

Summary

Comparison of WISC-R and WISC-IV Intelligence Scores in Gifted Children

İbrahim Yiğit

Ankara University

Cihat Çelik

Ankara University

Gülşen Erden

Ankara University

Gifted children have been identified as having higher performance in the areas of academic, intellectual ability, creative or productive thinking, arts, esthetic, sports and leadership than their peers (Marland, 1972; Pfeiffer, 2002; Sattler, 2002). Gifted children make rapid progress in these areas; and they need differentiated and specific programs beyond the regular education programs (Sattler, 2002; Smutny & Blocksom, 1990). In this respect, gifted children should be assessed and identified properly for being educated in schools tailored for their abilities.

Various assessment tools are used to assess and identify gifted children. The assessment tools generally include standardized tests (intelligence, achievement, social maturity and personality tests), parent and teacher reports, observations on child's behaviors and assessment of child's creativity (Alvino, McDonnell, & Richert, 1981; Sattler, 2002). Especially, individually administered intelligence tests are defined as the most valid and reliable tools. These tests are often used in order to identify intellectual skills of gifted children and place them into necessary education programs (Newman, Sparrow, & Pfeiffer, 2008; Pfeiffer, 2002; Sattler, 2002). Although there are many intelligence tests in the assessment of gifted children, it is reported that Wechsler intelligence scales are used widely (Alvino et al. 1981; Avant, 1987; Roid, 1990; Stinnet, Harvey, & Oehler-Stinnet, 1994; Wilson & Reschly, 1996). Wechsler Intelligence Scale for Children (WISC) was developed to identify intelligence levels of children aged between 6-16 years and yielded various subtests and composite or index scores (Wechsler, 2003). The key criterion in determining gifted children is intelligence score -total or general intelligence quotient (IQ) score. Accordingly, a child obtaining a total IQ higher than or equal to 130 is identified as gifted (Sattler, 2002; Pfeiffer, 2012).

Scores obtained from intelligence scales have been observed to having systematic and general increase around the world (Flynn, 1987; Kanaya, Scullin, & Ceci, 2003). This increase in the intelligence score may be re-

sulting from many factors such as schooling, enrichment in environmental stimulus, increased familiarity with the test, and developments in the nutrition and health fields (Lynn, 2009). According to the phenomenon known as Flynn effect, increase in general IQ scores over time cause intelligence scale norms to be obsolete. In other words, as time elapses and intelligence scale norms lose their validity, assessed children shows increasingly higher performance on the scale. Thus, there exists a 3-point increase in IQ scores in every ten years (Flynn, 1984, 1987, 1999; Steen, 2009). Considering Flynn effect, standardization and norm studies of intelligence scales should be revised periodically. These revisions come up score differences between old and new versions. Studies comparing two versions of intelligence scales in gifted children revealed that gifted children obtained higher scores from the old versions of WISC than the newer versions (Bryant, 1992; Larrabee & Holroyd, 1976; Sevier et al., 1994; Wechsler, 1991; Wheaton & Wandergriff, 1978).

The use of WISC in our country began with the introduction of WISC-R (Wechsler Intelligence Scale for Children-Revised). Adapted into Turkish by Savaşır and Şahin (1995), WISC-R has been used widely for years in hospitals, clinics, Guidance Research Centers and research areas. Specifically, in the assessment of children who need tailored education, it plays an important role as an assessment tool (Çelik, Yiğit, & Erden, 2015; Uluç et al., 2011). Standardization and norm study of WISC-IV that is the fourth version of Wechsler intelligence scales (Wechsler, 2003) has just been completed; and the process of its use has been started in our country (Uluç et al. 2011).

Compared to WISC-R, WISC-IV has several important modifications in its construct, content and interpretation (Flanagan & Kaufman, 2009; Wechsler, 2003). According to these modifications, two-factor structure in both WISC-R and WISC-III was eliminated completely and replaced four new index scores (composite scores) as Verbal Comprehension Index (VCI), Perceptual

Reasoning Index (PRI), Processing Speed Index (PSI), Working Memory Index (WMI). Picture Arrangement, Picture Completion and Mazes subtests were eliminated from the scale while Picture Concepts, Letter-Number Sequencing, Matrix Reasoning, Symbol Search, Cancellation and Word Reasoning subtests were added. Another change was in the Information and Arithmetic subtests, which were the core subtests in WISC-R. They were transferred to the section of additional subtests in WISC-IV. Moreover, item numbers of the subtests were increased by adding items at the low and high levels of difficulty. Finally, several changes about the test materials and administration instructions were made; and the administration manual was made more functional.

