Reliability and Validity of Wechsler Intelligence Scale for Children-Third Edition (WISC-III) in Iran

Jazayeri Ali Reza[1],PhD; Poorshahbaz Abas[2], PhD;

[1]Clinical Psychologist (PhD), Assistant Professor in Department of Psychology, The Science of Welfare and Rehabilitation University, Evin, Tehran, Iran.

[2]Clinical Psychologist (PhD), University teacher in Department of Psychology, The Science of Welfare and Rehabilitation University, Evin, Tehran, Iran.

ABSTRACT

The purpose of this study is to adapt, normalize and to determine reliability and validity of WISC-III. At adaptation stage, 835 students were selected through a systematic random sampling from 30 schools, after completion the test, suitable items were chosen according to Iranian culture. Then, at normalizing stage, 2456 boys and girls were selected through a systematic random sampling from 72 educational departments. The results showed that the reliability coefficients is higher than 0.95 and the validity is high the WISC-III has sufficient and satisfied reliability and validity in Iranian students.

Key Words: *RELIABILITY, VALIDITY, AND WISC-III*

Journal of Medical Education winter 2003; 2(2) 75-80

Introduction

One of the most noticeable problems in clinical psychology in our country is the diagnosis of mental retarded children from the other psychological disturbances and disorders in children with similar symptoms and signs. In the other hand, distinguishing the children with bright normal intelligence is also important. But, since this disorder has very unpleasant effects on both child and his/her family and the society, it is believed that these kinds of children should be Identified and provided with special education programs as soon as possible.

We should be careful that for diagnosing and measuring the levels of intelligence in children, we must use exact and appropriate instruments. A number of diagnostic and behavioral rating scales have been developed for those with mental retardation (2). In other word, we have to use an instrument with high reliability and validity. The Wechsler Intelligence Scale for Children (WISC) is one of the most valid and reliable diagnostic instruments for measuring intelligence levels (4).

The Weschler Intelligence Scale for Children-Third Edition (WISC-III) is the latest version of this scale. The WISC-III contains 13 subtests; three subtests are supplementary. Within the WISC-III verbal scale are information, comprehension, arithmetic, similarities, and vocabulary subtests. Digit span is a supplementary subtest. Within the WISC-III performance scale are the subtests of picture completion, picture arrangement, block design, object assembly, and coding (which parallels digit symbol-coding on the WAIS-III). The supplementary tests are a series of mazes that the child must traverse with a pencil and a symbol search consisting of paired groups of symbols. In this later subtest, each pair contains a target group and a search group, just as in the comparable WAIS-III subtest. These subtests, with the exception of mazes and symbol search, parallel the corresponding WAIS-III subtests in content and functions measured.

Reliability coefficients for the WISC-III were obtained. Split-half reliabilities for the WISC-III's Full-Scale, Verbal and Performance IQs average 0.96, 0.95 and 0.91 respectively; standard errors of measurement average 3.20, 3.53 and 4.54, respectively. Test-retest reliability coefficients are only slightly below those obtained using the split-half method.

Reliability coefficients for individual subtests are also comparable with most in the 0.70 s and 0.80 s. Test-retest coefficients were in the high 0.80s to low 0.99 s for the three IQs.

The correlation between the WISC-III and other Wechsler scales is rather good. Because the WISC-III overlaps with both the WAIS-III for 16-year-olds and the downward extension of the WISC-III, the WISC-III can be correlated with both the WAIS-III and the WIPPSI-R. All coefficients for the full-scale, Verbal, and performance IQs are in the 0.70 s and 0.80 s. The manual also reported correlations with the WISC-III and Stanford-Binet scale, with the majority of

coefficients in the 0.60 s and 0.70 s for individual subtests and in the 0.80 s to low 0.90 s for the three IQs (3).

