Available online at www.sciencedirect.com
SCIENCE DIRECT®
Intelligence 32 (2004) 481–498
INTELLIGENCE

Sex differences on the progressive matrices: A meta-analysis

Richard Lynn[a], Paul Irwing[b,*]

[a]*University of Ulster, Coleraine, Northern Ireland, UK*
[b]*Manchester School of Management, University of Manchester Institute of Science and Technology, P.O. Box 88, Manchester, M60 1QD, UK*

Received 19 November 2003; received in revised form 31 May 2004; accepted 8 June 2004

Abstract

A meta-analysis is presented of 57 studies of sex differences in general population samples on the Standard and Advanced Progressive Matrices (SPM and APM, respectively). Results showed that there is no difference among children aged 6–14 years, but that males obtain higher means from the age of 15 through to old age. Among adults, the male advantage is 0.33*d* equivalent to 5 IQ points. These results disconfirm the frequent assertion than there are no sex differences on the progressive matrices and support a developmental theory that a male advantage appears from the age of 15 years. A meta-analysis of 15 studies of child samples on the Colored Progressive Matrices showed that among children aged 5–11 years boys have an advantage of 0.21*d* equivalent to 3.2 IQ points.

Keywords: Intelligence; Progressive matrices; Sex differences

1. Introduction

During the last two decades, the issue of sex differences in cognitive abilities has been addressed by carrying out meta-analyses of studies on verbal abilities (Hyde & Linn, 1988), spatial abilities (Linn & Peterson, 1985; Voyer, Voyer & Bryden, 1995), and mathematical abilities (Hyde, Fennema, & Lamon,

* Corresponding author. Tel.: +161 2003419.
E-mail address: Paul.Irwing@umist.ac.uk (P. Irwing).

0160-2896/$ - see front matter
doi:10.1016/j.intell.2004.06.008

1990). No meta-analysis has yet been made of sex differences in reasoning ability. In this paper, we present the first study of this issue in the form of a meta-analysis of sex differences in nonverbal reasoning ability measured by the Progressive Matrices Test.

Raven's Progressive Matrices was constructed in the late 1930s as a test of nonverbal or abstract reasoning ability (Raven, 1939). It has become one of the leading and most frequently used tests of this ability and has been described as "the paradigm test of nonverbal, 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. Thus, Court (1983, p. 54) has written that it is "recognised as perhaps the best measure of *g*"; and Jensen (1998, p. 541) wrote, "the Raven tests, compared with many others, have the highest *g* loading." There are three versions of the test: the Standard Progressive Matrices (SPM) for the ages of 6 years to adulthood; the Colored Progressive Matrices, an easier version of the test designed for children aged 5 through 12; and the Advanced Progressive Matrices (APM), 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. It has been virtually universally concluded that there is no difference in the mean scores obtained by males and females. The first statement of this conclusion was made by Raven (1939, p. 30) who 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 difference in ability above the age of 14." The conclusion that there is no sex differences on the progressive matrices has been endorsed by numerous scholars. For instance, Eysenck (1981, p. 41) stated that the tests "give equal scores to boys and girls, men and women." 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 this, he concludes that there is no difference or, in a subsequent paper, that there is only a very small difference consisting of no more than a difference of 1–2 IQ points among adults "either way" (Mackintosh, 1998).

Both Jensen (1998) and Mackintosh (1998) rely for their conclusion that there is no sex difference in mean scores on the progressive matrices on a review by Court (1983). This review summarized 118 studies of sex differences on the progressive matrices and concluded that most showed no significant 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).

Court's (1983) review is a useful starting point for tackling the question of sex differences on the progressive matrices but it cannot be accepted as a satisfactory basis for the conclusion that no sex differences exist. The review has at least five deficiencies. First, it is over 20 years old and a number of studies of this question have appeared subsequently and need to be considered. Second, it is a literary review that does not attempt to quantify the magnitude of any sex difference that may exist. Third, it includes general population samples and a variety of convenience samples including those of psychiatric patients, deaf children, retarded children, shop assistants, clerical workers, British, Indian, and French university students, Native Americans, and Inuit. Many of these samples cannot be regarded as representative of males and females, and in some cases this is stated by the authors and reported by Court. For instance, two studies showing that among British military personnel in World War II female neurotic patients obtained higher scores on the progressive matrices than males; Court reports that the authors of the studies believed that the reason for this was "probably due to the biased selection procedures in favour of bright women in the British Armed Forces" (Court, 1983, p. 60). With such diverse and unrepresentative samples in Court's review, it is not surprising that in some studies higher scores were achieved by males, in others higher scores were achieved by females and in others there were no sex differences. Fourth, 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 give a statistically significant difference between males and females. To detect a statistically significant difference of between 2 and 5 IQ points such as may exist between males and females on the progressive matrices requires a sample size of around 500. Court's review gives only one study of adults with a sample size of this number or more. This is Heron and Chown's (1967) study (n=600) on which men obtained a significantly higher mean score than women of 0.31d, approximately 4.65 IQ points (see Table 1). Nine of the studies showing no statistically significant sex difference in Court's review have sample sizes of fewer that 100, such as those of 60 Welsh 11- to 12-year-olds and 22 American 5-year-olds. Studies with these small sample sizes that show no significant differences between males and females have no value for the resolution of the issue of whether there is a small but significant sex difference on the progressive matrices. We found that of the 118 studies covered in Court's review, only 21 met the criteria of being general population samples and having sample sizes on a minimum of 50 males and 50 females that we have adopted in the meta-analysis to be presented. Fifth, Court's review did not include all the studies; in our literature search, we found nine studies of general population samples that were not given in the review and these included two with the largest sample sizes (Hsu, 1976; Wilson et al., 1975). For all these reasons, Court's review cannot be accepted as an adequate basis for the conclusion that there are no sex differences on the progressive matrices.

The consensus that there are no sex differences on the progressive matrices and on other tests of abstract (nonverbal) reasoning ability has been challenged by Lynn (1994, 1998, 1999). He has proposed a developmental theory of sex differences in intelligence that states that boys and girls mature at different rates both physically and mentally during childhood and adolescence. Boys and girls mature at about the same rate up to the age of around 7 years; from the age of 8, girls begin a growth spurt in which there is an acceleration of their physical growth in respect of height, weight, and brain size; the growth rate of girls slows at the age of 14 and 15, while the growth of boys continues. The developmental theory states that intelligence follows the same trend. Evidence supporting the theory has been provided in Lynn (1994, 1998, 1999) and in Lynn, Allik, and Must (2000). In regard to abstract (nonverbal) reasoning ability, the theory as originally formulated in Lynn states that over the age range of around 9 through 12 years, girls have an advantage of approximately 1 IQ point; by the age of 16 years, this has changed to a small advantage in favor of boys and among adults the male advantage is 2.4 IQ points. These estimates were not derived from data on the progressive matrices but (in the case of adults) from the American standardization samples of the Differential Aptitude Test. In a subsequent compilation of studies, it was proposed that among adults the male advantage on abstract reasoning is approximately 5 IQ points (Lynn, 1999).

