Person. individ. Diff. Vol. 4, No. 2, pp. 223–226, 1983
Printed in Great Britain.

0191-8869/83/020223-04$03.00/0

Sex differences on the WISC-R

ARTHUR R. JENSEN

Institute of Human Learning, University of California, Berkeley, CA 94709, U.S.A.

and

CECIL R. REYNOLDS

Department of Educational Psychology, 704 Harrington Education Center, Texas A & M University, College Station, TX 77843, U.S.A.

(*Received 9 July 1982*)

Sex differences on 13 subtests of the Wechsler Intelligence Scale for Children—Revised (WISC-R), and in the factor structure of the 13 subtests, are examined in the white subsample ($N = 944$ males, 924 females) of the total national standardization sample of the WISC-R (Kaufman and Doppelt, 1976; Wechsler, 1974). This sample of children, whose ages range from 6 to 16.5 years, were chosen by a stratified, random-sampling procedure to be representative of the population of the United States, in accord with demographic features revealed in the 1970 census.

All of the WISC-R subtests, listed in Table 1, are highly familiar, except Tapping Span, which was not included in the published version of the test. It is ostensibly a non-verbal test of visual imitative memory. A straight row of four 1″ wooden blocks, spaced 1″ apart, is placed before the *S*. The examiner, holding a block between his thumb and index finger taps out a pattern on the row of four blocks, say, 1-4-2-3, if we imagine the blocks are consecutively numbered from left to right. The *S*'s task is to imitate the same pattern of taps on the row of blocks, tapping out the pattern just as the examiner had done. Task difficulty is increased by tapping out longer and more complex series.

The use of subtest scaled scores, with a mean of 10 and SD of 3 at every age in the total standardization sample, in effect obviates age differences in the test scores. All the analyses in this study are based on the age-standardized scaled scores.

Table 1 shows the means and SDs of the scaled scores for males and females, and the difference between the sexes, in scaled score units and in SD units. Despite a multivariate statistical test which shows the overall differences between the sexes to be significant beyond the 0.001 level, the differences are generally quite small, with the marked exception of the Coding test, on which females exceed males by about half of a SD. This one quite extreme sex difference in the WISC-R battery has also been noted by Samuel (1983) in large samples both of whites and blacks (independent of the present standardization sample). The Coding test (matching geometric symbols with numbers) seems to involve perceptual speed and accuracy, dexterity in copying symbols, and rote memory. The only other subtest which shows an appreciable difference (0.37 SD), in favor of males, is Information.

The sex differences can also be expressed in the form of the point-biserial correlation coefficient, r_{pb}, with males coded 0, females 1. The r_{pb}, in the range of low values found here, is virtually a linear function of the mean difference. Full-Scale IQ (FSIQ) can be partialed out of each r_{pb}, revealing the relative size of the sex differences on the various subtests when the sexes are statistically equated on FSIQ. These results are shown in Fig. 1. Partialing out FSIQ has only a slight

Table 1. Mean and SD of scaled score ($\mu = 10$, $\sigma = 3$) of WISC-R subtests and Verbal, Performance and FSIQ ($\mu = 100$, $\sigma = 15$) for males ($N = 944$) and females ($N = 924$)

	Male		Female		Difference*	
WISC-R Scale	Mean	SD	Mean	SD	$\bar{M} - \bar{F}$	Diff./$\bar{s}$
Information	10.94	3.01	9.87	2.71	1.07	0.37
Similarities	10.39	3.13	10.19	2.89	0.20	0.07
Arithmetic	10.46	2.99	10.28	2.69	0.18	0.06
Vocabulary	10.62	2.99	10.22	2.87	0.40	0.14
Comprehension	10.57	2.77	10.30	2.83	0.27	0.09
Digit Span	9.93	3.07	10.23	2.92	−0.30	−0.10
Tapping Span	10.03	2.88	10.15	2.86	−0.13	−0.04
Picture Completion	10.62	2.93	10.20	2.79	0.42	0.15
Picture Arrangement	10.52	2.98	10.21	2.84	0.31	0.11
Block Design	10.61	2.92	10.16	2.90	0.45	0.15
Object Assembly	10.63	2.99	10.09	3.02	0.54	0.18
Coding	9.44	3.00	11.01	2.86	−1.56	−0.53
Mazing	10.69	3.01	10.12	3.09	0.57	0.19
Verbal IQ	103.33	14.70	100.69	13.54	2.64	0.19
Performance IQ	102.28	14.33	102.07	13.94	0.21	0.01
FSIQ	103.08	14.54	101.41	13.55	1.68	0.12

*$\bar{M} - \bar{F}$ is the difference in mean scaled scores of males and females; Diff./$\bar{s}$ is the mean difference divided by the weighted average SD of the male and female samples, i.e.

$$\bar{s} = (N_M s_M^2 + N_F s_F^2)/(N_M + N_F),$$

where M and F stand for male and female, N is sample size, and s is the SD.