Since, it was for the first time that we wanted to use the WISC-III in our country, so we had to normalize it and measures its validity and reliability. Of course, Shahim et al (5,6) normalized the WISC-R in Shiraz, but nobody has determined the psychometric characteristic of WISC-III. In this study, we normalized the WISC-III on Iranian junior and senior students during 4 years (1376 to 1380). Also, we measured its validities and reliabilities.

Methods

This descriptive study was performed in two stages. At the first stage as a pilot-study, through a multi-stages sampling, 835 boys and girls in Tehran, whose age rang were 6 to 17 years, were selected. Then the Wechsler Intelligence Scale for Children, Third Revised (WISC-III) was completed by them. The items, which were appropriate and consistent with their educational-cultural situations, were chosen for screening and assigning the best items for the second stage. In second stage, the sorted items from the first stage were completed by 2456 boys and girls, which were selected through a multi-stage cluster sampling. These participants were divided into 11 age-group. The mean age of these groups is 10.91 ± 3.12. The mean age of 51% of girls' participants was 10.94 ± 3.08 and 49% of boys' participants was 10.89 ± 3.15, which there was no statistical significance among these age differences (table 1).

WISC-III Reliability

To assign the scale reliability, we computed the coefficients of reliability through split-half reliability and test-retest. Also standard error of measurement for subscales, intelligence quotients and indicators were estimated. In addition, the results of scoring consistency assessment about some subscales, which their scorings were due to intellectual judgment, are shown.

Split-half reliability: The reliability for each subscales except digit symbol and object assembly were estimated through split-half reliability, and then these coefficients were corrected by Spearman-Brown formula. The coefficients of reliability for intelligence quotients and indicators were calculated by compound reliability. The coefficients of reliability for intelligence quotients and indicators are higher than each subscale. Because, the scores of intelligence quotients and indicators were obtained according to subscales scores, it estimates the child's performance in a wide range and in comparison to each subscales are more accurate.

Standard error of measurement: One of the most important indicators of scale scores accuracy is standard error of measurement. The standard error of measurement for each subscales, intelligence quotients and indicators are showed in each age-groups. The standard error of measurement provides an estimation of the range of scale scores errors and has a converse relation with the coefficients of reliability. Although, subscales standard error of measurement were lower than intelligence quotients and indicators, but it didn't mean that the subscales provided more accurate measures than intelligence quotients and indicators. It means that, this accuracy in measuring was due to different standard error measurement of scores. For each subscale a standard deviation was set at 3 and for IQs and indicators a standard deviation was at 1.5. The scores confidence limits are estimated according to standard error measurement and confidence level. Confidence limits therefore showed the accuracy rate of obtained scores and through providing the

TABLE 1 THE NUMBER OF FEMALE AND MALE STUDENTS AND TOTAL SAMPLE BY AGE-Group

Gender / Age (years)	Female N=1253	Male N=1203	Total N=2456
6	111	111	222
7	113	119	232
8	108	111	219
9	114	113	227
10	127	109	236
11	127	110	237
12	121	118	239
13	114	102	216
14	111	104	215
15	107	104	211
16	100	102	202

range of child's true score, so the accurate interpretation became possible.

Test-retest reliability: The coefficients of reliability of subscales, IQs and indicators were estimated by test-retest procedure. For this purpose, 151 students including 79 girls and 72 boys in 6 age-groups (6, 7, 10, 11, 14, and 15) were retested with a time interval of 14-20 days. The correlations between two test performances were calculated by Pearson correlation coefficient. The scale scores demonstrated satisfactory reliability and as it was anticipated that the total IQ showed higher score of 10 in retest. Also performance IQ is affected by rehearsal effects and becoming familiar with test higher than verbal IQ.

Scoring consistency: Most of the Wechsler subscales have high scoring consistency because of objective and accurate scoring instruction, just in similarities, vocabulary, comprehension and mazes subscales, the error probability increases because scoring is based on subjective judgement. So the reliability of scores in these subscales was re-examined. For this purpose, 60 answer sheets were randomly selected among normalized sample, then they were scored independently by 4 specialists and finally, the scoring consistencies for these subscales were calculated by estimation of correlations between these 4 scores.