In addition to the theory that sex differences on the progressive matrices vary by age, there is also a theory proposed by Mackintosh (1998) that there are cohort differences such that among older generations men achieved higher means than women, but this is no longer the case among recent cohorts of young adults.

The resolution of these conflicting theories requires a meta-analysis of studies on sex differences on the progressive matrices such as we have carried out and now describe. The meta-analysis is designed to test three hypotheses. These are as follows:

1. The Eysenck (1981), Court (1983), Mackintosh (1996), Jensen (1998) hypothesis that there is no sex difference on the progressive matrices.
2. The Lynn (1994, 1998, 1999) hypothesis stating that there is no sex difference on the progressive matrices among young children up to the age of 8 years; that girls have a slight advantage from the ages of 9 through 12 years; that boys and girls obtain about the same mean scores at the ages of 13–15 years; that at the age of 16 years boys have a higher mean than girls; that this advantage increases up to the age of 18–19 years; and that from the age of 20 onwards the male advantage lies between 2.4 (Lynn, 1994) and 5.0 IQ points (Lynn, 1999).
3. The Mackintosh (1998, p. 538) hypothesis that the higher mean scores of men on the progressive matrices is a cohort effect such that it may have been present among older generations but that "the sex difference in general intelligence among young adults today in the USA, Britain or Israel" (and, presumably, elsewhere in the economically developed world) " is trivially small, surely amounting to no more than 1–2 IQ points either way," i.e., in favor of either males or females.

2. Method

2.1. Criteria for selection of studies

The meta-analyst has to address three problems before analyzing the data. These have been 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 aggregation shows different effects for different phenomena. For instance, Hyde and Linn (1988) have shown in a meta-analysis of sex differences in verbal abilities that the magnitude of the difference varies for different kinds of verbal ability and ranges from a female advantage of 0.33 *d* for speech production to a male advantage of 0.16 *d* for analogies. This indicates that verbal ability is too broad a concept for the calculation of sex differences and that it is more meaningful to calculate differences in narrower abilities such as essay writing, vocabulary, verbal reasoning, etc. The best way of dealing with this problem is to carry out meta-analyses in the first instance on narrow abilities 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.

The "File Drawer" problem is that studies producing significant effects tend to be published, while those producing nonsignificant effects tend not to be published and remain hidden in the file drawer. This is a serious problem for meta-analyses comparing effects of treatments, such as 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 in the file drawer. It is considered that this should not be a problem for our present inquiry because very few studies have ever been carried out with the primary objective of ascertaining whether there are sex differences 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 a by-product of studies concerned with other phenomena.

The "Garbage in-Garbage out" problem is concerned with what to do with poor quality studies. Meta-analyses that include many poor quality studies have been criticized by Feinstein (1995) as "statistical alchemy" that attempts to turn a lot of poor quality dross data 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" and therefore excluded is a difficult one for meta-analysts. For our own meta-analysis, poor quality studies are of two general kinds. First, those whose samples are unrepresentative of males and females, such as those of shop assistants, clerical workers, psychiatric patients in the military, and the mentally retarded included in Court's (1983) review. To deal with this problem, our meta-analysis is confined to general population samples. Many of the general population studies are standardization samples that have been selected to be representative of the general population. Others have been less rigorously selected but nevertheless have been obtained in such as way that there is no reason to suspect any bias in the selection of males and females. The second kind of poor quality study consists of those with small sample sizes that are liable to produce anomalously large chance effect sizes that obscure the true relationship. Some meta-analysts ignore differences in sample sizes and accord all studies equal weight irrespective of sample size. This is reasonable for certain data sets where all the studies have about the same sample sizes. Where this is not the case, some meta-analysts deal with this problem by ignoring studies with samples below a certain size, while others weight the studies by the sample sizes. These two solutions amount to much the same thing because weighting by sample size dilutes and may effectively eliminate the contribution of studies with small samples. Where the meta-analyst has a number of large samples, the simplest procedure is to ignore small samples and confine the analysis to studies where sample sizes are considered acceptable. Data providing sex differences on the progressive matrices differ considerably in the sample sizes, ranging from 43,825 for 5-year-olds on the Colored Progressive Matrices in Taiwan (Hsu, 1976) to 22 for 5-year-olds on the same test in the United States (Garrity & Donaghue, 1976). To accord these two studies, equal weight in a meta-analysis cannot be regarded as a reasonable procedure. The solution adopted in the present meta-analysis has been to use only data based on a minimum of 50 males and 50 females. The inclusion of such small sample sizes is probably vulnerable to the criticism of being too inclusionist. On the other hand, it is considered preferable to err on the side of overinclusion because this provides other meta-analysts with the references and data that can be reworked in various ways to reduce the effects of small samples.

2.2. Literature search

Once the meta-analyst has drawn up criteria for studies to be included in the analysis, it is necessary to obtain all the studies meeting these criteria. This is a difficult problem and one that it is rarely possible to solve completely. Meta-analysts attempt to find all the relevant studies of the phenomena being considered by examining previous reviews and searching 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. Many studies of this kind can only be found by searching through a large number of publications. It is virtually impossible to identify all relevant studies. 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 et al. (e.g., Raven, 1981; Raven, Court, & Raven, 1996; Raven, Raven, & Court, 1998) and from *Psychological Abstracts* from 1937 (the year the progressive matrices was first published). In addition to consulting these bibliographies, which are widely regarded as comprehensive, we conducted computerized data base 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 2002. Finally, we contacted active researchers in the field and made a number of serendipitous discoveries in the course of researching this issue. In totality, our review of the literature covers the years 1939–2002.

2.3. Organization of meta-analyses

The studies have been organized into two categories. The first consists of studies of general population samples for the SPM and APM and presents sex differences for individual years for the ages 6 through 19 and for 10-year age groups of 20–29, 30–39, etc., through 80–89. Data for the SPM and APM are combined because it is considered that the APM measures the same nonverbal reasoning ability as SPM. There are a few data sets whose numbers are too small to meet the criterion of a minimum of 50 males and 50 females for individual years (i.e., for 5, 6 s, etc.) but for which data for several years can be averaged to meet the criterion. In these cases, the sex differences for several years are averaged and entered for the mean year (for instance, data for fifty 6-, fifty 7-, and fifty 8-year-olds would be averaged and entered as 7-year-olds). There are a few data sets where the sex difference is given for a sample with an age range but not for individual years (e.g., for 11- to 13-year-olds but not separately for 11-, 12-, and 13-year-olds). In these cases, the sex differences are entered for the average year. The second category consists of data for the Colored Progressive Matrices. The same methodology is used as for the SPM and APM.

