Negligible Sex Differences in General Intelligence

ROBERTO COLOM
MANUEL JUAN-ESPINOSA
FRANCISCO ABAD
LUÍS F. GARCÍA
Universidad Autónoma de Madrid (UAM), Madrid, Spain

The general factor, *g*, can be extracted from a correlation matrix of a battery of mental ability tests. *g* is common to all mental abilities. A key question in the research on cognitive sex differences is whether, on average, females and males differ in *g*. This question is technically the most difficult to answer and has been the least investigated. Cognitive batteries were applied in the present study to independent samples totaling 10,475 adult subjects (4,256 females and 6,219 males). The scores were factor-analyzed by sex to obtain separate *g* factors. The congruence coefficients (r_c) suggested a near identity of these factors. Then, three methods were used to know if the standardized sex differences (*ds*) are explained by *g*: (1) the method of correlated vectors; (2) the sex loading in *g* was computed including the point-biserial correlation between sex and each of the subtests in the full matrix of subtest intercorrelations for factor analysis; and (3) the correlation between sex and *g* factor scores. The results suggest a negligible sex difference in *g*. The present study includes the largest sample on which a sex difference in *g* has ever been tested. The findings are consistent with those using quite different test batteries and subject samples.

INTRODUCTION

Some issues on sex differences in IQ as an estimate of general intelligence have been prompted by the finding of a significant sex difference in brain size (Ankey, 1992, 1995; DeLacoste, Adesanya, & Woodward, 1990; Lynn, 1994; Willerman, Schultz, Rutledge, & Bigler, 1991). A "paradox" concerning sex differences in intelligence and brain size has been noted by Ankey (1992): males have, on average, larger brains than females and brain size is positively correlated with IQ. It would be expected that males would have a higher average level of IQ than females. Yet it is generally stated than there are no overall differences in the scores obtained by males and females on IQ (Brody, 1992; Colom, 1998; Halpern, 1992; Juan-Espinosa, 1997). There seems to be a logical inconsistency among the findings of larger male brain, the association of brain size with IQ, and the absence of a sex difference in overall IQ that calls for a resolution (Lynn, 1994).

Direct all correspondence to: Roberto Colom, Facultad de Psicología, Universidad Autónoma de Madrid, 28049 Madrid, Spain. E-mail: roberto.colom@uam.es

INTELLIGENCE 28(1): 57–68
ISSN: 0160-2896

Ankey (1992) accepts the view that there is no sex difference in IQ and that females obtain higher means on verbal abilities, while males obtain higher means on spatial abilities (Hyde & Linn, 1988; Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). The solution to the "paradox" is that spatial ability may require more brain tissue than verbal ability. McGlone (1980) has proposed that females have verbal abilities located in their right as well as in their left hemispheres. If this is correct, females must have about the same amount of brain tissue devoted to verbal abilities as males that explains the fact that female verbal abilities are about the same as those of males. Female spatial abilities will be rather substantially weaker than those of males, because the female right hemisphere is smaller than that of a male (and some of it has been given over to verbal abilities).

Rushton (1992) has proposed another solution to the "paradox": males and females do not have the same mean IQ. This is the solution favored by Lynn (1994) who has argued that a difference of 4 IQ points favoring males would be consistent with the prediction from the calculated sex differences in brain size. In a review of the literature, he has found an overall sex difference of precisely 4 IQ points. Males have a brain size advantage of 0.78 *SD* units. The correlation between brain size measured by magnetic resonance imaging and intelligence has been calculated by Willerman et al. (1991) at 0.35 (see Rushton & Ankey, 1996, for a review). The male advantage for intelligence accruing from greater brain size is therefore $0.78 \times 0.35 = 0.27$ *SD* units = 4 IQ points.

However, as Jensen (1998) has stated, any overall difference on a collection of tests, even if significant, like the one reported by Lynn (1994), has questionable generality across batteries and cannot answer the question concerning a sex difference in general ability defined as *g*.

