there is a sex difference in intelligence defined as IQs obtained on the Progressive Matrices (or on any other measure) and suggest that the larger male brain can be explained by other ways than the need to accommodate higher intelligence. Mackintosh (1996) proposes that the larger average male brain is a by-product of the larger male body, while Jensen proposes that females have the same number of neurones as males but these are smaller and more closely packed, and Anderson (2004) asserts that the theory is 'idle speculation' for a number of reasons including '(a) Neanderthals had a bigger brain than current humans but nobody wants to make a claim that they were more intelligent than modern people; (b) the relationship between brain size and IQ within species is very small; (c) the causal direction is ambiguous (IQ and its environmental consequence may affect brain size rather than the other way round); (d) the brain does far more than generate IQ differences and it may be those other functions that account for any male/female differences in size'. Nyborg (2003), however, accepts the argument that the male/female difference in brain size can explain the difference in intelligence and shows empirically that there is a sex difference among adults of 5.55 IQ points in *g* measured by a battery of tests, rather than by the Progressive Matrices. It has been shown by both Posthuma *et al.* (2002), and Tompson *et al.* (2001) that the evidence is that the association between both grey-matter and white-matter volume and *g* is largely or even completely attributable to genetic factors, which seems to rule out an environmental explanation of the sex difference in *g*. However, a high heritability of a trait in men and women does not necessarily imply that the sex difference is genetic. There may be environmental explanations of the small sex difference in general cognitive ability observed in adult populations as proposed by Eagly and Wood (1999). It is evident that these issues will have to be considered further before a consensus emerges.

Fourth, a number of those who have asserted that there is no sex difference in intelligence have qualified their position by writing that there is no sex difference 'worth speaking of' Mackintosh (1996, p. 567), 'only a very small advantage of boys and men' (Geary, 1998, p. 310), ' no practical differences in the scores obtained by males and females' (Halpern, 2000, p. 90), 'no meaningful sex differences' (Lippa, 2002), and 'negligible differences' (Jorm *et al.*, 2004, p. 7). These qualifications raise the question of whether a sex difference in average general intelligence of 5 IQ points should be regarded as 'not worth speaking of', 'not a practical difference', 'not meaningful' and 'negligible'. We do not think that a 5 IQ point difference can be so easily dismissed. The effect of an 5 IQ difference between men and women in the mean with equal variances is to produce a male-female ratio of 2.3:1 with IQs of 130 + , 3:1 with IQs of 145 + , and 5.5:1 with IQs of 155 + . These different proportions of men and women with high IQs are clearly 'worth speaking of' and may go some way to explaining the greater numbers of men achieving distinctions of various kinds for which a high IQ is required, such as chess grandmasters, Fields medallists for mathematics, Nobel prize winners and the like.

However, while a small sex difference in *g* may be more significant for highly complex tasks (Gottfredson, 2003; Nyborg, 2003) than might first appear, caution should be exercised in generalizing to sex inequalities in the labour market as a whole. In the UK, in contrast to the situation 25 years ago, women now outnumber men at every level of educational achievement with the sole exception of the numbers registered for doctorates (Lynn & Irwing, 2004). There is an argument that the factors responsible for educational achievement are highly similar to the corresponding qualities required for significant occupational achievement (Kunzel, Hezlett, & Ones, 2004). It is perhaps not surprising then that work force participation has seen similar degrees of change, although inevitably there is something approaching a 40-year lag between educational change and the evidence of its full effects on the work force. For example, in the US, by 1995, women accounted for 43% of managerial and related employment, nearly double their share of 22% in 1975 and women have become increasingly represented in senior positions throughout North America and Europe (Stroh & Reilly, 1999).

In terms of higher education, in 1980 only 37% of first degrees in Britain were obtained by women (Ramprakash & Daly, 1983), but by 2001 this figure had risen to 56% (Matheson & Babb, 2002). Up to 1997, men obtained more first class honours degrees than women, but in 1998, women, for the first time, gained more first class honours degrees than men (Higher Education Statistics Agency, 1998).

