FACTORIAL SIMILARITY AND ACCURACY OF MEASUREMENT OF THE SAUDI VERSION OF THE WISC-R ACROSS SEX AT SIX AGE GROUPS

Abdullah Qataee
Department of Psychology
King Saud University

The Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) has become "one of the most frequently used instruments in the individual assessment of school-aged children" (Gutkin & Reynolds, 1981, 227). It has also gained acceptance in other countries such as Egypt (Ismael & Melaika 1983), Jordan (Olian & Kailani, 1988), and Kuwait (Abu Allam, 1989). Because of these reasons, the WISC-R was adapted to the society of Saudi Arabia.

The initial validation studies support its construct and criterion-related validity (Note1). Acknowledging Cronbach's contention that construct validation is a never ending process (Cronbach, 1971), the Saudi version of the WISC-R should go through a thorough investigation of its construct validity.

In the United States, there is a growing concern about test bias and its implication on testing. In fact, "there is little doubt that the so-called IQ Controversy has generated the lion's share of public scrutiny over the years" (Reynolds & Kaiser, 1990,: 487). More specifically, one among the various criticisms of intelligence tests is that "the tests measure different attributes when used with children outside of the mainstream" (Reynolds, 1983, P.243).

Accordingly, several investigators have studied structural invariance of the WISC-R for children from different ethnic backgrounds (Reschly, 1978; Vance & Wallbrown, 1978; Oakland & Feigenbaum, 1979; Gutkin & Reynolds, 1981; Reynolds & Gutkin, 1980).

Reschly (1978) investigated factorial similarity of the WISC-R for whites, blacks, Mexican-Americans, and Native American Papagos. Reschly's findings showed substantial congruence of the two-factor solutions

across groups. Reschly (1978) concluded that the interpretation of the WISC-R for whites, blacks, Mexican-Americans, and Native American Papagos as a measure of general ability is essentially the same.

Oakland and Feigenbaum (1979) investigated several sources of test bias on the WISC-R and Bender-Gestall tests. First factor loading were correlated for whites, blacks and Mexican American. Oakland and Feigenbaum reported very high correlations for the first factor loadings. The correlations range between.95-.97 for race and.94 for Socioeconimic status (SES) (1979: 973). They concluded that the WISC-R did not exhibit any sign of bias with regard to its construct (Oakland & Feigenbaum, 1979,: 974).

Further support for the factorial similarity of the WISC-R for different ethnic groups was reported by Gutkin and Reynolds (1981). Using the data gathered from the normative sample of the WISC-R Gutkin and Reynolds investigated factorial similarity of the WISC-R. Their findings showed that the congruence coefficients for "the g-factor, the two-factor, and the three-factor solutions across racial groups ranged only from.98 to.99" (Gutkin & Reynolds, 1981:229). They concluded that the WISC-R factor structure for blacks and whites was similar, thus calling for uniform interpretation of test scores for all children regardless of their ethnic background (1981).

Silverstein (1977) investigated the stability of the eleven age groups of the WISC-R standardization sample. For the two-factor solution, the reported median coefficient of congruence between matching factors was.93. However, the median coefficient for the three-factor solution was.87 for the matching factors. Silverstein 1977: 121) concluded that "the two-factor solution proved somewhat more stable".

Reynolds and Gutkin (1980) investigated factor similarity across sex at two age levels of the normative sample of the WISC-R. The results showed that the coefficient of congruence for the three-factor solution ranged from.97 to.99. Reynolds and Gutkin then concluded that the WISC-R factor structure was invariant across several demographic variables. In fact "the apparent stability of the three-factor solution across sex and age suports the uniform interpretation of the WISC-R factor score independent of the child's sex and age" (1980:777).

The WISC-R was only recently adapted to the Saudi society. Its psychometric properties have not been investigated. Furthermore, there is a systematic division of boys and girls in social settings and in education in Saudi Arabia. In addition, Jensen pointed out that "Sex bias in mental test

arises from consideration of the fact that boys and girls, from early age, are socialized in traditional masculine and feminine roles, involving different types of toys, games,...and developed skills. Varieties of test items are certainly not all independent of these experiential sex differences, and any given test could consists of a biased sample of the total possible pool of masculine and feminie-slanted items" (Jensen, 1980: 621). Therefore, the purpose of the present study was to determine, first factorial similarity of the Saudi version of the WISC-R for boys and girls across various age levels; second the constancy of errors due to domain sampling in the choice of items for males and females.