The empirical fact that all mental abilities are positively correlated calls for an analytic taxonomy of mental abilities based on some form of correlation analysis. The dimensions found in the factor analysis of the correlations among a large variety of mental ability measurements can be arranged hierarchically according to their generality (Carroll, 1993, 1997). The *g* factor is the most general of all and is common to all mental abilities. *g* may be thought of as a distillate of the common source of individual differences in all mental tests. *g* can be roughly likened to a computer's central processing unit. The knowledge and skills tapped by mental test performance merely provide a *vehicle* for the measurement of *g* (Jensen, 1992). *g* is best regarded as a source of variance in performance associated with individual differences in the speed or efficiency of the neural processes that affect the kinds of behavior called mental abilities (Jensen, 1998).

Concerning sex differences in general intelligence defined as *g*, the statement of McArdle (1996) that equality of factor loadings should be established before other group comparisons (e.g., mean differences) were considered is worth noting. If not, the psychological constructs being measured may be qualitatively different for the groups being compared.

Several studies of factorial similarity have been conducted for cognitive ability (Carreta & Ree, 1995; DeFries et al., 1974; Humphreys & Taber, 1973; Loehlin, Lindzey, & Spuhler, 1975; Michael, 1949; Ree & Carretta, 1995). These studies found no difference in factor structure across groups. Carretta and Ree (1995) found that the correlation of the *g* loadings for males and females for the hierarchical *g* factor was +0.97. Carretta and Ree (1997) found that the correlation between the male and female factor loadings on *g* was

+0.999. Therefore, their results add to the findings that the structure of cognitive abilities is nearly identical across sex.

The most problematic question in the research on cognitive sex differences is whether, on average, males and females differ in *g*. This question is technically the most difficult to answer and has been the least investigated. The vast majority of studies have been looked at sex differences in more specialized abilities (Jensen, 1998).

The method of correlated vectors is one way of testing whether the *g* factor extracted from a battery of diverse tests is related to some variable, *X* (Jensen, 1998, Appendix B). The significance level is determined from the rank-order correlation between the elements in the column vector of the various tests' *g* loadings and the elements in the column vector of the variable *X*. Concerning sex differences, the Spearman rank-order correlation (r_s) of the column vector of subtests' *g* loadings with the vector of the sex differences (*d*) on the subtests indicates the degree to which *g* is related to the rank-order of the sex differences on the various subtests. Jensen (1998) has found that "the method of correlated vectors shows that in no case is there a correlation even approaching significance between subtests's *g* loadings and the mean sex difference on the various subtests" (p. 546). His conclusion is that the sex difference in psychometric *g* is nonexistent.

The current study investigated whether the null hypothesis can be rejected, that is, whether there is no sex difference in general intelligence defined as *g*. It is important to have cumulative evidence from different batteries and subject samples, because as Carroll (1997) has stated, "*g* ... is likely to be present, in some degree, in nearly all measures of cognitive ability. Furthermore, it is an important factor, because on the average over many studies of cognitive ability tests it is found to constitute more than half of the total common factor variance in a test" (p. 31).

For testing the hypothesis, the method of correlated vectors was used. Furthermore, (1) the sex difference in *g* was represented on each of the subtests in terms of a point-biserial correlation, including these correlations with the full matrix of subtests intercorrelations for factor analysis, and (2) it was computed a single *g* factor score, for testing a sex difference.

Method

Participants

The participants were applicants for admissions to a private university between 1989 and 1995. They were 6,879 adult subjects (2,743 females and 4,136 males) in the first sample, and 3,596 adult subjects (1,513 females and 2,083 males) in the second sample. Their ages at time of testing have a mean of 23.12 years and the standard deviation was 2.17. Therefore, the present study tests two independent samples totaling 10,475 adult subjects (4,256 females and 6,219 males).

Measures and Procedures

The batteries of cognitive tests were applied collectively to independent samples and in groups of no more than 30 subjects each. The batteries were applied as part of a selection procedure to applicants for a private university in Madrid. The list with the tests in the batteries follows.