2.4. Strategy of analysis

The analysis broadly followed procedures developed by Hunter and Schmidt (1990). Cohen's d (the difference between the male and female means divided by the within group standard deviation) was adopted 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. In a minority of cases, estimates of the standard deviation were obtained from tables of percentiles. Otherwise, in a very few cases, the effect size was derived from a t ratio using the conversion formula provided by Rosenthal (1991).

2.4.1. Meta-analysis of effect sizes

First, the mean of effect sizes was calculated for each age group, weighted by sample size ($N-1$) (see Table 2). These estimates were then corrected for measurement error. The weighted artefact distributions used in this calculation were derived from those reliability studies reported in Court and Raven (1995) for the Standard, Advanced, and Colored 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 (1990) 75% rule. Each corrected mean *d* score was fitted with a confidence interval. We did not compute credibility intervals since the 95% confidence intervals for corrected mean *d* scores were generally wide (see Table 2), which would suggest that the boundaries of credibility intervals might have been in error by some margin.

3. Results

The results of the studies on sex differences on the SPM and APM are shown in Table 1. The table gives data derived from 57 studies analyzed to provide effect size estimates for 195 samples, with participants numbering a total of 80,928. Samples were considered to be independent within age categories since no study provided more than one estimate per age group. The table gives the location of the study; the size of the male and female samples; the male–female difference in *d* scores with positive signs denoting higher means by males and negative signs higher means by females; the reference; and whether the Standard or Advanced form of the test was used, identified in the notes to the table. The number of effect sizes for each age varied from 1 to 23, and the pooled sample sizes ranged from 200 to 10,708 across age groups. For some age groups (ages 6, 19, 70–79, and 80–89 years old), the sample sizes are small. The reader should therefore exercise caution in interpreting some point estimates, and use the width of confidence intervals as an indicator of the accuracy of corrected mean *d* scores (see Table 2).

Hunter and Schmidt's (1990) corrections for unreliability are controversial (e.g., Rosenthal, 1991). For this reason, Table 2 gives uncorrected mean *d* scores, in addition to the corrected mean *d* coefficients and their associated confidence intervals. The combined reliabilities and variance of reliabilities for the two types of test (APM and SPM), which are necessary in order to calculate the corrected *d* coefficients, were computed using standard weighting procedures (Hunter & Schmidt, 1990) for each age group separately. The estimates for reliability and variance, in this instance, have narrow confidence intervals. This suggests that the corrected *d* coefficients provide a better estimate of the population difference as compared with the uncorrected coefficients: in any case there is little difference between them.

The general trend of the data on the SPM and APM is that boys obtain slightly but not significantly higher means over the ages 6 through 9 years; this is followed by a slight but not significant shift towards higher means for girls over the age range from 10 through 13 years. At the age of 14, an advantage of 0.08*d* for boys begins to appear, increasing to 0.10*d* and becoming statistically significant at the age of 15 and increasing further to a statistically significant 0.17*d* at age 19. The male advantage becomes 0.33*d* among young adults aged 20–29 and remains at approximately this size through all later age groups to 80–89. The sex difference is statistically significant for all the adult age groups except among the 80- to 89-year age group, probably because of the small sample size.

The shape of the growth curve is not established precisely in the current data, but it is clear that whereas there is no sex difference on the progressive matrices among younger children, a difference in favor of males appears in adolescents at about 14–15 years of age and this difference increases in later adolescence and among adults. As indicated by the tests for homogeneity (see Table 2), the fluctuations in the point estimates among adults, within and across age groups, should probably be attributed to sampling errors. The aggregated mean of the corrected *d* scores for the adult age groups is 0.33*d*,

Table 1
Studies of sex differences for SPM and APM

Location	No. of males	No. of females	Age (years)													
			6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Iceland	270	279		.12		.00		.00		−.21		−.68				
2. Estonia	951	882		−.06	−.24	−.16	−.25	−.33								
3. Britain	1625	1625	.18	.12	.01	−.06	−.21	−.29	.13	.00	.10	.09				
4. Poland	2008	1998	.07	−.05	−.11	−.06	−.14	.09	−.07	−.03	−.08	−.04				
5. Mexico	472	448		.16	.10	.02	−.03									
6. Malaysia	2763	2621		−.01	.13	.16	.10	.03	.00							
7. New Zealand	1526	1605			−.02	−.01	−.22	−.14	−.28	.01	.06	−.18	.04			
8. Italy	587	572						.22								
9. Australia	1948	2120			.02	.02	−.15	−.07	.03	.01	−.10	−.04	.19			
10. India	100	100					.28									
11. Argentina	840	840				.06	−.07	−.26	−.18	.11	.62	.46				
12. Iran	2206	2355				.25	.22	.28	.30	.20	.13	.12	.22	.03	.01	
13. North Ireland	93	112				.22										
14. Ireland	605	594				.06			−.17							
15. Ireland	1732	1732				−.06										
16. New Zealand	375	433				−.18		.25		−.43		.19				
17. Lithuania	122	137				.30		−.14								
18. Hong Kong	120	77					.15									
19. Britain	75	95					.00									
20. Denmark	299	329							−.05							
21. Italy	1218	1218						−.04	−.03	.17	.08	.04	−.05	.05		
22. United States	184	172						.00								
23. Israel	370	383						.11	−.49	−.21	−.02	.30	.32	.14		
24. Spain	1621	1650						.02	−.22	.05	.02	.03	−.14	.03	.29	
25. England	2200	2699							−.04							
26. Cuba	640	878							−.02	−.07	.10	.06	.17			
27. India	185	185								−.07						
28. Estonia	1250	1441							−.54	−.26	−.03	−.19	.16	.08	.04	
29. Tanzania—A	413	314								.14	.12	.21	.38			
30. Tanzania—B	1804	836									.15	.30	.11	.06	−.07	
31. Hawaii	503	525									.15	−.05	.10	.05	.10	
32. United States	80	80									−.04					
33. Brazil	80	80									.26					
34. India	1025	1019										.13	.38	.79	.42	
35. England	239	229										.14				
36. South Africa—W	490	566										−.03				
37. South Africa—C	386	381										.19				
38. South Africa—I	530	533										.25				
39. Hong Kong	903	594										.42	.36	.16	.38	
40. Singapore	229	230										.12				
41. South Africa—B	554	539											.31			
42. Italy	303	303												.13		
43. Croatia	174	263												.15		
44. Spain	1147	1565												.13	.10	.16
45. Argentina	53	112												.69		
46. Spain	303	301													.30	
47. Belgium	214	64													.36	