METHOD

Sample:

The standardization sample of the Saudi version of the WISC-R furnished the subjects for the study. The total sample was 1100 boys and girls drawn at random from Riyadh Province. 50 boys and 50 girls were included in each of the eleven age groups (6-16 yrs). A full description of the sample can be found elsewhere (Note 2).

For the purpose of the study, male and female subjects were divided into five age groups (6,7-8, 9-10, 11-12, 13-14, and 15-16), with 50 males and 50 females in the first age group and 100 males and 100 females in each of the remaining groups.

Dividing males and females into six age groups allowed the investigation of possible bias due to sex by age interaction effects on factorial stability (Reynolds & Gutkin, 1980).

Instrument:

Due to cultural differences, the Saudi version of the WISC-R has eliminated some items from the original test, and included some new items. The criteria for deletion were cultural inadequacy; extreem difficulty; and poor discrimination.

The test manual was written in a simple arabic language. However, instruction was similar to that of the original test.

Test administeration is similar to the original test administration. That is the test begins with a verbal test followed by a performance test and so on. However. The testing time ranged from 1.5 to 2 hours (Note 1).

Reliability: Reported reliability indices were test-retest, split-half and coefficient alpha. The indices were similar in magnitude to the original test

reliablities. In fact, test-retest reliabilities were.96.,95., 96 for verbal-, performance-, and total-scale, respectively. Split-half reliabilities for the same scales were.93, 84, and.95 respectively; and alpha coefficients were.98.,88.,.97, for the same tests, respectively (Note 2).

Validation: test manual included different types of validation procedures and all were supportive of its validity (Note 2).

Analysis:

A principal-factor analysis of the corelation matrix with R^2 as the initial communality estimate was performed, with a varimax rotation, for males and females separately in each age group. Deriving factors from the correlation matrix rather than the covariance matrix would not have a major impact on the accuracy of comparison (Reynolds and Harding, 1983).

Herman (1976) pointed out that the coefficient of congruence is the most appropriate test for factor similarity. However, Reynolds and Harding (1983) compared six methods usually used to assess factorial similarity. These methods were a coefficient of congruence drawn from the correlation matrix,and another drawn from the covariance matrix, the Pearson correlation between factor loadings, the Pearson correlation between Fisher-transformed factor loadings, the salient variable similarity index, and the Pearson correlation between factor scores. Their findings showed a great deal of similarity in the outcome of these methods. Accordingly Reynolds and Harding (1983: 728) concluded that "the differences between these indexes is not nearly so great as their similarity". They recommended, as a conservative approach, the use of a congruence coefficient drawn from a correlation matrix and the salient variable similarity index.

In the light of the aforementioned findings, the coefficient of congruence and the pearson correlation of the factor loading scores were used to study factor similarity for males and females across age groups.

In addition, Feldt's (1969) index (W) was used to compare reliability coefficients of males and females on the twelve subtests. The degrees of freedom were n-1 for both the numerator and denominator due to the large sample size ('Feldt, 1969).

RESULTS AND DISCUSSION

Using both eigenvalue (of one) - to determine the number of factors to be extracted-and scree plot criteria for factor rotation, a two-factor solution emerged for males and females across age groups, as indicated in tables 1 and 2.

Table (1)

Factor score for males (M) and

females (F) across six age groups (verbal factor)

	Age groups											
	6		7-8		9-10		11-12		13-14		15-16	
	N=100		N=200		N=200		N=200		N=200		N=200	
	M	F	M	F	M	F	M	F	M	F	M	F
INF*	.75	.66	.70	.34	.69	.79	.76	.64	.61	.70	.77	.72
SIM	.77	.09	.55	.55	.60	.42	.56	.49	.46	.62	.66	.65
AR	.59	.45	.65	.29	.53	.61	.52	.33	.69	.68	.72	.60
VOCAB	.78	.80	.83	.87	.88	.79	.80	.83	.57	.76	.76	.78
COM	.69	.82	.83	.89	.80	.75	.71	.75	.53	.72	.78	.86
DS	.13	.26	.42	.34	.23	.66	.73	.56	.78	.72	.61	.48
PC	.62	.68	.49	.30	.26	.28	.36	.28	.37	.55	.19	.28
PA	.67	.54	.43	.35	.18	.21	.29	.30	.38	.51	.27	.21
BD	.17	-.05	.28	.23	.15	.20	.29	.21	.35	.45	.26	.08
OA	.09	.23	.21	.19	.15	.01	.07	.15	-.18	-.05	.12	.27
COD	.15	-.09	-.16	-.04	.14	.31	.51	.68	.54	.39	.15	.07
MAZ	.25	.17	.42	.33	.09	.23	.16	.15	.27	.32	.32	.24