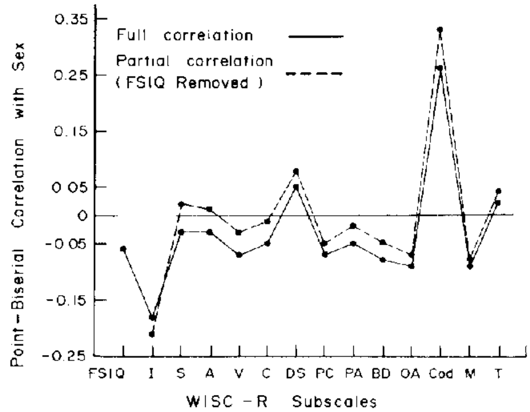

Fig. 1. Point-biserial correlation as an index of female–male difference on FSIQ and on each of 13 subtests of the WISC-R. (Because males and females were quantitized 0 and 1, respectively, a positive correlation indicates female superiority.) The solid-line profile reflects the actual group differences. The dashed-line profile reflects the female-male differences on the 13 subtests after FSIQ has been partialed out, in effect equating the sexes on general intelligence. With 1866 *df*, correlations falling outside the range of ± 0.04 differ significantly from zero beyond the 5% level of confidence. (I, Information; C, Comprehension; A, Arithmetic; S, Similarities; DS, Digit Span; V, Vocabulary; Cd, Coding [Digit Symbol]; PC, Picture Completion; BD, Block Design; PA, Picture Arrangement; OA, Object Assembly; M, Mazes; T, Tapping [Knox Cubes].)

effect on the r_{pb} profile of sex differences. When FSIQ is partialed out, the largest sex differences favoring males are on Information, Block Design, Picture Completion, Object Assembly and Mazes. The largest differences favoring females are on Coding, Digit Span and Tapping Span. The tests showing the smallest (and non-significant) differences, after FSIQ is partialed out, are Similarities, Arithmetic, Vocabulary, Comprehension and Picture Arrangement.

In a review of the evidence on a sex difference in the variability of mental test scores, Jensen (1980, pp. 627–628) concluded that the most reliable data indicate greater variability for males, equivalent to about a 1 IQ point larger standard deviation for males than for females. The present IQ data are all entirely consistent with this generalization, for Verbal, Performance, and FSIQ. The variance ratios (male variance/female variance) are significant beyond the 0.05 level for the Verbal and FSIQs.

Table 2 shows the intercorrelations among subtests for males and females, and the general factor (first principal factor) of the correlation matrix for each sex. The pattern of correlations is highly similar for males and females. A sensitive statistical test (Jennrich, 1970) of the equality of the two correlation matrices shows they do not differ significantly (χ^2 78.25, $df = 78$, $P = 0.47$). This means also that the male and female correlation matrices cannot differ significantly in factor structure or in their factor loadings. Hence it is not surprising that the coefficient of congruence between the first principal factors of the two matrices is 0.997.

In a previous study (Jensen and Reynolds, 1982), a hierarchical factor analysis, with Schmid–Leiman (1957) orthogonalization, was performed on the white standardization sample (both sexes combined). From this analysis, factor scores, scaled to a mean of 0 and SD of 1 for the entire (white and black) standardization sample, were obtained for every individual on each of the four orthogonalized factors extracted from the 13 WISC-R subtests: the general factor (g), a verbal factor (V), a non-verbal performance factor (P), and a memory factor (M). (The details of this factor analysis are presented in Jensen and Reynolds, 1982). Table 3 shows the means and SDs of the factor scores for each

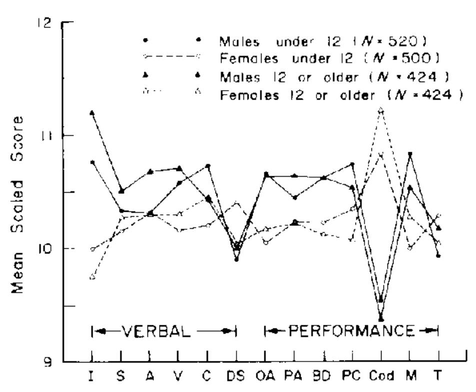

Fig. 2. Mean scaled scores ($\mu = 10$, $\sigma = 3$) on WISC-R subtests for males and females in age groups (1) under 12 years and (2) 12 years and older.