Table 1. Descriptive Data and *ds* (Standardized Mean Differences) in the First Sample

	Males			*Females*			
Tests in the first battery	*M*	*SD*	*N*	*M*	*SD*	*N*	*d*
Vocabulary (PMA)	28.6	7.2	4,136	27.5	6.9	2,743	0.15
Verbal Fluency (PMA)	48.02	11.76	4,132	49.82	11.55	2,739	−0.15
Spatial Rotation (PMA)	30.92	11.56	4,134	26.32	11.47	2,740	0.4
Inductive Reasoning (PMA)	19.05	4.92	4,131	19.56	4.75	2,742	−0.10
Monedas	25.54	6.65	4,086	20.45	6.67	2,716	0.76

The first battery includes five cognitive tests. Four of them were from the Primary Mental Abilities (PMA) battery: Vocabulary, Verbal Fluency, Spatial Rotation, and Inductive Reasoning. The fifth test was developed in Spain and is called "Monedas." "Monedas" is based on the combination of the size of a series of coins, the digits put inside the coins to specify the number of them that the subject must take into account, and some numerical operations to make the necessary calculations to arrive at a given response (adding, subtracting, and so forth). "Monedas" correlates $r = +0.64$ with the Numerical Ability (NA) Scale from the DAT. The second battery includes six cognitive tests. Three of them are from the PMA battery: Vocabulary, Spatial Rotation, and Inductive Reasoning. The fourth one is "Monedas." The other two scales are: Verbal Reasoning, the VR scale from the Differential Aptitude Test (DAT), and Dominoes, a test of abstract reasoning.

There are two basic criteria for extracting a good *g* (Jensen & Weng, 1994): number and variety of tests. The tests in the present study's batteries call for more than three primary factors (variety) and the number of tests is enough to a factor solution based on the first unrotated principal factor (the extraction of *g* as a second-order factor in a hierarchical analysis requires a minimum of nine tests from which at least three primary factors can be obtained (Jensen & Weng, 1994); therefore, we cannot perform a hierarchical analysis).

Note that most of the applicants do not reach the score level required for admissions to the State University. What this could imply is that the sample is more representative of the general population than samples taken from applicants to the State University. Anyhow, there is perhaps some statistical sampling error that could be attributed to the present study. The *g* factor is extracted from a sample taken from a segment of the population that does not display the full range of mental ability that exists in the total population. Therefore, the *g* extracted could be smaller than it would be if extracted from data for the general population. But note that this sampling error is not higher than the one that could be attributed to studies like, for instance, the one conducted by Ree and Earles (1991) at Brooks Air Force Base with 9,173 recruits. Looking at the test manuals (Spanish standardizations), the *SD* values are: Vocabulary = 7.95; Verbal Fluency = 9.15; Spatial Rotation = 11.87; Inductive Reasoning = 5.20; Monedas = 5.64; Verbal Reasoning = 7.63; Dominoes = 6.63. Therefore, there are no sharp differences with the *SD* values of the samples in the present study (see Tables 1 and 2). On the other hand, there is the possibility that females are self-selecting to apply to the University. Although there are more males than females, all the subjects are applicants for admissions to a wide range of careers

Table 2. Descriptive Data and *ds* (Standardized Mean Differences) in the Second Sample

	Males			*Females*			
Tests in the second battery	*M*	*SD*	*N*	*M*	*SD*	*N*	*d*
Vocabulary (PMA)	27.27	6.8	2,002	26.1	6.8	1,477	0.17
Verbal Reasoning (DAT)	31.57	7.79	1,999	29.85	7.64	1,476	0.22
Monedas	26.27	6.4	2,005	21.45	6.4	1,480	0.75
Inductive Reasoning (PMA)	18.63	4.76	1,997	19.55	4.8	1,471	−0.19
Dominoes	26.21	6.3	1,993	25.62	6.3	1,480	0.09
Spatial Rotation (PMA)	28.56	10.75	2,002	24.97	11	1,479	0.33

(faculties): humanities, social science, natural science, engineering, and so forth. Thus, it seems that the probability that the subjects are representing different portions of the general population is very low.