Table 1 (*continued*)

Location	No. of males	No. of females	Age (years)						
			20–9	30–9	40–9	50–9	60–9	70–9	80–9
48. Hawaii	939	971	.16	.49	.39	.44			
49. Britain	300	240	.31	.59	.37	.39	.05	.14	
50. Belgium	850	979	.21	.20	.30	.32	.36	.36	.39
51. Brazil	1921	741	.32	.24					
52. United States	63	80		.16					
53. Hungary	250	250		.04		.31			
54. Israel	100	100			.31				
55. Belgium	101	174					.40	.40	.33
56. United States	92	114				.31			
57. France	564	802					.21		

Minus signs denote higher means obtained by females. 1: Iceland—Pind et al. (2003), age ranges, 6.5–8, 8.5–10, 10.5–12, 12.5–14, and 14.5–16. 2: Estonia—Lynn (2002). 3: Britain—Raven (1981). 4: Poland—Jaworowska and Szustrowa (1991). 5: Mexico—Lynn, Allik, Pullmann, and Laidra (2004), Lynn, Backhoff, and Contreras-Nino (2004). 6: Malaysia—Chaim (1994). 7: North Zealand—Reid and Gilmore (1989). 8: Italy—Young, Tagiuri, Tesi, and Montemagni (1962). 9: Australia—De Lemos (1989). 10: India—Lynn and Jindal (1993). 11: Argentina—Rimoldi, Velasco, San Martin, and Buhrer (1947). 12: Iran—Baraheni (1974). 13: North Ireland—Lynn, Cooper, and Topping (1990). 14: Ireland—Lynn and Wilson (1993); age ranges 8–10 and 11–13. 15: Ireland—Raven (1981); age range 6–12. 16: North Zealand—Reid and Gilmore (1989); age ranges 8–10, 11–12, and 13–15. 17: Lithuania—Lynn 2002. 18: Hong Kong—Lynn, Pagliari, and Chan (1988). 19: Britain—Lynn et al. (1988). 20: Denmark—Vejleskov (1968). 21: Italy—Tesi and Young (1962). 22: United States—Tulkin and Newbrough (1968). 23: Israel—Nathan and Schnabl (1976). 24: Spain—Albade Paz and Muñoz Cantero (1993); APM. 25: England—Adams (1952). 26: Cuba—Alonso (1974). 27: India—Sinha (1968). 28: Estonia—Lynn, Allik et al. (2004), Lynn, Backhoff et al. (2004). 29: Tanzania—Asians—Klingelhofer (1967). 30: Tanzania—Blacks—Klingelhofer (1967). 31: Hawaii—Wilson et al. (1975). 32: United States—Natalicio (1968). 33: Brazil—Natalicio (1968). 34: India—Rao (1975). 35: England—Conrad (1979). 36: South Africa—White—Lynn (2002). 37: South Africa—Colored—Lynn (2002). 38: South Africa—Indian—Lynn (2002). 39: Hong Kong—Lynn and Tse-Chan (2003); APM. 40: Singapore—Lim (1994); APM. 41: South Africa—Black—Lynn (2002). 42: Italy—Young et al. (1962); age range 15–17. 43: Croatia—Matesic (2000), age range 15–19. 44: Spain—Raven (1996); APM. 45: Argentina—Cortada de Kohan (1998); APM. 46: Spain—Colom and Garcia-Lopez (2002). 47. Belgium—Florquin (1964); APM. 48: Hawaii—Wilson et al. (1975). 49: Britain—Heron and Chown (1967). 50: Belgium—Deltour (1993). 51: Brazil—Campos (1999). 52: United States—Sitkei and Michael (1996); age range 16–49. 53: Hungary—Szegedi (1974); age range 15–60. 54: Israel—Guttman (1974); age range 35–61. 55: Belgium—Deltour (1993). 56: United States—Salthouse (2001); APM; age range 18–64. 57: France—Dufouil, Ducimetiere, and Alperovitch (1997); APM.

equivalent to 5 IQ points. In view of the homogeneity test results, this is the best estimate for the male advantage on the progressive matrices among adults.

Table 3 gives results for the Colored Progressive Matrices. There are 15 studies that yielded 42 estimates that were independent within age groups and comprised a total sample with n=60,168. The study of Hsu (1976) provided a disproportionate number of the 6-year-olds. However, inspection of Table 2 shows that the median d (0.14) for that age group does not differ markedly from the weighted mean d. It may therefore be concluded that this study does not distort the estimate of d for 6-year-olds. Examination of the sex differences for each year shows that there was no sex difference at 5 years of age, boys obtain significantly higher means ($P<.05$) of 0.21d, 0.24d, and 0.34d for children aged 6, 7, and 8 years, respectively; there is no difference at ages 9 and 10 years; and boys obtain a significantly higher mean of 0.30d at age 11 years. (A mean effect size is statistically significant when the confidence intervals do not include zero.) The homogeneity tests (see Table 2) indicate that within age group differences largely reflect sampling error and to a lesser extent variation due to attenuation.

Table 2
Magnitude of sex differences on the SPM, APM, and CPM as a function of age

Age	n		k	d	Percentage of variance explained by artefacts	δ	95% confidence interval for δ
	Males	Females					
6	282	285	2	.10	100	.11	−.06 to .27
7	915	945	6	.03	100	.03	−.06 to .09
8	1649	1570	7	.01	58	.01	−.08 to .10
9	4445	4310	15	.01	57	.01	−.05 to .06
10	2337	2227	12	−.03	39	−.03	−.12 to .06
11	2862	2754	14	.05	36	.06	.00 to .14
12	5025	5683	15	−.06	26	−.06	−.15 to .02
13	2321	2275	15	−.02	47	−.02	−.10 to .08
14	2370	2507	16	.07	47	.08	.00 to .16
15	4758	5219	23	.10	33	.10	.03 to .17
16	3025	3184	14	.21	39	.22	.14 to .31
17	2781	2439	13	.15	25	.16	.06 to .27
18	2334	1499	10	.16	40	.18	.08 to .28
19	310	310	1	.16	–	.17	–
20–29	1791	898	4	.30	100	.32	.23 to .46
30–39	1023	830	6	.30	44	.31	.16 to .46
40–49	768	707	4	.37	100	.39	.28 to .50
50–59	560	500	5	.34	100	.37	.24 to .50
60–69	775	1088	4	.24	100	.27	.17 to .37
70–79	202	289	4	.33	100	.34	.16 to .53
80–89	61	139	3	.37	100	.39	−.08 to .87
All adults	5180	4451	10	.30	86	.33	.28 to .37
CPM							
5	347	350	4	.00	100	.00	−.15 to .15
6	25,770	24,880	7	.19	99	.21	.19 to .23
7	1326	1320	8	.21	100	.24	.16 to .31
8	1512	1500	9	.30	35	.34	.21 to .46
9	575	575	6	.06	100	.06	−.04 to .17
10	510	436	4	.02	100	.02	−.11 to .15
11	488	371	4	.28	100	.30	.17 to .44
All children	30,736	29,432	15	.19	45	.21	.19 to .23