However, despite this evidence of rapid advance in the educational and occupational achievements of women, there is still evidence of vertical segmentation of the labour market. Within all countries, the proportion of women decreases at progressively higher levels in the organizational hierarchy (Parker & Fagenson, 1994). Although the proportion of women in higher management has increased over time in individual countries (Hammond & Holton, 1994), nevertheless, this proportion is still reported as less than 5%, however top management is defined (e.g. 1–2% UK & Australia, Davidson & Cooper, 1992). Although our best estimate of a .33*d* male advantage in *g* may contribute to this imbalance, a number of considerations suggest that it only does so partially. Firstly, in accordance with the argument that abilities that contribute to educational success are similar to those that lead to occupational achievement, it seems likely that once the current crop of educationally high achieving women fully saturate the labour market, the under-representation of women at top management levels will be substantially reduced. Secondly, Gottfredson (1997) estimates that an IQ of 125 is adequate to ascend to all levels in the labour market. Thirdly, there is some evidence to suggest that, for any given level of IQ, women are able to achieve more than men (Wainer & Steinberg, 1992), possibly because they are more conscientious and better adapted to sustained periods of hard work. The small male advantage in *g* is, therefore, likely to be of most significance for tasks of high complexity such as 'complex problem-solving in mathematics (Benbow, 1992), engineering and physics (Lubinski, Benbow, & Morelock, 2000), and in other areas calling for high spatial ability (Shea, Lubinski, & Benbow, 2001)' (Nyborg, 2003, p. 215).

Fifth, the finding in this meta-analysis that there is no sex difference in variance on the Advanced Progressive Matrices and that females show greater variance on the Standard Progressive Matrices is also contrary to the frequently made contention, documented in the introduction, that the variance of intelligence is greater among males. This result should be generalizable to the general population of normal intelligence. The greater male variance theory may, however, be correct for general population samples that include the mentally retarded. It is not wholly clear whether males are overrepresented among the mentally retarded. According to Mackintosh (1998a, p. 187) 'studies of the prevalence of mental retardation have found little or no evidence' for an overrepresentation of males. He does not, however, cite Reed and Reed (1965) who made the largest study of sex differences in the rate of mental retardation in a sample of 79,376 individuals, among whom the percentages of retardates were .022 among males and .016 among females. This difference would be expected because there are more males than females among those with autism and several X-linked disorders including Fragile X syndrome, Duchenne's muscular dystrophy, and some other rare disorders. If this is correct there should be greater variance of intelligence among males in general population samples that include the mentally retarded because of the higher male mean, producing more males with high IQs, and the greater prevalence of mental retardation among males, producing more males with low IQs. The issue of whether there is greater male variance for intelligence in general population samples needs to be addressed by meta-analysis.

## References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's advanced progressive matrices: Evidence for bias. *Personality and Individual Differences*, *36*, 1459–1470.

Abdel-Khalek, A. M. (1988). Egyptian results on the Standard Progressive Matrices. *Personality and Individual Differences*, *9*, 193–195.

Anderson, M. (2004). Sex differences in general intelligence. In R. L. Gregory (Ed.), *The Oxford companion to the mind*. Oxford, UK: Oxford University Press.

Ankney, C. D. (1992). Sex differences in relative brain size: The mismeasure of women, too? *Intelligence*, *16*, 329–336.

Backhoff-Escudero, E. (1996). Prueba matrices progressives de Raven: normas de universitarios Mexicanos. *Revista Mexicana de Psicologia*, *13*, 21–28.

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

Benbow, C. P. (1992). Academic achievement in mathematics and science of students between ages 13 and 23: Are there differences in the top one percent of mathematical ability? *Journal of Educational Psychology*, *84*, 51–61.

Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first year university students and the development of a short form. *Educational and Psychological Measurement*, *58*, 382–398.

Brody, N. (1992). *Intelligence*. San Diego, CA: Academic.

Caplan, P. J., Crawford, M., Hyde, J. S., & Richardson, J. T. E. (1997). *Gender differences in human cognition*. Oxford: Oxford University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.