Analyses

The method of correlated vectors compares the vectors defined by the *g* loadings of a variety of tests and the standardized mean group differences in those tests (*ds*). It must observe several conditions (Jensen, 1998): (1) the samples must be representative of their respective populations and should be sufficiently large that the sampling error of the correlations among tests is small enough to yield unambiguous and reliable factors; (2) the samples should not have been selected on any variables related to cognitive abilities (for instance educational or occupational level) that might *significantly* restrict the variance with respect to performance on the battery of tests subjected to factor analysis; (3) the factor analysis should be based on large enough number of tests to permit the extraction of a stable and reliable *g* factor, as would be indicated by a very high coefficient of congruence between the *g* factor obtained in independent samples from the same population; (4) any test that is demonstrably biased with respect to the groups of interest should be excluded; (5) the tests must be sufficiently diverse in content, task demands, and factor structure to allow significant differences between the *g* loadings of the various tests; (6) the tests' reliability coefficients should be known so that each test's *g* loadings (and also the standardized mean group difference) can be corrected for attenuation (measurement error); (7) the factor analysis must be carried out separately within either sample or separately for both, but not in the combined samples, so that any psychometric differences between them cannot possibly enter into the factor analysis; and (8) the vector of *g* loadings extracted separately from each group must be sufficiently similar across groups to assure that the same factor is represented for both, as indicated by a congruence coefficient of above +0.95. The groups cannot be compared meaningfully on a nominal factor unless it is the same factor for both groups (McArdle, 1996).

The present study fits most of the prerequisite conditions to apply the method of correlated vectors: (1) the *g* factor is the same for both groups in both of the independent samples (see below); (2) the tests' reliability coefficients were known; (3) the tests are sufficiently diverse in content and task demands; (4) the tests are not biased; and (5) the samples are large enough. Although there is probably some restriction of range, the subjects are applicants not previously selected.

The congruence coefficient (r_c) is an index of factor similarity. A value of r_c of $+0.90$ is considered a high degree of factor similarity; a value greater than $+0.95$ is generally interpreted as practical identity of the factors (Jensen, 1998). The r_c is preferred over the Pearson r for comparing factors, because r_c estimated the correlation between the factors themselves, whereas the Pearson r gave only the correlation between the two column vectors of factor loadings. The n factor loadings of each of the n tests for each sample can be arranged as two-column vectors. The congruence coefficient is computed through the next formula: $r_c = \sum XY / \sqrt{\sum X^2 \sum Y^2}$.

The statistical test of the hypothesis concerning mean group differences is the correlation between the vector of the tests' *g* loadings and the vector of standardized mean differences between the groups on each of the tests (*ds*), taking the tests' reliability coefficients into account. The hypothesis and the method for testing it depends only on the relative magnitudes of the group difference across various tests that differ in their *g* loadings.

The Pearson r and Spearman's rank-order correlation, r_s, are suitable measures of the degree of relationship between the two vectors (it is most informative to report both). The test of significance of r_s is a stringent statistical test of the hypothesis, because the n (number of tests) is typically small. This does not mean that the number of subjects (N) is unimportant: the larger the N, the smaller will be the standard error of the *g* loadings and the larger will be the standardized mean differences, and hence, the more reliable will be their vectors. Given the large N in the present study, the *g* loadings can be seen as highly reliable (note the similarity between the *g* loadings of the same test in both samples; see Tables 5 and 6).

If the groups show virtual identity of the *g* factor, the reliability of the vector of *g* loadings can be increased by combining the two vectors. The two groups' factor loadings on a given test are averaged as follows: Average loading $= \sqrt{(a_1^2 + a_2^2)/2}$.

It seldom makes any difference whether *g* is represented by the highest-order factor in a hierarchical factor analysis or by the first unrotated principal factor in a principal factor analysis (these typically have a congruence coefficient of $+0.99$ or more) because the *g* factor is found to be remarkably invariant across all the various methods of factor analysis and relatively invariant across different batteries of diverse tests of mental ability (Jensen, 1998; Jensen & Weng, 1994).

Test unreliability must be considered because it has the effect of decreasing both the *g* loadings and the standardized mean differences. If the various tests' reliability coefficients differ significantly then, because they affect each test's factor loadings and standardized mean group difference to the same relative degree, a correlation between the vector of loadings and the vector of differences could be entirely an artifact of differences in the reliability of the various tests. On the other hand, if the vector of tests' reliability coefficients were negatively correlated with either the vector of factor loadings or the vector of mean group differences, this fact could obscure the possibility of a significant correlation between the vector of *g* loadings and the vector of mean group differences. There would be no problem if the tests all had the same reliability. The two methods for controlling for the effect of the tests' unequal reliability coefficients are partial correlation and correction for attenuation.