The results showed that the reliability among 4 scores was 0.95 for similarities, 0.98 for vocabulary, 0.97 for comprehension and 0.99 for mazes. These results illustrated the subscales, which are related to subjective judgment, have high reliability in scoring.

Scale Validity

The validity evidence for each scale is obtained by a host of methods. These data should demonstrate that (A) the test measures exact structures that it is designed for; (B) there are significant correlations between the results of this test and the results of similar tests; (C) this test can discriminate special groups of children.

In this part, we demonstrate the results of validation measurements.

Construct validity

In internal validity the experimental evidence about the structure of scale scores are assessed. For this purpose, we used the correlation between subscales, IQs and indicators, and we have also done confirmatory factor analysis.

The results of correlation among subscales, IQs and indicators in total sample showed that there are high correlations among verbal subscales with each other and with verbal intelligence quotient. Also, the performance subscales showed a high correlation with each other and with performance intelligence quotient.

The one-factor, two-factor, three-factor and four-factor analyses were done. As the coefficients of each subscales show, one-factor analysis approves general factor, two-factor analysis approves the verbal and performance sections in scale, three-factor analysis adjusts on the results of exploratory factor analysis, in which subscales of similarities, information, comprehension, vocabulary and arithmetic are variables expressed by the first factor; subscales of picture completion, picture arrangement, block design, object assembly and mazes are variables expressed by the second factor; and subscales of digit span and coding are in the variables expressed by the third factor. Four-factor analysis is adjusted on four factors of verbal comprehension, perceptual organizing, freedom from distractibility and the speed of processing in WISC-III. In four factors analysis subscale of mazes is a variable expressed by the second factor.

Criterion Validity

To assess the convergent validity of WISC-III, the results of this scale were compared with other intelligence scales and academic progress indicators. A sample consists of 48 students from 6 to 13 years, 21 girls and 27 boys, with the mean age of 10, assessed after measuring by WISC-III with a 14-20 days interval, were assessed by WISC-R. The means and standard deviations for two scales and their coefficients of correlation were measured. The correlation among subscales scores and IQs' scores show the scales' structural similarities with each other. Also, these two scales have lower correlation in performance subscales, which might be resulted by the significant differences in performance section of WISC-III compared with WISC-R. Although, the IQs of WISC-R compared with WISC-III are higher in verbal section about 12, in performance about 22 and in total IQ about 19 score, but, since these two scales aren't matched with each other according to their administration order, we can't claim surely about these differences.

A sample consists of 93 students with the age of 6 to 13 years, 41 boys and 52 girls, with the mean age of 10 years 1 hour after WISC-III, were assessed by Raven scale for students in age group of 6-8 years we used colorful Raven and the

students with age-group of 9-16 years were assessed by white & black version of Raven. The results showed that the correlation between verbal IQ and Raven is 0.32, performance IQ and Raven is 0.54 and total IQ and Raven is 0.49. This scale in WISC-III verbal section has the highest correlation with similarities (r=0.40) and in performance section have the highest correlation with block design (r=0.63). To assess the relation between students' academic abilities and their IQ, the correlation between the IQs and Indicators of 40 students including 19 boys and 21 girls with mean age of 12 years with their scores in reading, arithmetic, science, dictation and writing were estimated. As it is showed in table 2, WISC-III verbal IQ has the highest correlation with these subjects. These results are consistent with previous studies in this area.

On the other hand, reading ability has the highest correlation with verbal comprehension indicator (r=0.40); dictation has the highest correlation with speed of processing (r=0.49), writing has the strongest correlation with verbal comprehension (r=0.42), and arithmetic has the highest correlation with freedom from distractibility (r=0.63). These results approve the WISC-III validity and are consistent with another similar study.