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

Colom, R., Escorial, S., & Rebello, I. (2004). Sex differences on the progressive matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, *37*, 1289–1293.

Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, *32*, 445–452.

Colom, R., Garcia, L. F., Juan-Espinoza, M., & Abad, F. J. (2002). Sex differences in general intelligence on the WAIS 111: Implications for psychological assessment. *Spanish Journal of Psychology*, *5*, 29–35.

Colom, R., Juan-Espinoza, M., Abad, F. J., & Garcia, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*, 57–68.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, *36*, 75–82.

Court, J. C. (1980). *Researchers' bibliography for Raven's progressive matrices and Mill Hill vocabulary scales*. Australia: Flinders University.

Court, J. H. (1983). Sex differences in performance on Raven's progressive matrices: A review. *Alberta Journal of Educational Research*, *29*, 54–74.

Court, J. C., & Raven, J. (1995). Normative, reliability and validity studies. Association between alcohol consumption and cognitive performance. *American Journal of Epidemiology*, *146*, 405–412.

Crucian, G. P., & Berenbaum, S. A. (1998). Sex differences in right hemisphere tasks. *Brain and Cognition*, *36*, 377–389.

Davidson, M. J., & Cooper, C. L. (1992). *Shattering the glass ceiling: The woman manager*. London: Paul Chapman.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence*, *31*, 533–542.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior. *American Psychologist*, *54*, 408–423.

Ellis, H. (1904). *Man and woman: A study of human secondary sexual characteristics*. London: Walter Scott.

Eysenck, H. J. (1981). In H. J. Eysenck & L. Kamin (Eds.), *Intelligence: The battle for the mind: H.J. Eysenck versus Leon Kamin* (pp. 11–89). London: Pan.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, *48*, 71–79.

Ferguson, G. A. (1989). *Statistical analysis in psychology and education* (6th ed). London: McGraw-Hill.

Filipek, P. A., Richelme, C., Kennedy, D. N., & Caviness, V. S. (1994). Young-adult human brain – An MRI-based morphometric analysis. *Cerebral Cortex*, *4*, 344–360.

Florquin, F. (1964). Les P.M. 1947 de J.C.Raven au niveau des classes terminals du cycle secondaire. *Revue Belge de Pyschologie et de Pedagogie*, *26*, 108–120.

Geary, D. C. (1998). *Male, female*. Washington, DC: American Psychological Association.

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency and arithmetical reasoning. *Journal of Experimental Child Psychology*, *77*, 337–353.

Glass, G. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, *51*, 3–8.
Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.
Gottfredson, L. S. (2003). *g*, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence*. Amsterdam: Elsevier.
Grieve, K. W., & Viljoen, S. (2000). An exploratory study of the use of the Austin Maze in South Africa. *South African Journal of Psychology*, *30*, 14–18.
Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain grey and white matter in healthy young adults: Correlations with cognitive performance. *Journal of Neuroscience*, *1*, 4065–4072.
Halpern, D. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Erlbaum.
Hammond, V., & Holton, V. (1994). The scenario for women managers in Britain in the 1990s. In N. J. Adler & D. N. Izraeli (Eds.), *Competitive frontiers: Women managers in a global economy* (pp. 224–242). Cambridge, MA: Blackwell.
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.
Heron, A., & Chown, S. (1967). *Age and function*. London: Churchill.
Herrnstein, R., & Murray, C. (1994). *The bell curve*. New York: Random House.
Higher Education Statistics Agency Online Information Service (1998). *Student table 14a – HE qualifications obtained in the UK by mode of study, domicile, gender and subject area 1997/98*. Retrieved July 4, 2002, from http://www.hesa.ac.uk/holisdocs/pubinfo/student/quals78.htm.
Hunter, J. S., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139–155.
Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 53–69.
Jensen, A. R. (1998). *The* g *factor*. Westport, CT: Praeger.
Jensen, A. R., & Reynolds, C. R. (1983). Sex differences on the WISC-R. *Personality and Individual Differences*, *4*, 223–226.
Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence* (pp. 139–242). Norwood, NJ: Ablex.
Jorm, A. F., Anstey, K. J., Christensen, H., & Rodgers, B. (2004). Gender differences in cognitive abilities: The mediating role of health state and health habits. *Intelligence*, *32*, 7–23.
Kanekar, S. (1977). Academic performance in relation to anxiety and intelligence. *Journal of Social Psychology*, *101*, 153–154.
Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: Massachusetts Institute of Technology.
Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods*, *3*, 23–31.
Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.
Lehrke, R. (1997). *Sex linkage of intelligence*. Westport, CT: Praeger.
Linn, M. C., & Petersen, A. C. (1985). Emergence and characterisation of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.
Lippa, R. A. (2002). *Gender, nature and nurture*. Mahwah, NJ: Erlbaum.
Lovaglia, M. J., Lucas, J. W., Houser, J. A., Thye, S. R., & Markovsky, B. (1998). Status process and mental ability test scores. *American Journal of Sociology*, *104*, 195–228.
Lubinski, D. (2000). Scientific and social significance of assessing individual differences. *Annual Review of Psychology*, *51*, 405–444.