Table 2. Intercorrelations (decimals omitted) among WISC-R subtests for males (above diagonal) and females (below diagonal) and first principal factor loadings

Subtest	I	S	A	V	C	DS	TS	PC	PA	BD	OA	Cd	M	FPF
Information		60	54	67	51	37	28	33	39	46	30	34	21	75
Similarities	57		44	62	56	31	22	41	40	48	31	27	25	73
Arithmetic	49	40		51	41	45	35	30	28	44	21	33	26	65
Vocabulary	64	63	43		62	38	27	39	42	46	30	35	23	79
Comprehension	51	53	40	61		23	21	35	37	38	26	27	25	65
Digit Span	34	35	39	34	24		38	14	16	29	14	28	18	48
Tapping Span	24	17	29	21	19	35		17	20	30	16	27	22	42
Picture Completion	35	39	30	36	34	20	16		37	47	37	17	29	54
Picture Arrangement	34	33	24	35	30	21	20	30		44	33	27	25	55
Block Design	40	42	38	39	37	29	23	47	38		55	36	36	73
Object Assembly	37	40	25	34	32	22	14	45	40	56		24	21	49
Coding	30	28	28	29	24	26	23	17	23	30	25		19	46
Mazes	20	22	22	19	21	20	18	29	28	40	36	23		40
First principal factor	72	72	59	75	65	49	36	55	51	68	61	43	42	

sex, and the sex differences. Males are slightly, but significantly ($P < 0.01$), higher on *g*, V and P, while females are rather markedly higher on M. A multiple regression analysis, with the factor scores as the independent variables and sex as the dependent variable, yielded a multiple R of 0.171, $P < 0.001$, indicating that the pattern of factor scores on the WISC-R is significantly different for males and females. Factor scores on the *g* factor, which accounts for the largest percentage (29.7%) of the total WISC-R variance, are correlated only -0.08 with sex (favoring males). Although statistically significant ($p < 0.01$) because of 1867 *df*, this size of correlation is practically negligible. It means that sex accounts for only about 0.6% of the variance in *g* factor scores. This small difference, equivalent to about 0.16 SD, could reasonably be viewed as a psychometric artifact due to the particular selection of subtests in the WISC-R, or it could reflect a true, albeit practically negligible, difference in Spearman's *g*. The present data cannot resolve this uncertainty, which could be reduced only by looking at sex differences in the *g* factors derived from a sizeable number of different collections of diverse cognitive ability tests, especially tests and test batteries which were not constructed with the aim of minimizing or balancing out sex differences by means of item selection.

The females' superiority (by 0.26 SD) on the memory factor, and particularly on the Coding test, is consistent with other studies (Samuel, 1983).

The fact that the smallest sex difference is on the performance factor, however, may seem surprising in view of the considerably larger male–female difference on the performance tests in the adult standardization sample of the WAIS, especially on the Block Design and Object Assembly tests, which involve spatial visualization ability. Also, in the adult sample, females exceed males on the verbal factor, which is opposite to the differences found in the present sample (Jensen, 1980, pp. 624–625).

Samuel (1983) has reviewed evidence that the sex difference in spatial ability (which, aside from *g*, is a large part of the WISC-R performance factor) becomes accentuated after puberty, under the influence of hormonal changes. In samples of *S*s between the ages of 12 and 16, Samuel found significant sex differences (favoring males) on the more spatially-loaded performance tests of the WISC (Object Assembly, Picture Arrangement, Block Design, and Picture Completion) which average about 0.25 scale-score units larger than the corresponding sex differences found in the present sample, of which slightly more than half of the *S*s are below age 12.