Given that our country directly have begun to use WISC-IV, skipping WISC-III (Uluç et al., 2011), it is important to examine the validity and effectiveness of WISC-IV in the identification of gifted children and to determine score differences between WISC-R and WISC-IV. Thus, in the light of the studies showing the score differences between two versions of WISC, the purpose of the present study is to compare WISC-R and WISC-IV scores in a sample of gifted children. In other words, it aims to determine whether expected score differences between these versions exist or not.

Method

The sample of the study consisted of 50 gifted students (ages 6-16) in the city of Ankara. WISC, developed by Wechsler in 1949, is an individually administered intelligence test for children between the ages of 5 and 15. The scale was revised in 1974 as WISC-R and the age range of the scale was changed to 6-16. In our country, the standardization study of WISC-R was performed with a sample of 1639 children who were chosen from 11 different city center and it was adapted to Turkish culture by Savaşır and Şahin (1995). WISC-R consists of 10 core and 2 supplemental subtests and organizes into Verbal (Information, Similarities, Arithmetic, Comprehension, Vocabulary, Digit Span) and Performance (Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, Mazes) composite scores.

WISC-IV is an individually administered intelligence test for children between the ages of 6 and 16 (Wechsler, 2003). It consists of totally 10 core subtests and 5 additional subtests, including 4 factors. WISC-IV provides 4 index scores as VCI (Similarities, Vocabulary and Comprehension core subtests; and Information and Word Reasoning supplemental subtests), PRI (Block Design, Picture Concepts and Matrix Reasoning core subtests; and Picture Completion supplemental subtest), WMI (Digit Span and Letter-Number Sequencing core

subtests; and Arithmetic supplemental subtest) and PSI (Coding and Symbol Search core subtests; and Cancellation supplemental subtest). FSIQ score is obtained by the sum of 10 core subtest scores.

Before the children were included in the study, necessary information about the test was conveyed to their parents and school board. Then, those wished to participate were given informed consent forms. The administrations were conducted by psychologists who have both WISC-R and WISC-IV Administration Certificate being given by Turkish Psychological Association. Children who obtained Full Scale IQ higher than or equal to 120 on WISC-R and had not been administered WISC-R for at least 6 months were included in the study.

Results

A paired sample t-test was conducted to test whether there were significant differences between WISC-R IQs and subtest scores; and WISC-IV index and subtest scores. Results indicated that WISC-R Full Scale IQ score was significantly higher than WISC-IV FSIQ ($t = 3.99, p < .001$); and WISC-R Performance IQ score was significantly higher than WISC-IV PRI ($t = 10.70, p < .001$). Similarly, WISC-R Similarities, Block Design and Coding scores were significantly higher than same scores on WISC-IV ($t = 2.06, p < .05$; $t = 4.80, p < .001$; $t = 7.52, p < .001$, respectively). In contrast, WISC-IV Vocabulary score was significantly higher than WISC-R Vocabulary score ($t = -7.57, p < .001$).

The Pearson correlation coefficients were calculated for examining relationships between WISC-R IQ scores and WISC-IV index scores. Accordingly, relationships between WISC-R Performance IQ and WISC-IV PRI; WISC-R Full Scale IQ and WISC-IV FSIQ were found to be significant ($r = .41, p < .01$; $r = .49, p < .01$, respectively). However, WISC-R Verbal IQ was not significantly related to WISC-IV VCI ($r = .27, p > .05$).

Discussion

In this study, children identified as gifted with WISC-R were administered WISC-IV and scores from both of the two versions were compared. The results of the study showed that these children scored higher on WISC-R Performance IQ and Full Scale IQ than WISC-IV PRI and FSIQ scores. Although there is no study comparing WISC-R and WISC-IV intelligence scores in gifted children, the findings of studies comparing other versions of WISC have supported the findings of the present study (e.g., Larrabee & Holroyd, 1976; Sabatino et al., 1995; Sevier et al., 1994), indicating that old version of the scale produced higher scores.