Besides the method of correlated vectors, the sex difference on each of the subtests of the batteries was also represented in terms of a point-biserial correlation

Table 3. Correlation Matrix for the First Sample

	1	*2*	*3*	*4*	*5*
1. Vocabulary (PMA)	0.91	0.186 (0.174)	0.230 (0.175)	0.297 (0.297)	0.223 (0.175)
2. Verbal Fluency (PMA)	*0.175*	0.73	0.153 (0.104)	0.169 (0.163)	0.265 (0.258)
3. Spatial Rotation (PMA)	*0.219*	*0.116*	0.73	0.389 (0.408)	0.274 (0.307)
4. Inductive Reasoning (PMA)	*0.292*	*0.170*	*0.379*	0.92	0.435 (0.450)
5. Monedas	*0.217*	*0.218*	*0.332*	*0.395*	0.94

Note: The correlations for the male and female subsamples are presented in the top half (female correlations are in parentheses). The correlations for the subsamples combined are presented in the bottom half. Reliability coefficients (r_{xx}) at the diagonal.

(adjusted for the inequality of *N* and *SD* values), including these correlations with the full matrix of subtest intercorrelations for factor analysis. The results of this analysis reveal the factor loading of sex on *g*. The *g* factor loading of sex is equivalent to the point-biserial correlation between *g* and the sex variable. This method is preferable to the use of *g* factor scores, because *g* factor scores are not a pure measure of the *g* factor of the test battery from which it was extracted. An individual's *g* factor score is calculated as a *g*-weighted mean of the individual's standardized scores on each of the subtests. Therefore, it is contaminated by other factors (and/or test specificity), either increasing or decreasing the mean sex difference, depending on the types of subtest in the battery (Jensen, 1998). However, there is an important caution note on these analyses: we are grateful to Lloyd Humphreys, who notes in his review that "the point-biserial for each subtest with sex is indeed determined jointly by the general factor and other nonerror factor content. When several subtests are added together, whether with unit weights or with weights used to estimate a score on the hypothetical general factor, *total* nonerror uniqueness shrinks in its contribution to variance."

Results

The descriptive data for the first and the second samples are shown in Tables 1 and 2. Also shown are the standardized sex differences (*d*) used for later studying the relationships with the average *g* loadings.

Tables 3 and 4 present the correlation matrix for the first and the second samples. Each correlation matrix shows the correlations separately for males and females (top half), as well as for the combined subsamples of males and females (bottom half).

A principal axis factoring was computed separately for females and males. The *g* factor was represented by the first principal unrotated factor. As Jensen (1998) wrote: "whatever variation exists among the myriad estimates of *g* that have been reported since the beginning of factor analysis, exceedingly little of it can be attributed to differences in the methods of factor analysis employed" (p. 83). The KMO statistic was between 0.691 and 0.815 depending on the subsample being analyzed, whereas the Bartlett's tests were all significant ($p = 0.000$).

Table 5 shows the *g* loadings for males, females, as well as the average loadings in the first battery. The congruence coefficient computed after the factor loadings in Table 5 for males and females is +0.978 (partial correlation controlling for $r_{xx} = 0.9962$, $p < 0.01$).

Table 4. Correlation Matrix for the Second Sample

	1	*2*	*3*	*4*	*5*	*6*
1. Vocabulary (PMA)	0.91	0.369 (0.444)	0.3 (0.287)	0.337 (0.399)	0.151 (0.196)	0.255 (0.270)
2. Verbal reasoning (DAT)	*0.406*	0.89	0.478 (0.454)	0.397 (0.398)	0.382 (0.376)	0.275 (0.284)
3. Monedas	*0.304*	*0.475*	0.94	0.471 (0.460)	0.425 (0.433)	0.3 (0.317)
4. Inductive Reasoning (PMA)	*0.353*	*0.383*	*0.402*	0.92	0.444 (0.467)	0.392 (0.418)
5. Dominoes	*0.174*	*0.382*	*0.417*	*0.446*	0.91	0.339 (0.338)
6. Spatial Rotation (PMA)	*0.271*	*0.291*	*0.340*	*0.380*	*0.341*	0.73

Note: The correlations for the male and female subsamples are presented in the top half (female correlations are in parentheses). The correlations for the subsamples combined are presented in the bottom half. Reliability coefficients (r_{xx}) at the diagonal.