* INF = Information; SIM = Similarity; AR = Arithmetic

VOAB = Vocabulary; COM = Comprehension; DS = Digit Span;

PC = Picture Completion; PA = Picture Arrangement;

BD = Block Design; OA = Object Assembly;

COD = Coding; MAZ = Mases.

Table (2)

Factor score for males (M) and

Females (F) across six age groups (performance factor)

	Age groups											
	6 N=100		7-8 N=200		9-10 N=200		11-12 N=200		13-14 N=200		15-16 N=200	
	M	F	M	F	M	F	M	F	M	F	M	F
INF*	.26	.41	.30	.69	.19	.27	.36	.46	.49	.43	.24	.43
SIM	.21	.01	.26	.31	.33	.51	.41	.45	.52	.39	.39	.41
AR	-.01	.26	.38	.69	-.003	.29	.55	.63	.35	.29	.26	.33
VOCAB	.16	.15	.12	.26	.06	.21	.21	.25	.58	.33	.22	.20
COM	-.24	.19	.01	.13	.12	.24	.34	.31	.48	.17	.08	.08
DS	.03	.71	.45	.61	-.28	.08	.19	.36	-.06	-.05	.26	.25
PC	.11	.30	.45	.65	.27	.67	.58	.57	.57	.45	.70	.65
PA	.35	.46	.65	.67	.41	.56	.61	.61	.52	.36	.60	.66
BD	.87	.87	.65	.68	.78	.72	.69	.73	.63	.63	.74	.79
OA	.74	.59	.58	.59	.76	.70	.82	.77	.82	.89	.76	.68
COD	.15	-.01	.66	.54	.30	.51	.08	.05	.08	.22	.44	.08
MAZ	.26	.57	.56	.62	.45	.67	.62	.67	.62	.60	.49	.52

* INF = Information; SIM = Similarity; AR = Arithmetic
VOB = Vocabulary; COM = Comprehension; DS = Digit Span;
PC = Picture Completion; PA = Picture Arrangement;
BD = Block Design; OA = Object Assembly;
COD = Coding; MAZ = Mases.

As can be seen from tables 1 and 2, the two-factor solution resembles Wechsler's (1974) grouping of the WISC-R into Verbal and Performance scales. Although the three-factor solution emerged in a validation study of the Saudi version (Note 1), it did not show up with any regularity in the present study for all age groups. Similar results were reported by Kaufman (1975). In fact, Kaufman found a three-factor solution for only 5 of 11 age

levels of the WISC-R normative sample. Table 3, shows a g-factor scores for males and females across age groups.

However, the order in which the factors emerged was different for males and females in some age groups.

Table (3)

g-factor score for males (M) and females (F)

across six age groups

	Age groups											
	6		7-8		9-10		11-12		13-14		15-16	
	N=100		N=200		N=200		N=200		N=200		N=200	
	M	F	M	F	M	F	M	F	M	F	M	F
INF*	.69	.75	.73	.76	.81	.76	.81	.78	.78	.82	.74	.80
SIM	.70	.25	.59	.57	.60	.66	.70	.66	.70	.73	.75	.78
AR	.76	.66	.75	.73	.68	.64	.76	.69	.73	.72	.71	.69
VOCAB	.70	.71	.73	.72	.70	.72	.74	.75	.81	.80	.72	.72
COM	.53	.48	.65	.62	.72	.71	.76	.74	.71	.69	.65	.67
DA	.51	.65	.61	.61	.48	.53	.67	.65	.49	.55	.63	.64
PC	.65	.68	.66	.70	.59	.66	.66	.61	.67	.72	.60	.63
PA	.73	.75	.74	.75	.66	.54	.62	.64	.64	.62	.59	.52
BD	.51	.49	.62	.69	.57	.64	.68	.68	.70	.73	.68	.65
OA	.52	.58	.53	.59	.46	.49	.60	.67	.48	.48	.58	.68
COD	.57	.14	.28	.42	.52	.58	.43	.50	.53	.58	.35	.35
MAZ	.60	.59	.68	.70	.62	.64	.54	.59	.63	.61	.56	.55

*INF = Information; SIM = Similarity; AR = Arithmetic

VOAV = Vocabulary; COM = Comprehension; DS = Digit Span;

PC = Picture Completion; PA = Picture Arrangement;

BD = Block Design; OA = Object Assembly;

COD = Coding; MAZ = Mases.