These results were also supported by Flynn's studies (1984; 1987) proposing that intelligence scale norms should be revised in every ten years. Flynn (1987) reported 5-25 score differences between Full Scale IQ scores. Given that WISC-R norm study for our country was 30 years ago, 10-point increase might be expected in the intelligence scores according to the Flynn effect. In this study, score difference of 5.88 was found and this difference was consistent with Flynn's findings. Additionally, there was a score difference of 15.18 between WISC-R Performance IQ and WISC-IV PRI. According to one of Flynn's (1987) findings, scores obtained from WISC-R subtests in Performance IQ increased more than twice of scores obtained from Verbal IQ subtests. In this sense, the score difference found in our study is consistent with the related literature.

The correlation analysis was conducted to determine relationships between WISC-R IQ scores and WISC-IV index scores. The results of the analysis showed that there were significant relationships among Full Scale IQ/FSIQ and Performance IQ/PRI. These strong and significant relationships between IQ and index scores were thought to address the continuity of both versions and support each other in assessment of the intelligence. In our country, studies with normal (Uluç et al., 2014) and clinical (ADHD-diagnosed children) (Çelik ve ark., 2017) samples produced similar correlation coefficients with the present study.

Another type of score used in the assessment of gifted children is General Ability Index (GAI) score. GAI score is obtained from the sum of the subtest scores in VCI and PRI and there is a norm table for calculation of GAI. When high inconsistencies between the index scores exist, GAI is suggested to be used (Flanagan & Kaufman, 2004). These two indexes which gifted children obtained higher scores and evaluate abstract and visual reasoning, have been revealed to be better indicators in identifying gifted children than both WMI (manipulated auditory memory tasks) and PSI (speed in paper and pencil tasks) indexes (Falk, Silverman, & Moran, 2004; Flanagan & Kaufman, 2004; Weiss, Saklofske, Prifitera, & Holdnack, 2006). In this regard, conducting a norm study for calculation of GAI scores in our country may be useful in terms of both reducing administration time and cost; and revealing intellectual abilities of gifted children more precisely.

The use of additional assessment tools in the identification of gifted children may contribute to the assessment to be able to make it in a healthier way. The studies in the literature emphasized multiple assessments in the identification of gifted children. In a study in our country, it was reported that the use of Raven Progressive Matrices Test as well as WISC-R presented better

results for the identification of gifted children (Bildiren & Uzun, 2007). According to Kaufman and Harrison (1986), when intelligence scales are used properly, they are important clinical tools in the assessment of gifted children. However, these authors also reported that intelligence scales did not generate precise results and were limited to measure abilities such as creativity and artistic, suggesting that the use of multiple assessment criteria was important. Consequently, gifted children exhibits musical, creative, athletic, esthetic and artistic abilities (Pfeiffer, 2002; Sattler, 2002); and the use of multiple assessment tools may enable diverse intelligence areas to be assessed together. This study, as it revealed score differences between two versions, have been addressed the need for an additional assessment tool in the assessment of gifted children.

The intelligence score differences in the present study may often be thought as a result of the radical changes in the structure and content of new version (WISC-IV). Therefore, although both of the two versions measure same structures, several issues should be considered. The first issue is the use of different subtest scores in the calculation of Full Scale IQ/FSIQ score in the two test versions. The core subtests of Information and Arithmetic that gifted children had higher scores on WISC-R were transferred to the section of supplemental sub-tests in WISC-IV. In a study with WISC-IV, it was found that Arithmetic and Information subtests were amongst the strongest measures of giftedness (Silverman et al. 2004). Moreover, our study showed that these children obtained average scores from Picture Concepts, Matrix Reasoning, Letter-Number Sequencing and Symbol Search subtests, which were newly added to WISC-IV. In this regard, it is most likely to be score differences between the two versions. It is believed that more studies are necessary for supporting these findings and determining whether the score differences may be due to a number of changes in the new version. The second issue is that the revision of WISC-R for our country after a period of 30 years may cause such differences between the two versions. As reported by Uluç and colleagues (2013), our country has a foreign-source dependency since it has not yet produced its own intelligence scale. It is obvious for Turkey to be late for revising WISC-R's norm sample and standardization. In conclusion, revising test versions periodically (in every 10 years) would contribute to the test to be strong and valid assessment tool.