Exceptionally, at age 8, artefacts explained only 45% of variance in d scores, suggesting, somewhat anomalously, the presence of a moderator. Additionally, the homogeneity test for all children (45% of variance explained by artefacts) suggests that differences among age groups reflect both real differences and sampling error. Since it is not possible to disentangle real differences from sampling error, a reasonable treatment of the data is to average the results for the six age groups giving a corrected weighted mean of 0.21*d*, equivalent to a higher mean for boys of 3.2 IQ points. The advantage of boys on the SPM averaged over the same age range is only 0.02*d*. Homogeneity of effect sizes is also supported for children of 6–8 years of age, whereas heterogeneity of effect sizes is observed for children and adolescents from 9 to 18 years of age, indicating the presence of one or more moderator variables.

We now examine Mackintosh's (1998) suggestion that while there may have been a sex difference on the progressive matrices in earlier times, it became virtually zero in the late 20th century. In addition, we

Table 3
Sex differences on the Colored Progressive Matrices

Location	No. of males	No. of females	Age (years)						
			5	6	7	8	9	10	11
1. Germaniy	717	704	−.03	.22	.18	.53			
2. New Zealand	64	87			.53				
3. United States	49	55	.02						
4. Australia	350	350	−.02	.14	.31	.02	.11	.20	.32
5. Brazil	773	773	.06	.25	.12	.31	.06	−.04	.50
6. India	546	471			.21				
7. Switzerland	146	144		.13					
8. Taiwan	22,542	21,283		.20					
9. Hong Kong	2217	2641		.12					
10. Taiwan	991	974		.14	.25	.32			
11. United States—Black	164	185			.09	.12	.31		
12. United States—White	225	215			.01	.01	.04		
13. India	940	594				.11	−.04	−.06	.17
14. Belgium	477	368				.91	.19	.19	.00
15. Kenya	583	639				.30			

1: Germany—von Winkelmann (1972). 2: North Zealand—Freyberg (1966). 3: United States—Levinson (1960). 4: Australia—Reddington and Jackson (1981). 5: Brazil—Angelini, Alves, Custodio, Duarte, and Duarte (1999). 6: India—Rao and Reddy (1968). 7: Switzerland—Dupont (1970). 8: Taiwan—Hsu (1976). 9: Hong Kong—Chan and Lynn (1989). 10: Taiwan—Hsu (1971). 11: United States—Black—Higgins and Sivers (1958). 12: United States—White—Higgins and Sivers (1958). 13: India—Despande (1971). 14: Belgium—Goosens (1952). 15: Kenya—Costenbader and Ngari (2000).

examine a possible explanation of the heterogeneity of effect sizes observed for children and adolescents from 9 to 18 years of age, which may reside in different rates of maturation across ethnic groups. Analyses to test for these possibilities were carried out using maximum likelihood-based weighted regression analysis, as implemented within the ANOVA and regression programs of Stata. Since these procedures produce correct estimates of sums of squares, the corrections prescribed by Hedges and Becker (1986) for standard weighted regression were unnecessary.

In the first analysis, at Step 1, age was entered as a control variable. The midpoint of the range served as the estimate of age in adult samples. At Step 2, we entered the birth date of the cohort, estimated by subtracting the age of each sample from the date on which the study was published in order to test for generational effects. As previously established, the analysis revealed a significant effect of age on effect size [ΔR^2=.19, β=.44, $t_{(1,162)}$=5.14, P=.00]. However, once age was controlled, there was no evidence of a significant effect of the birth date of each cohort [ΔR^2=.00, β=.01, $t_{(1,162)}$=.07, P=.95] on the magnitude of the sex difference with respect to the progressive matrices. Consequently, Mackintosh's (1998) suggestion that sex differences on the progressive matrices have declined among the younger generation is not supported.

To provide a further check, we examined the correlation between the birth date of cohorts and effect size among adults in the age range 20–89 years. Since the data are consistent with a sex difference in reasoning ability, among adults (which is approximately constant in magnitude) a zero correlation would indicate that the magnitude of the sex difference in reasoning ability among younger cohorts has not declined. The obtained correlation of −.15 (P=.44) is thus supportive of this conclusion.

In the second analysis, at Step 1 we again entered age and at Step 2 we entered ethnicity, which was coded (1=East Asian, 2=White, 3=South Asian, 4=Black, Missing=Other). Finally, at Step 3, we entered an interaction term of ethnicity with age. In this instance, there was evidence of a significant interaction between age and ethnicity [$F_{(3,140)}$=3.2, P=.03]. These results suggest that the developmental trends in the magnitude of sex differences on the progressive matrices differ across ethnic groups. However, post hoc tests show that it is only the Black (n=6) and South Asian groups (n=11) that exhibit a significantly different developmental trend from the Whites. These samples are small and may reflect nothing more than sample bias. For example, in Africa, dropout rates from school are high so school samples may be biased in favor of either sex, depending on local customs.

4. Discussion

There are eight points of interest in the results of this meta-analysis. First, it was designed to test the Eysenck (1981)–Court (1983)–Mackintosh (1996)–Jensen (1998) hypothesis that there is no sex difference in mean scores on the progressive matrices. The results given in Tables 1 and 2 show that this is correct for the SPM for the age range 6–14 years. However, it is incorrect for the age of 15 years onwards. From this age into adulthood, males consistently obtain higher means than females and all of these differences are statistically significant except for the 80–89 age group, for which the difference is of approximately the same magnitude and the lack of statistical significance is attributable to the small sample size. The advantage of boys begins to appear at the age of 14 and increases in size among adults when it reaches an average for the whole age range of 20–29 through 80–89 of 0.33d, equivalent to 5 IQ points. It is proposed that this is the best estimate of the male advantage on the progressive matrices among adults.

Second, the hypothesis that there is no difference on the progressive matrices also fails for the Colored Progressive Matrices for the age range 5 though 11 years shown in Tables 2 and 3. The results show that boys obtain significantly higher means than girls at ages 6, 7, 8, and 11 years, and that for the whole age range boys have a mean advantage of 0.21d, equivalent to 3.2 IQ points. However, on the SPM over this age range, boys have a negligible advantage of 0.02d.