Congruence coefficients for the g-factor and the two-factor solution as well as the correlation coefficients for the same factors for 6 age groups are presented in table 4. Congruence coefficients for the g-factor ranged from.97 to.999. Checking their significance against the critical values reported by Cattell (1978) indicated that they were all significant beyond the.001 level of significance.

The congruence coefficients for the two-factor solution ranged from.83 to.99 and were all significant beyond the.01 level of significance.

Table: (4)

Congruence and Correlation coefficients of male and female g-factor scores and the two-factor solution across six age groups.

age-group	Congruence	factor score correlations for male-female		
	g-factor aF1 bF2	g-factor	aF1	bF2
6	. 970. 90. 83	.86	.72	.69
7-8	. 997. 95. 96	.95	.78	.84
9-10	. 996. 95. 91	.83	.83	.89
11-12	. 998. 98. 99	.92	.96	.91
13-14	. 999. 96. 89	.96	.90	.81
15-16	. 998. 98. 99	.93	.93	.84

a = First Factor.

b = Second factor

Factor score correlation coefficients of males and females across age groups ranged from .83 to .96 for the g-factor and from .69 to .96 for the two-factor solution and were all significant beyond the nominal level of.001.

Accordingly, the two methods of comparison of factorial similarity indicated factorial invariance of the Saudi version of the WISC-R across sex and age groups. Such findings give credence to the international application of the WISC-R, and add to the growing body of research regarding its construct validity.

Silverstein (1977) reached similar results when the two factor solution was considered. In fact Silverstein pointed out that the two-factor solution is more stable for different ethnic groups.

The present findings regarding factorial invariance of the WISC-R across sex and age were consistent with the findings of Reynolds and Gutkin (1980:777). They pointed out that "the three factor Varimax solution for the WISC-R is essentially invariant across sex and age with regard to the pattern of factor loading".

Table 5 indicates the factorial similarity of the g-factor of males and females as compared to the standardization sample's g-factor. The congruence coefficient was.99 for the male and.99 for the females. Both coefficients were significant beyond the.001 level of significane. The g-factor score correlations were.97 for the males and.97 for the females, and were all highly significant (P..001).

Table: (5)

Relation of the male and the female

g-factors to the g-factor of the normative sample

	standardization sample g-factor		female g-factor
	Congruence	Correlation	Correlation
male			.90
g-factor	. 99	.97	
Female			
g-factor	. 99	.97	

Accordingly, both congruence and correlation coefficients showed a great degree of similarity between the g-factor of the standardization sample and the g-factor scores of males and females.

Table 5 also indicates the correlation coefficient between the g-factor scores of total males and total females. As can be seen from table 5, the correlation was .90 which indicates a high degree of similarity of the g-factor scores of males and females.

Using Feldt's test (W) to compare the reliabilities of male and female scores across 10 subtests (Digit span and Coding were not included) yielded a nonsignificant value of W for all but Mazes subtest (W = 1.375) which was significant at the.05 level of significance (critical value at the .05 level is approximately 1.145).

Given the fact that Mazes subtest is among the supplementary tests, one can say that the Saudi version of the WISC-R is unbaised for either males or females, and the accuracy of measurement is equivalent across sex. Similar findings were reported by Oakland and Feigenbaum (1979). The largest difference reported by Oakland and Feigenbaum in the reliability coefficient between males and females was.03 which led to the conclusion that the WISC-R reliabilities for males and females are highly similar.

In conclusion, the Saudi version of the WISC-R appears to hold against factorial dissimilarity across sex and age as indicated by the congruence and correlation coefficient results. Furthermore, the findings of Feldt's W test seems to warrant the conclusion that the accuracy of measurement is about the same for males and females. Such findings give additional support to the construct validity of the Saudi version of the WISC-R. The WISC-R is newly adapted to Saudi society, more studies are needed to assure its validity and usablity.