TABLE 2 CORRELATION COEFFICIENTS BETWEEN IQS SCORES AND WISC-III INDICATORS WITH SUBJECTS

Subject / IQ indicator	Reading	Dictation	Writing	Arithmetic	Science
Verbal IQ	0.38	0.40	0.42	0.48	0.67
Performance IQ	0.29	0.16	0.25	0.25	0.27
Total IQ	0.32	0.38	0.40	0.46	0.56
Verbal perception	0.40	0.41	0.42	0.40	0.62
Perceptual organizing	0.07	0.10	0.13	0.22	0.15
Freedom of distractibility	0.23	0.42	0.26	0.63	0.66
Processing speed	0.29	0.49	0.40	0.40	0.44

Assessing special groups: In table 3 the means and standard deviations of IQs and indicators of bright normal children, students with high and low academic progress and mental retarded children are shown.

WISC-III was performed on 70 bright normal student including 35 boys and 35 girls with the mean age of 13 years. Their total IQs' mean was 134.8, verbal IQ 136.1 and performance IQ 124.7. All of the three IQs were high normal in the range of bright normal. Only 1.5 percent of these students, obtained total IQ lower than 120. These students got the highest scores in freedom from distractibility. In other word, These students have a high speed of processing as well as a high ability of concentration and attention. In addition, WISC-III can discriminate bright normal students from students with high academic progress. As it is shown in table 3, there is a statistical significant difference between IQ and indicator's scores of bright normal students and students with high academic progress. The most important differences are their difference in speed of processing. This indicator in students with high academic progress is in the normal range but in bright normal students s more than 2 standard deviation higher than mean.

WISC-III was performed on 60 mental retarded students, who were study in special schools; including 30 girls and 30 boys with the mean age of 11.3 years. Their mean of total IQ was 53.4, verbal IQ was 55.8 and performance IQ was 54.8. All three IQs were in the range of IQ that can be taught. None of the students had total IQ higher than 69 and all of them were labelling as mental retarded accurately. These students got the scores lower than 69 in all indicators. To the contrary of normal students, these students got the lowest scores in freedom from distractibility and had failure in concentration and attention abilities. WISC-III can discriminate mental retarded students from students with low academic progress. As it is shown in table 3, there are significant differences between IQ and indicator scores in students with low academic progress and in educatable mental retarded children. Students with low academic progress had the most differences in verbal perception and perceptual organizing with mental retarded children.

Table 4 shows the means and standard deviations of WISC-III IQ and indicators scores in children with behavioral disturbances, and in deaf and blind children.

In the group of children with behavioral disturbances, there were 39 male students with mean age of 11.6 years. They had disturbances such as attention deficit hyperactive disorder (ADHD), conduct disorder, and oppositional

TABLE 3 IQS SCORE AND WISC-III INDICATORS IN BRIGHT NORMAL, HIGH AND LOW ACADEMIC PROGRESS, AND MENTAL RETARDED CHILDREN

Groups / Scales	Bright normal	High academic progress	Low academic progress	Mental Retarded
Total IQ	134.8 ± 15.2	122.1 ± 14.4	85 ± 16.9	53.4 ± 11.8
Verbal IQ	136.8 ± 11.9	123.9 ± 15.5	85 ± 17.4	55.8 ± 12.1
Performance IQ	124.7 ± 13.2	112.8 ± 13.8	86.9 ± 17.9	54.8 ± 12.7
Verbal perception	131.5 ± 11.6	121.4 ± 15.6	88.2 ± 15.3	55.8 ± 12.1
Freedom of distractibility	136.6 ± 14.4	120.8 ± 12.6	82.8 ± 16.6	53.5 ± 11.8
Perceptual organizing	123.1 ± 14.2	113.2 ± 14.8	89.9 ± 17.7	55.6 ± 11.1
Processing speed	136.8 ± 11.3	108.6 ± 13.2	82.2 ± 14.2	56.8± 10.3

defiant disorder. It was possible that some mental retarded students were in this group but they weren't screened. Based on their IQs and indicators scores they were educatable mental retarded students. Except freedom from distractibility, there were significant differences among their IQs and indicators means with mental retarded children (α=0.05). It means that their cognitive abilities were higher than mental retarded children.