Lubinski, D., Benbow, C. P., & Morelock, M. J. (2000). Gender differences in engineering and the physical sciences among the gifted: An inorganic-organic distinction. In K. A. Heller, F. J. Monks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook for research on giftedness and talent* (pp. 627–641). Oxford, UK: Pergamon Press.

Lynn, R. (1994). Sex differences in brain size and intelligence: A paradox resolved. *Personality and Individual Differences*, *17*, 257–271.

Lynn, R. (1998). Sex differences in intelligence: A rejoinder to Mackintosh. *Journal of Biosocial Science*, *30*, 529–532.

Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, *27*, 1–12.

Lynn, R., Allik, J., & Irwing, P. (2004). Sex differences on three factors identified in Raven's Standard Progressive Matrices. *Intelligence*, *32*, 411–424.

Lynn, R., Allik, J., & Must, O. (2000). Sex differences in brain size, stature and intelligence in children and adolescents: Some evidence from Estonia. *Personality and Individual Differences*, *29*, 555–560.

Lynn, R., Allik, J., Pullmann, H., & Laidra, K. (2002). Sex differences on the progressive matrices among adolescents: Some data from Estonia. *Personality and Individual Differences*, *34*, 669–679.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481–498.

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–572.

Mackintosh, N. J. (1998a). *IQ and human intelligence*. Oxford.

Mackintosh, N. J. (1998b). Reply to Lynn. *Journal of Biosocial Science*, *30*, 533–539.

Mann, C. (1990). Meta-analysis in the breech. *Science*, *249*, 476.

Matteson, J., & Babb, P. (2002). *Social trends*, No. 32. London: The Stationary Office.

Mohan, V. (1972). Raven's standard progressive matrices and a verbal test of general mental ability. *Journal of Psychological Researches*, *16*, 67–69.

Mohan, V., & Kumar, D. (1979). Performance of neurotics and stables on the standard progressive matrices. *Intelligence*, *3*, 355–368.

Nopoulos, P., Flaum, M., O'Leary, D., & Anreasen, N. C. (2000). Sexual dimorphism in the human brain: Evaluation of tissue volume, tissue composition and surface anatomy using magnetic resonance imaging. *Psychiatry Research Neuroimaging*, *98*, 1–13.

Nyborg, H. (2003). Sex differences *g*. In H. Nyborg (Ed.), *The scientific study of general intelligence*. Amsterdam: Elsevier.

Parker, B., & Fagenson, E. A. (1994). An introductory overview of women in corporate management. In M. J. Davidson & R. J. Burke (Eds.), *Women in management: Current research issues* (pp. 11–28). London: Paul Chapman.

Passe, T. J., Rajagopalan, P., Tupler, L. A., Byrum, C. E., MacFall, J. R., & Krishnan, K. R. R. (1997). Age and sex effects on brain morphology. *Progress in Neuropharmacology and Biological Psychiatry*, *21*, 1231–1237.