To determine whether sex differences on the spatially-loaded performance subtests may be larger for adolescents than for prepubertal children, as suggested by Samuel, we have divided the WISC-R standardization sample into two age groups: (1) children under 12 years and (2) children of 12 years and older. The mean ages of the younger and older groups are 8.51 (SD = 1.71) and 14.00 (SD = 1.42), respectively. The profiles of all of the WISC-R subtest scaled score ($\mu = 10$, $\sigma = 3$) means for males and females within each age group are shown in Fig. 2. The results for the Performance subtests may be compared directly with the results in Samuel's (1983) Fig. 1. As in Samuel's study, there is a small sex difference favoring males on all of the Performance subtests except Coding, on which the sex difference—by far the largest in both studies—favors females. But the pattern of sex differences on the Performance subtests is virtually the same in both the younger and older groups. On none of the subtests is there a significant interaction between the variables of sex and age. The observed sex difference in these data is not just a postpubertal phenomenon. It is also

Table 3. Mean and SD of factor scores ($\mu = 0$, $\sigma = 1$), based on Schmid–Leiman hierarchical factor analysis in the WISC-R standardization sample, for males ($N = 944$) and females ($N = 924$)

	Males		Females		Difference*	
Factor	Mean	SD	Mean	SD	M − F	Diff./*s*
g	0.202	0.873	0.066	0.819	0.136	0.161
Verbal	0.089	0.596	−0.015	0.592	0.104	0.175
Performance	0.104	0.641	0.011	0.652	0.093	0.144
Memory	−0.086	0.674	0.080	0.622	0.166	0.256

*See footnote to Table 1.

Table 4. Sex and age contrasts on WISC-R subtests significant beyond 0.05 level*

		Verbal						Performance							IQ	
	Contrast	I	S	A	V	C	DS	OA	PA	BD	PC	Cd	M	T	V	P
(A)	Under 12: Male vs Female	M				M	F	M				F	M		M	
(B)	12 or older: Male vs Female	M									M	F			M	
	Interaction (A × B)															
	Males 12 or older vs under 12															
	Females: 12 or older vs under 12															

*Significant contrasts indicated by M (males superior) and F (females superior). Non-significant contrasts indicated by single dash.

evident in Fig. 2 that the sex difference is not confined to the performance tests, but also appears on the verbal tests as well. With the exception of Coding, however, the sex differences are all quite small, averaging less than one-sixth of a standard deviation, with little variation across the subtests.

All of the possible sex and age contrasts (and their interactions) depicted in Fig. 2 were tested for statistical significance by the Dunn–Bonferroni (1961) method for contrasting means, setting the overall level of confidence for rejecting the null hypothesis at 0.05 for all of the contrasts on each subtest. (It should be noted that these statistical tests of the contrasts are not statistically independent from one subtest to another, because the subtests are intercorrelated.) The results of this analysis are fully displayed in Table 4. It can be seen that although there are significant sex differences on several subtests, there are no significant interactions between sex and age group in the WISC-R white standardization sample. Moreover, one of the hypotheses mentioned by Samuel to explain the females' superiority on the Coding subtest, viz., that the Coding task

> "encourages verbal associations to be formed between symbols and numbers and so allows the more highly developed verbal skills of females to be brought to bear on the task,"

appears inconsistent with the finding in the present data that, although females perform markedly better than males on Coding, females obtain slightly but significantly ($P < 0.05$) *lower* Verbal IQs than males. (There is no significant sex difference on Performance IQ.) Female superiority on the Coding test is probably better explained by this test's substantial loading on the memory factor, which shows by far the largest sex difference (in favor of females) of any of the factors in the WISC-R.

REFERENCES

Dunn, O. J. (1961) Multiple comparisons among means. *J. Am. statist. Ass.* **56,** 52–64.

Jennrich R. I. (1970) An asymptotic χ^2 test for the equality of two correlation matrices. *J. Am. statist. Ass.* **65,** 904–912.

Jensen A. R. (1980) *Bias in Mental Testing.* The Free Press, New York.

Jensen A. R. and Reynolds C. R. (1982) Race, social class and ability patterns on the WISC-R. *Person. individ. Diff.* **3,** 423–438.

Kaufman A. S. and Doppelt J. (1976) Analysis of WISC-R standardization variables in terms of the stratification variables. *Child Dev.* **47,** 165–171.

Samuel W. (1983) Sex differences in spatial ability reflected in performance on IQ subtests by black or white examinees. *Person. individ. Diff.* **4,** 219–221.

Schmid J. and Leiman J. M. (1957) The development of hierarchical factor solutions. *Psychometrika* **22,** 53–61.

Wechsler D. (1974) *Wechsler Intelligence Scale for Children–Revised.* Psychological Corporation, New York.