In addition, the comparison among performance and verbal subscales in behavioral disturbed children with mental retarded shows that these children have higher abilities in subscales of information, arithmetic, vocabulary, comprehension, blocks design, picture completion, object assembly, digit symbol and mazes than mental retarded children (P=0.01). So, these subscales can be helpful and appropriate in differentiating behavioral disturbances from mental retardation. There were no significant differences in subscales of similarities; digit span, coding and picture arrangement between behavioral disturbed children with mental retarded group.

Deaf children (n=35) including 15 boys and 20 girls with mean age of 11.6 years. Based on the means of IQs and indicators scores, their scores in verbal section and verbal perception and freedom from distractibility were like borderline mental retarded children, and in performance section and perceptual organizing and the speed of processing they were like normal children. Finally, their total IQs were dull normal. Based on their IQs, deaf children get higher scores in performance sections especially in comprehension and vocabulary; their abilities were weak. This shows their partial deficits in verbal perception and freedom from distractibility.

Blind children,(n=40) including 20 boys and 20 girls with the mean age of 11.9 years. For this group, performance section was impossible. In verbal section, their abilities were like normal children, especially, they got the lowest score in freedom from distractibility. It might be for their higher concentration on

auditory stimulus and nonintervrtion of visual distractive stimulus. Also, this group had the highest abilities in arithmetic and digit span subscales, but in subscales of comprehension, which is influenced by social judgement, their performances were weak.

Discussion

This research showed that WISC-III is a very valid and reliable scale for measuring intelligence functions in children. This scale has high reliability and as it was shown, there is a high internal consistency and a wide admission between specialists for scoring it.

The correlation between verbal and performance

TABLE4 THE MEANS AND STANDARD DEVIATION OF IQ SCORES AND WISC-III INDICATORS IN CHILDREN SPECIAL GROUPS

Special groups / Scales	Behavioral Disturbance	Deaf	Blind
Total IQ	59 ± 17.9	81± 18.1	-
Verbal IQ	61 ± 17.4	69 ± 16.2	95 ± 15.1
Performance IQ	62 ± 18.1	102 ± 17.8	-
Verbal perception	63 ± 14.8	69 ± 16	93 ± 12.1
Perceptual organizing	63 ± 15.3	103 ± 17.2	-
Freedom of distractibility	56 ± 16.9	77 ± 16.3	107 ± 7.8
Processing speed	62 ± 15.6	97 ± 15.9	-

sections and also the correlation among total IQ and each section in all age-groups as statistical significant. This correlation model provides the evidences of concurrent validity (1). The lower correlation among performance and verbal subscales shows discriminative validity that demonstrates some factors in this scale.

Evidence from factor analysis and the correlation of the scales results with other intelligence scales and academic progress scales approve construct validity. Also this scale is a reliable instrument for clinical judgments. The data in specific groups shows the usefulness of this scale as a diagnostic instrument. Scale indicators approved their efficacies in assessments and they are also appropriate in differential diagnosis.

References

1- Golombok, S. & Rust, J. (1992). Wechsler Intelligence Scale for Children. Third Edition, UK, the psychological corporation.

2-Kaplan, H. & Sadock, B. (2000). Comprehensive Textbook of Psychiatry, William & Willkins, USA.

3-Kaplan. R. M. & Saccuzzo, D. P. (2001). Psychological Testing, Fifth Edition, Wadsworth, Canada.

4-Marnat, G. G. (2000). Neuropsychological Assessment in Clinical Practice, A Guid to Test Interpretation and Integration, Jhon Wiley & Sons, USA.

5-Shahim, S. (1369). Normalization and Adaptation of WISC-R in Iran. Shiraz University.

6-Shahim, S. (1371). Appling the WISC-R in Iran, psychological researches, No 3,4.