It is proposed that the explanation for the greater advantage of boys on the Colored Progressive Matrices than on SPM is probably that the progressive matrices is not a pure measure of reasoning ability, as has been frequently asserted, but also measures visualization. This was suggested by Van der Ven and Ellis (2000) who have shown in a Rasch analysis for unidimensionality of the SPM, in a sample of 905 Dutch 12- to 15-year-olds, that the test contains two factors identified as (1) “gestalt continuation” present in Set A and items 1–6 in Set B for which “the correct solution must be found according to some Gestalt continuation rule”; and (2) “analogical reasoning,” present in items B 8–12 and in most of the later items. We prefer the word “visualization” to “gestalt continuation” as the widely accepted term for this ability in hierarchical factor models such as that of Carroll (1993). It is known from the meta-analysis of sex differences in spatial abilities carried out by Linn and Peterson (1985) that boys tend to perform a little better on average than girls on visualization. The Colored Progressive Matrices contain more visualization items than the SPM, namely, most of Set Ab, which is absent in the SPM. We think it probable that boys perform better than girls on these visualization items and that this is the explanation for the greater male advantage on the Colored Progressive Matrices. This is a hypothesis that needs to be examined.

Three, the results provide tests of two components of the developmental theory of the sex differences in the maturation of intelligence. The component of the theory for the age range 6–14 years is not supported by the results. This states that the sex difference is negligible up to the age of about 8 years; moves in favor of girls at the age of about 9 years as the accelerated growth spurt of girls speeds up their development in a number of characteristics including height, weight, and brain size; and remains in favor of girls for the next 2 or 3 years up to an age of around 12 years, following which there is little difference between boys and girls up to the age of 15. The results set out in Table 1 show trends supporting this theory but these are not statistically significant. The second component of the theory stating that from the age of 15 years onwards males obtain higher means than females is confirmed by the results.

Four, the results do not support Mackintosh's (1998) contention that the male advantage has been greater among older cohorts of adults than among young adults. Mackintosh's contention is that among adults the size of the sex difference has declined such that it is negligible in the 20–29 age group even though substantial in older age groups. Informally, it can be seen from inspection of the magnitude of the sex differences for the age groups 20–29 through 80–89 set out in Table 2 that there is no tendency for the difference among the younger age groups to be smaller than among older. The sex differences for the 20- to 49-year-olds and for the 60- to 89-year-olds are virtually identical at 0.34 and 0.33, respectively. The finding of a nil correlation between age of cohort and effect size among adults provided a more precise test supportive of the same conclusion. Mackintosh's hypothesis is refuted more formally by the regression analysis, which showed no significant effect of generational age on the magnitude of the sex difference in scores on the progressive matrices. According to Hunter and Schmidt's (1990) 75% rule, there was evidence of moderator variables for children and adolescents of 8–18 years of age, though not among adults, except for an anomalous finding for the groups aged 30–39 years. However, our test for moderator variables using multiple-regression excluded generational effects as a possible moderator.

Five, the conclusion that from the age of 15 years males have higher mean scores than females has implications for the question of sex differences in general intelligence, fluid intelligence (Gf), and Spearman's *g*. Both Jensen (1998) and Mackintosh (1998) draw significant conclusions from their reading of the evidence that there is no sex difference on the progressive matrices. Jensen (p. 541) argues that because the progressive matrices is the best measure available of Spearman's *g*, it follows that there is no sex difference in Spearman's *g*. Mackintosh (p. 538) draws a similar conclusion that because the progressive matrices is an excellent measure of Gf (fluid intelligence) and that Gf can be identified with "general intelligence," there is no sex difference in either Gf or in general intelligence. The results of our meta-analysis show that if these identifications are correct, males from the age of 15 years onwards have higher average *g*, fluid intelligence, and "general intelligence" than females by approximately 5 IQ points.

Six, we now consider the issue of possible cross-cultural variation in growth curves. There was some evidence that the patterns in growth curves of sex differences on the progressive matrices are culture specific. However, it is evident that all cultures examined in the analysis converge on the same end point by late adolescence and early adulthood. It should be noted that 9 of the 10 studies of adults from the age of 20–80 are from economically developed nations. The remaining study is from Brazil and the sex difference of 0.28*d* in the Brazil sample is a little less than the 0.33*d* average for all 10 studies.

Seven, while our results support Lynn's (1994, 1998, 1999) developmental theory of sex differences in intelligence, other explanations of the gradual increase in the mean scores obtained by males relative to those of females from age 15 should be considered. First, there is the possibility of differential experience of men and women. In all societies, women are predominantly responsible for the care of children (Whiting & Whiting, 1975), and even professional women have been shown to carry out the majority of domestic work (Yogev, 1982). It may be possible that this specialization could affect scores on tests of cognitive ability. Lippa (1998) has shown that there are male and female differences on Prediger's (1982) People–Things dimension such that males are more interested in things and females in people. It has been shown further that genetic variation explained 53% of the variance in within-sex differences in gender diagnosticity, a variable highly correlated with the People–Things dimension (Lippa & Herschenberger, 1999). A possible interpretation of these findings might be that women's genetic predispositions are such as to maximize their satisfaction with their traditional adult roles and these predispositions are subsequently amplified by gender role socialization (Lippa, 2002; Maccoby, 2000; Ridley, 1993). That the resulting differences in experience can influence scores on intelligence tests is suggested by a study of Ackerman, Bowen, Beier, and Kanfer (2001), which showed that social potency, social closeness, and traditionalism–worry are all related to lowered scores on general knowledge, a measure of crystallized intelligence (Carroll, 1993), i.e., those who are socially oriented, among whom women predominate, invest their cognitive resources differently from those who are not socially oriented, and this difference accounts for the lower scores on general knowledge.

Whether this difference would affect nonverbal reasoning ability may be more doubtful. However, studies of experts also suggest how experience might influence fluid intelligence (Ericsson & Kintsch, 1995). They have shown that the major factor that appears to differentiate experts at chess and medical diagnosis from those who are average is the existence of long-term memory structures that augment the capacity of working memory. Since Kyllonen and Christal (1990) have provided the adult evidence that fluid intelligence can be roughly equated to the capacity of working memory, a difference in long-term memory structures due to different adult experiences among men and women may, by augmenting working memory capacity, be responsible for an observed difference in scores on fluid intelligence. The above theory, which posits the joint action of biological and social factors as proximal causes of gender differences, corresponds to what Feingold (1994) describes as a biosocial model.

While differential experience may provide an explanation for the adult sex differences in scores on the progressive matrices, there are two aspects of the data that seem incompatible with this. First, if differential experience comprises the underlying mechanism, we would expect cross-cultural variation in sex differences on the progressive matrices since adult sex roles differ across cultures. Yet we did not find this. Second, given the large changes in the adult roles of women in western industrialized societies postwar, we should expect a reduction in the sex difference over time, yet we found no evidence of generational change.