Thus, the g factors are the same for males and females, so we can compute the average g loading to apply the method of correlated vectors.[1]

Table 6 shows the factor loadings for males and females, as well as the average loadings for the second battery of cognitive tests. The congruence coefficient (r_c) computed after the factor loadings in Table 6 for males and females is $r_c = +0.995$ (partial correlation controlling for $r_{xx} = 0.9611$, $p < 0.01$). Therefore, the g factors are the same for males and females.

Jensen (1998) has applied the method of correlated vectors for comparing the g loadings in some batteries of mental abilities (WISC-R, WAIS, GATB, ASVAB, and the British Ability Scales, BAS) and the standardized sex difference (d) for the scales included in those batteries. The Spearman rank-order correlation (r_s) of the column vector of subtests' g loadings with the vector of the sex differences (d) on the subtests indicates the degree to which g is related to the rank order of the sex differences on the various subtests. The mean correlation found by Jensen (1998) was $r_s = +0.116$, suggesting a negligible sex difference in g (p. 539).

The Spearman rank-order correlation computed after Table 5 is $r_s = +0.4$ ($p = 0.505$), a value suggesting a negligible sex difference in g (Pearson $r = +0.251$, $p = 0.684$; partial correlation controlling for $r_{xx} = +0.1733$, $p = 0.827$). The Spearman rank-order correlation (r_s) computed after Table 6 is $r_s = -0.143$ ($p = 0.787$) suggesting a negligible sex difference in g (Pearson $r = -0.088$, $p = 0.869$; partial correlation controlling for $r_{xx} = -0.0025$, $p = 0.997$).

However, using five or six tests in a correlated vector analysis is relatively weak and risks a Type II error (accepting the null hypothesis when it is false), because the statistical test is based on the number of variables, which is small here. To surpass this problem, the correlated vectors derived from the two independent samples were combined to yield a statistical test based on g and d vectors with an N of 11. This gives the null hypothesis a fair chance of rejection. The Pearson r between the vectors is +0.122 ($p < 0.721$). With the vector of reliability coefficients partialled out, the g and d vectors are correlated +0.051. The Spearman rank-order correlation is 0.000 ($p > 0.999$). This result is a definitive failure to reject the null hypothesis.[2]

Table 5. Males *g* Loadings, Females *g* Loadings, and Average *g* Loadings (First Battery)

Tests in the first battery	*Males (g loadings)*	*Females (g loadings)*	*Average g loadings*
Vocabulary (PMA)	0.434	0.363	0.4
Verbal Fluency (PMA)	0.336	0.3	0.318
Spatial Rotation (PMA)	0.509	0.512	0.51
Inductive Reasoning (PMA)	0.698	0.748	0.723
Monedas	0.599	0.607	0.6
Eigenvalue	1.406	1.412	
Percent Variance	28.115	28.238	
N	4,136	2,743	6,879

Table 6. Male *g* Loadings, Female *g* Loadings, and Average *g* Loadings (Second Battery)

Tests in the second battery	*Males (g loadings)*	*Females (g loadings)*	*Average g loadings*
Vocabulary (PMA)	0.457	0.515	0.486
Verbal Reasoning (DAT)	0.641	0.644	0.642
Monedas	0.688	0.655	0.671
Inductive Reasoning (PMA)	0.706	0.722	0.714
Dominoes	0.592	0.599	0.595
Spatial Rotation (PMA)	0.506	0.523	0.514
Eigenvalue	2.198	2.262	
Percent Variance	36.635	37.703	
N	2,083	1,513	3,596

The point-biserial correlations between sex and the subtests in each sample were also computed. For the first sample, these correlations were: Vocabulary = 0.07, Verbal Fluency = −0.07, Spatial Rotation = 0.20, Inductive Reasoning = −0.05, Monedas = 0.36. For the second sample, the correlations were: Vocabulary = 0.08, Verbal Reasoning = 0.11, Monedas = 0.35, Inductive Reasoning = −0.09, Dominoes = 0.04, Spatial Rotation = 0.16. These correlations were included with the full matrix of subtest correlations for factor analysis. The *g* loadings of sex were 0.244 and 0.188, respectively. Both loadings are smaller than the highest one reported by Jensen (1998), that is, −0.255.