Paul, S. M. (1985). The advanced Raven's progressive matrices: Normative data for an American university population and an examination of the relationship with Spearman's *g*. *Journal of Experimental Education*, *54*, 95–100.

Penrose, L. S. (1963). *The biology of mental defect*. New York: Grune and Stratton.

Pitariu, H. (1986). Analiza de itemi si standardizarea Matricilor Progressive avansate. *Revista de Psihologie*, *32*, 33–43.

Posthuma, D., De Geus, E. J. C., Baare, W. F. C., Pol, H. E. H., Kahn, R. S., & Boomsma, D. I. (2002). The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience*, *5*, 83–84.

Powell, G. N. (1999). *Handbook of gender and work*. London: Sage.

Rabinowicz, T., Dean, P. E., Petetot, J. M. C., & de Courten-Myers, G. M. (1999). Gender differences in the human cerebral cortex: More neurons in males; More processes in females. *Journal of Child Neurology*, *14*, 98–107.

Ramprakash, D., & Daly, M. (1983). *Social trends*, No 13. London: HMSO.
Raven, J. C. (1939). The RECI series of perceptual tests: An experimental survey. *British Journal of Medical Psychology*, *18*, 16–34.
Raven, J. (1981). *Irish and British standardisations*. Oxford: Oxford Psychologists Press.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced progressive matrices*. Oxford, UK: Oxford Psychologists Press.
Raven, J. C., Court, J. H., & Raven, J. (1996). *Raven matrices progressivas manual*. Madrid: Publicaciones De Psicologia Aplicada.
Reed, E. W., & Reed, S. C. (1965). *Mental retardation: A family study*. Philadelphia: Saunders.
Rimoldi, H. J. A. (1948). Study of some factors related to intelligence. *Psychometrika*, *13*, 27–46.
Rushton, J. P. (1992). Cranial capacity related to sex, rank and race in a stratified sample of 6,325 military personnel. *Intelligence*, *16*, 401–413.
Rushton, J. P., & Skuy, M. (2000). Performance on Raven's matrices by African and white university students in South Africa. *Intelligence*, *28*, 251–266.
Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, *17*, 881–901.
Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614.
Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex difference. *Evolution and Human Behavior*, *21*, 201–213.
Spearman, C. (1923). *The nature of intelligence and the principles of cognition*. London: Macmillan.
Spearman, C. (1946). Theory of general factor. *British Journal of Psychology*, *36*, 117–131.
Stroh, L. K., & Reilly, A. H. (1999). Gender and careers: Present experiences and emerging trends. In G. N. Powell (Ed.), *Handbook of gender and work* (pp. 307–324). Thousand Oaks, CA: Sage.
Terman, L. M. (1916). *The measurement of intelligence*. Boston, MA: Houghton Mifflin.
Tompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V. P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C. G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C. I., & Toga, A. W. (2001). Genetic influences on brain structure. *Nature Neuroscience*, *4*, 1253–1258.
Van Dam, F. (1976). Le Advanced Progressive Matrices de J.C.Raven au niveau des premieres candidatures en sciences. *Revue Belgie de Psychologie*.
Vernon, P. A., Wickett, J. C., Bazana, P. G., & Stelmack, R. M. (2000). The neuropsychology and neurophysiology of human intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence*. Cambridge, UK: Cambridge University Press.
Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial ability: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.
Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the mathematics section of the Scholastic Aptitude Test: A bi-directional validity study. *Harvard Educational Review*, *62*, 323–335.
Witelson, S. F., Glezer, I. I., & Kigar, D. L. (1995). Women have greater density of neurons in posterior temporal cortex. *Journal of Neuroscience*, *15*, 3418–3428.
Yates, A. J., & Forbes, A. R. (1967). *Raven's advanced progressive matrices: Provisional manual for Australia and New Zealand*. Hawthorn, Australia: Australian Council for Educational Research.

*Received 14 July 2004; revised version received 15 April 2005*