Eight, whatever the causes of the observed pattern of sex differences on the progressive matrices, it would seem desirable to place the overall difference of 0.33*d* in adults in context. There is considerable recent evidence that from a position that obtained about 20 years ago, when boys and men consistently outperformed girls and women in terms of educational performance, on many indices girls and women are increasingly either equaling or overtaking boys and men. For example, in the United Kingdom, in 1980, 13% of young women obtained two or more "A" levels or their equivalent compared with 14% of young men (Ramprakash & Daly, 1983), whereas by the year 2000 the respective figures were 39% and 31% in favor of young women (Matteson & Babb, 2002).

In terms of higher education, in 1980 only 37% of first degrees were obtained by women (Ramprakash & Daly, 1983), but by 2001 this figure had risen to 56% (Matteson & Babb, 2002). Up to 1997, men obtained more first class honors degrees that women but in 1998 women for the first time gained more first class honors degrees than men (Higher Education Statistics Agency, 1998). Given that this increasing female advantage in educational achievement coexists with somewhat lower scores among adult women on the progressive matrices, it can be inferred that there are other factors predominantly possessed by women that facilitate this achievement. Possibly, this may be stronger work motivation. Thus, it has been found in the United States that women obtain lower mean scores on the SAT-M but they did not obtain lower math grades (Wainer & Steinberg, 1992). The most probable explanation is that women's stronger work motivation compensates for their lower test scores.

References

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*, 797–825.

Adams, E. A. (1952). *Analysis of Raven's matrices scores*. Unpublished.

Albalde Paz, E., & Muñoz Cantero, J. (1993). *El test PMS de Raven y los escolares de Galicia*. Universidade da Coruña: Servicio de Publicaciones.

Alonso, O. S. (1974). Raven, *g* factor, age and school level. *Havana Hospital Psiquiatrico Revista*, *14*, 60–77.

Angelini, A. L., Alves, I. C., Custodio, E. M., Duarte, W. F., & Duarte, J. L. (1999). *Matrizes progressivas Coloridas de Raven*. Sao Paulo: Centro Editor de Testes.

Baraheni, M. N. (1974). Raven's Progressive Matrices as applied to Iranian children. *Educational and Psychological Measurement*, *34*, 983–988.

Campos, F. (1999). *Teste das Matrizes Progresivas Escala Geral*. Rio de Janeiro: Cepa.

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.

Chaim, H. H. (1994). *Is the raven progressive matrices valid for Malaysians?* Unpublished.

Chan, J., & Lynn, R. (1989). The intelligence of six year olds in Hong Kong. *Journal of Biosocial Science*, *21*, 461–464.

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

Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, *32*, 445–451.

Conrad, R. (1979). *The deaf schoolchild: Language and cognitive function*. New York: Harper & Row.

Cortada de Kohan, N. (1998). Logos en educacion secundaria y su relacion con intelligencia. *Revista Latinoamericana de Psicologia*, *30*, 293–310.

Costenbader, V., & Ngari, S. M. (2000). A Kenya standardisation of the Coloured Progressive Matrices. *School Psychology International*, *22*, 258–268.

Court, J. C. (1980). *Researchers' bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary Scales*. Australia: Flinders University.

Court, J. C., & Raven, J. (1995). *Normative, reliability and validity studies*. Oxford, UK: Oxford Psychologists Press.

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

De Lemos, M. (1989). *Standard progressive matrices: Australian manual*. Hawthorn, Victoria: Australian Council for Educational Research.

Deltour, J. J. (1993). *Echelle de Vocabulaire Mill Hill de J.C. Raven.* Braine de Chateau. Belgium: Editions L'Application des Techniques Modern S.P.R.L.

Despande, M. V. (1971). Sex differences on Raven's Matrices Test. *Journal of Psychological Researches*, *15*, 101–103.

Dufouil, C. D., Ducimetiere, P., & Alperovitch, A. (1997). Sex differences in the association between alcohol consumption and cognitive performance. *American Journal of Epidemiology*, *146*, 405–412.