Finally, it was computed a single *g* factor score. This score was then correlated with sex. For the first sample, the Pearson *r* was 0.166 ($p < 0.000$), Rho was 0.167 ($p < 0.000$), and Tau was 0.136 ($p < 0.000$). For the second sample, the Pearson *r* was 0.154 ($p < 0.000$), Rho was 0.153 ($p < 0.000$), and Tau was 0.125 ($p < 0.000$). These significant correlations could be explained by the non-*g* variance included in the *g* factor scores. It should be remembered that the *g* factor scores are not a pure measure of *g*.

Discussion

The negligible sex difference in *g* reported in the present study comes from the largest sample on which a sex difference in *g* has ever been tested (Jensen, personal communication). Furthermore, the findings are entirely consistent with those using quite different batteries and subject samples. Therefore, what we have here is a replication of a consequential phenomenon, both because the question about differences between the sexes in *g* is technically the most difficult to answer within this field and because it has been the least investigated.

As was put forth by Jensen (1998), testing a hypothesis by means of correlated vectors is an extremely severe test, because in the existing studies, the number of variables in each vector is typically about 12 and rarely more than 20 (p. 416). The results in the present study does not reject the null hypothesis and hence the conclusion that there are no sex differences than can be attributed to general intelligence defined as *g* is ready to be made. Even more important, beyond the method of correlated vectors, the findings derived from other statistical analyses are consistent (except for the *g* factor scores; but remember that these scores are not a fine-grained representation of *g*).

Lynn (1994) has noted that the 4 IQ points male advantage he had obtained is the advantage that can be predicted from their larger brains. This prediction is based on a correlation of 0.35 between in vivo brain size and WAIS IQ and a sex difference of 0.78 *SD* in adult brain size. Three issues on the Lynn's study are worth noting: (1) the sex differences in intelligence vary at different ages so that aggregating them for all ages and concluding that no differences exist is not a satisfactory approach; (2) sex differences fluctuate in tandem with sex differences in brain size; and (3) sex differences are about 1.5 IQ points up to the age of 14 to 15 years and thereafter increase steadily to approximately 4 IQ points among adults.

However, considering the data of Lynn (1994), Ankey (1995) has calculated that 88 percent of the 4 IQ points favoring males is derived from their higher performance on reasoning and spatial tests. Almost all the differences are due to the male advantage in spatial and math reasoning ability. This suggests that these tasks must require more brain tissue than do those tasks at which women have higher scores (verbal ability and perceptual speed). Moreover, the sex difference in brain size could be explained considering the greater packing density of neurons in the female brain. This sexual dimorphism allows the same number of neurons in the male and female brains despite their differences in gross size. Variation in total brain size accounts for only a minor part the total variance in *g* or in IQ. Thus, other physiological factors unrelated to brain size must also contribute a large part of the *g* variance (Jensen, 1998).

On the other hand, the mean *d* that can be calculated after Tables 1 and 2 of the present study are $d = 0.2$ and 0.23, respectively. These values translate into 3 and 3.4 average IQ points favoring males, not so far from the 4 IQ points calculated by Lynn (1994). The important point is that the method of correlated vectors contradicts the conclusion that could be derived from a direct calculation made after the standardized mean group differences (*d*). If the method of correlated vectors is favored to test the null hypothesis concerning sex differences in general intelligence defined as *g*, then we can conclude that there is no sex difference in general intelligence. This conclusion is congruent with the findings reported by Jensen (1998, Chap. 13) as well as with textbooks considering the topic (Brody, 1992; Colom, 1998; Halpern, 1992). The conclusion also disagrees with the procedures of Lynn (1994) or Rushton (1992) to resolve the "paradox" concerning sex differences in intelligence and brain size.