NOTES

1-Nafie, A.; Qataee, A. & Saleem J. Test development for the identification of the gifted and talented (Part one). Reliability and Validity of the Saudi version of the WISC-R. Part of an on-going project titled "Identification and caring of the gifted and talented in Saudi Arabia" which is sponsored by King Abdulaziz City for Science and Technology (1990 -)

2-Nafie, A,; Qataee, A & Saleem, J. Test development for the identification of the gifted and talented (part one). Standardization of the Saudi version of the WISC-R. Part of an on-going project titled "Identification and caring of the gifted and talented in Saudi Arabia" which is sponsored by King Abdulaziz City for Science and Technology (KACST) (1990)

However, Permission was granted from KACST to use the normative data; and the author is grateful for that.

REFERENCES

AbuAllam, R.

1989 Manual of the Kuwait WISC-R. Kuwait: Psychological Service center; Ministry of Education (in Arabic).

Cattell, R.B.

1978 The Scientific use of factor analysis in behavioral and life sciences. New York: Plenum.

Cronbach, L.

1971 Test Validation. p.p. 443-507. In R. Thorndike (Ed.) Educational Measurement (2d ed). Washington D.C. American Council on Education.

Feldt, L.S.

1969 "A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. Psychometrika, 34 (4), 363-373.

Gutkin, T.B. & Reynolds, C.R.

1981 "Factorial similarity of the WISC-R for white and black children from the standardization sample". Journal or Educational Psychology, 73 (2), 277-231.

Herman, H.

1976 Modern Factor Analysis. Chicago: The University of Chicago Press.

Ismael, M. & Melaika. L

1983 Manual for the Wechsler Intelligence Scale for Children-Revised. Cairo: Librarei La Renaissance D'Egypt. (in Arabic)

Jensen, A.

1980 Bias in Mental Testing New York: The Free Press.

Kaufman, A.S.

1975 "Factor analysis of the WISC-R at 11 age levels between 6.5 and 16.5 years". Journal of Consulting and Clinical Psychology, 43 (2), 135-147.

Oakland, T. & Feigenbaum, D.

1979 "Multiple source of test bias on the WISC-R and Bender-Gestalt test". Journal of Consulting and Clinical Psychology, 47 (5), 968-974.

Olian, K & Kailani, A.

1988 "Psychometric Properties of the adapted Jordanian version of the WISC-R. Derasat 15 (1), 9-29. (in Arabic).

Reschly D.J.

1978 "WISC-R factor structure among Anglos, blacks, Chicanos, and Native Papagos". Journal of Consulting and Clinical Psychology, 46, 417-422.

Reynolds, C.R.

1983 "Test bias:In God we trust; all others must have data". Journal of

Special Education, 17 (3), 241-260.

Reynolds, C.R. & Gutkin, T.B.

1980 "Stability of the WISC-R factor structure across sex at two age levels". Journal of Clinical Psychology, 36 (3), 775-777.

Reynolds, C.R. & Harding R.E.

1983 "Outcome in two large sample studies of factorial similarity under six methods of comparison". Educational and Pscychological Measurement, 43 (3), 723-729.

Reynolds, C.R. & Kaiser, S.M.

1990 Test bias in psychological assessment. In T.B. Gutkin & C.R. Reynolds (Eds). The handbook of school psychology, New York: Wiley

Silverstein, A.B.

1977 Alternative factor analytic solution for the Wechsler Intelligence Scale for Children-Revised". Educational and Psychological Measurement, 37 (1), 121-124.

Vance. H.B. & Wallbrown, F.

1978 "The structure of intelligence for black children: A hierarchical approach". Psychological Record, 28, 31-39.

Wechsler, D.

1974 Wechsler Intelligence Scale for Children Revised. New York: Psychological Corporation.

Submitted December 1992.

Accepted Jul 1993.

FACTORIAL SIMILARITY AND ACCURACY OF MEASUREMENT OF THE SAUDI VERSION OF THE WISC-R ACROSS SEX AT SIX AGE GROUPS

Abdullah Qataee

Factorial similarity across sex and age for the Saudi version of the Wechsler Intelligence Scale for Children-Revised (WISC-R-) was investigated. Separate factors for males and females across age groups were arrived at through the use of principal-factor analysis on the normative sample's data. Similarities of factor structure for males and feamles across age groups were assessed via the use of congruence and correlation coefficients. Furthermore, Feldt's test (W) was used to assess constancy of errors due to domain sampling. The results indicated similarity of factors for males and females across age groups. Accuracy of measurement was about the same for males and females. Implication of the results for test validity is discussed.