Dupont, J. B. (1970). *Sex differences on the PM 47 in Switzerland.* Unpublished.
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
Eysenck, H. J. (1981). *Intelligence: The battle for the mind*. London: Pan.
Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*, 429–456.
Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, *48*, 71–79.
Florquin, F. (1964). Les PM 1947 de J.C.Raven au niveau des classes terminals du cycle secondaire. *Revure Belge de Psychologies et de Pedagogie*, *26*, 108–120.
Freyberg, P. S. (1966). The efficacy of the Coloured Progressive Matrices as a group test with young children. *British Journal of Educational Psychology*, *36*, 171–177.
Garrity, L. I., & Donaghue, J. T. (1976). Preschool children's performance on the Raven's Coloured Progressive Matrices and Peabody Picture Vocabulary Test. *Educational and Psychological Measurement*, *36*, 1043–1047.
Goosens, G. (1952). Etalonnage du Matrix 1947 de J.C. Raven. *Revue Belge de Psychologie et de Pedagogie*, *14*, 74–80.
Guttman, R. (1974). Genetic analysis of analytic spatial ability. *Behavior Genetics*, *4*, 273–284.
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.
Higgins, C., & Sivers, C. H. (1958). A comparison of Stanford-Binet and colored Raven progressive matrices IQs for children with low socioeconomic status. *Journal of Consulting Psychology*, *22*, 465–468.
Higher Education Statistics Agency Online Information Service. *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.
Hsu, C. (1976). The learning potential of first graders in Taipei city as measured by Raven's Coloured Progressive Matrices. *Acta Pediatrica Sinica*, *17*, 262–274.
Hsu, C. C. (1971). Chinese children's responses on the Coloured Progressive Matrices. *Journal of the Formosan Medical Association*, *70*, 51–63.
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.
Jaworowska, A., & Szustrowa, T. (1991). *Podrecznik Do Testu Matryc Ravena*. Warsaw: Pracownia Testow Psychologicznych.
Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.
Klingelhofer, E. L. (1967). Performance of Tanzanian secondary school pupils on the Raven Standard Progressive Matrices Test. *Journal of Social Psychology*, *72*, 205–215.
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.
Kyllonen, P. C., & Christal, RE. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.
Levinson, B. M. (1960). A comparative study of the verbal and performance ability of monolingual and bilingual native born Jewish preschool children of traditional Jewish parentage. *Journal of Genetic Psychology*, *97*, 93–112.
Lim, T. K. (1994). Gender-related differences in intelligence: Application of confirmatory factor analysis. *Intelligence*, *19*, 179–192.
Linn, M. C., & Peterson, A. C. (1985). Emergence and characterisation of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498.
Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the People–Things dimension. *Journal of Personality and Social Psychology*, *74*, 996–1009.
Lippa, R. (2002). *Gender, nature and nurture*. Mahwah, NJ: Lawrence Erlbaum Associates.
Lippa, R., & Herschenberger, S. (1999). Genetic and environmental influences on individual differences in masculinity, femininity, and gender diagnosticity: Analyzing data from a classic twin study. *Journal of Personality*, *67*, 127–155.
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. (2002). Sex differences on the progressive matrices among 15–16 year olds in South Africa. *Personality and Individual Differences*, *33*, 669–677.
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. (2004). Sex differences on the Progressive Matrices among adolescents: Some data from Estonia. *Personality and Individual Differences*, *36*, 1249–1255.
Lynn, R., Backhoff, E., & Contreras-Nino, L. A. (2004). Sex differences on *g*, reasoning and visualization tested by the Progressive Matrices among 7 to 10 year olds: Some normative data for Mexico. *Personality and Individual Differences*, *36*, 779–787.
Lynn, R., Cooper, C., & Topping, S. (1990). Reaction times and intelligence. *Current Psychology Research and Reviews*, *9*, 264–276.
Lynn, R., & Jindal, S. (1993). Positive correlations between brain size and intelligence: Further evidence from India. *Mankind Quarterly*, *34*, 109–115.
Lynn, R., Pagliari, C., & Chan, J. (1988). Intelligence in Hong Kong measured for Spearman's *g* and the visuospatial and verbal primaries. *Intelligence*, *12*, 423–434.
Lynn, R., & Tse-Chan, P. W. (2003). Sex differences on the progressive matrices: Some data from Hong Kong. *Journal of Biosocial Science*, *35*, 145–154.
Lynn, R., & Wilson, R. G. (1993). Sex differences in cognitive abilities among Irish primary and secondary school children. *Irish Journal of Psychology*, *14*, 293–300.
Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–572.
Mackintosh, N. J. (1998). Reply to Lynn. *Journal of Biosocial Science*, *30*, 533–539.
Maccoby, E. E. (2000). Parenting and its effect on children: On reading and misreading behaviour genetics. *Annual Review of Psychology*, *51*, 1–27.
Matesic, K. (2000). Relations between results on Raven's Progressive Matrices plus sets and school achievement. *Review of Psychology*, *7*, 75–82.
Matteson, J., & Babb, P. (2002). *Social Trends, No. 32*. London: The Stationary Office.
Natalicio, L. (1968). Aptidatao general, status social e sexo: Um estudio de adolescentes Brasilieros e norte-Americanos. *Revista Interamericana de Psicologia*, *2*, 25–34.
Nathan, M., & Schnabl, A. (1976). The effect of Kibbutz children's age and sex on performance in the Raven Progressive Matrices Test. *The Kibbutz*, *3–4*, 105–113.
Pind, J., Gunnarsdóttir, E. K., & Jóhannesson, H. S. (2003). Raven's standard progressive matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences*, *34*, 375–386.
Prediger, D. J. (1982). Dimensions underlying Holland's hexagon: Missing link between interests and occupations? *Journal of Vocational Behavior*, *21*, 259–287.
Ramprakash, D., & Daly, M. (1983). *Social Trends*, *vol. 13*. London: HMSO.
Rao, S. N., & Reddy, I. K. (1968). Development of norms for the Raven's Coloured Progressive Matrices Test. *Psychological Studies*, *13*, 105–107.
Rao, Y. R. (1975). Distribution of progressive matrices scores among pupils of XI standard. *Indian Journal of Applied Psychology*, *12*, 76–82.
Raven, J. (1981). *Irish and British standardisations*. Oxford: Oxford Psychologists Press.
Raven, J. (1996). *Raven matrices progressivas manual*. Madrid: Publiciones de Psicologia Aplicada.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced Progressive Matrices*. Oxford, UK: Oxford Psychologists Press.
Raven, J. C. (1939). The RECI series of perceptual tests: An experimental survey. *British Journal of Medical Psychology*, *18*, 16–34.
Raven, J. C., Court, J. H., & Raven, J. (1996). *Raven matrices progressivas manual*. Madrid: Publicaciones De Psicologia Aplicada.
Reddington, M. J., & Jackson, K. (1981). The Raven's Coloured Progressive Matrices: A Queensland standardisation. *Australian Council for Educational Research Bulletin*, *30*, 20–26.
Reid, N., & Gilmore, A. (1989). The Raven's standard progressive matrices in New Zealand. *Psychological Test Bulletin*, *2*, 25–35.
Ridley, M. (1993). *The red queen*. London: Penguin.

Rimoldi, H. J., Velasco, E. S., San Martin, R., & Buhrer, L. (1947). *Tipicacion de los Progressive Matrices de Raven*. Mendoza, Argentina: Universidad Nacional de Cuyo.

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

Salthouse, T. (2001). Personal communication.

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.

Sinha, U. (1968). The use of Raven's Progressive Matrices in India. *Indian Educational Review*, *3*, 75–88.

Sitkei, G. E., & Michael, W. B. (1966). Predictive relationships between items on the Revised Stanford-Binet Intelligence Scale and total scores on Raven's Progressive Matrices. *Educational and Psychological Measurement*, *26*, 501–506.

Szegedi, M. (1974). A Raven-fele intelligencia test. *Magyar Pszichologiai Szemle*, *31*, 194–201.

Tesi, G., & Young, H. B. (1962). A standardisation of Raven's Progressive Matrices. *Archive de Psicologia Neurologia e Pscichologia*, *5*, 455–464.

Tulkin, S. R., & Newbrough, J. R. (1968). Social class, race and sex differences on the Raven standard progressive matrices. *Journal of Consulting and Clinical Psychology*, *32*, 400–406.

Van der Ven, A.H.G.S., & Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences*, *29*, 45–64.

Vejleskov, H. (1968). An analysis of Raven matrix responses in fifth grade children. *Scandinavian Journal of Psychology*, *9*, 177–186.

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.

Whiting, B. B., & Whiting, J. W. M. (1975). *Children of six cultures: A psycho-cultural analysis*. Cambridge, MA: Harvard University Press.

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

Winkelmann, W. von (1972). Normen fur den Mann-Zeichen-Test von Ziler und die Coloured Progressive Matrices von Raven fur 5–7 jahrige kinder. *Psychologische Beitrage*, *17*, 80–94.

Yogev, S. (1982). Are professional women overworked? Objective versus subjective perception of role loads. *Journal of Occupational Psychology*, *55*, 165–170.

Young, H. B., Tagiuri, R., Tesi, G., & Montemagni, G. (1962). Influence of town and country upon children's intelligence. *British Journal of Educational Psychology*, *32*, 151–158.