The theoretical importance of a negligible sex difference in *g*, as the one we have found in the present study with independent samples totaling 10,475 adult subjects (see also Aluja-Fabregat, Colom, Abad, & Juan-Espinosa, in press, for a study with 1,565 young adolescents) suggests that: (1) the factor (*g*) that is present in nearly all measures of cognitive ability (and that accounts for more than half of the total common factor variance in a test), does not differ across sex; (2) non-*g* factors (and/or test specificity) are responsible for the observed cognitive sex differences; and (3) with respect to the "paradox" (the findings of larger male brain, the association of brain size with IQ, and the absence of a sex difference in overall IQ), the sex differences could reside "in the modular aspects of brain functioning rather than in whatever conditions of the brain's information processing capacity cause positive correlations among all of the modular functions on which there is normal variation and which accounts for the existence of *g*" (Jensen, 1998, pp. 541–542).

Notes

1. Lloyd Humphreys notes in his review that "qualitatively equality can still produce a difference in means of scores estimating the general factor."
2. We thank Arthur Jensen for suggesting this combination of vectors.

References

Aluja-Fabregat, A., Colom, R., Abad, F. J., & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as *g* among young adolescents. *Personality and Individual Differences*, *28*(4), 813–820.

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

Ankey, C. (1995). Sex differences in brain size and mental abilities: Comments on R. Lynn and D. Kimura. *Personality and Individual Differences, 18*(3), 423–424.

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

Carretta, T. R., & Ree, M. J. (1995). Near identity of cognitive structures in sex and ethnic groups. *Personality and Individual Differences, 19*, 149–155.

Carretta, T. R., & Ree, M. J. (1997). Negligible sex differences in the relation of cognitive and psychomotor abilities. *Personality and Individual Differences, 22*(2), 165–172.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor analytic studies*. Cambridge: Cambridge Univ. Press.

Carroll, J. B. (1997). Psychometrics, intelligence, and public perception. *Intelligence, 24*(1), 25–52.

Colom, R. (1998). *Psicologá de las diferencias individuales. Teorá y práctica* [Psychology of individual differences. Theory and practice]. Madrid: Pirámide.

DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. C., & Johnson, R. C. (1974). Near identity of cognitive structure in two ethnic groups. *Science, 183*, 338–339.

DeLacoste, M., Adesanya, T., & Woodward, D. (1990). Measures of gender differences in the human brain and their relationship to brain weight. *Biological Psychiatry, 28*, 931–942.

Halpern, D. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.

Humphreys, L. G., & Taber, T. (1973). Ability factors as a function of advantaged and disadvantaged groups. *Journal of Educational Measurement, 10*, 107–115.

Hyde, J., & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53–69.

Jensen, A. (1992). Vehicles of g. *Psychological Science, 3*, 275–277.

Jensen, A. (1998). *The g factor*. London: Praeger.

Jensen, A., & Weng, L. (1994). What is a good *g*? *Intelligence, 18*, 231–258.

Juan-Espinosa, M. (1997). *Geografá de la inteligencia humana. Las aptitudes cognitivas* [Geography of human intelligence. Cognitive abilities]. Madrid: Pirámide.

Linn, M., & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. San Francisco, CA: Freeman and Co.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences, 17*(2), 257–271.

McArdle, J. J. (1996). Current direction in structural factor analysis. *Current Directions in Psychological Science, 5*, 11–18.

McGlone, J. (1980). Sex differences in human brain assymetry: A critical survey. *Behavioral Brain Science, 3*, 215–263.

Michael, W. B. (1949). Factor analysis of tests and criteria: A comparative study of two AAF pilot populations. *Psychological Monographs, 63*, 55–84.

Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement, 55*, 268–277.

Ree, M. J., & Earles, J. A. (1991). The stability of convergent estimates of *g*. *Intelligence, 15*, 271–278.

Rushton, J. P. (1992). Cranial capacity related to sex, rank and race in a stratified sample of 6,325 U.S. military personnel. *Intelligence, 16*, 401–414.

Rushton, J. P., & Ankey, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin and Review, 3*(1), 21–36.

Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250–270.

Willerman, L., Schultz, R., Rutledge, J., & Bigler, E. (1991). In vivo brain size and intelligence. *Intelligence, 15*, 223–228.