


**CLINICAL INTERPRETATION OF
THE WOODCOCK-JOHNSON
TESTS OF COGNITIVE ABILITY—
REVISED**

Kevin S. McGrew
St. Cloud State University

1994

Allyn and Bacon
Boston • London • Toronto • Sydney • Tokyo • Singapore

The information summarized in Table 8-5 can be used to explain differences between full-scale scores on the WJTCA-R and other intelligence batteries. For example, if an individual is administered both the WJTCA-R and the K-ABC and significant full scale score differences are found, the individual tests of each battery should be carefully inspected. Generally, similar levels of performance might be expected on the WJTCA-R and K-ABC tests that measure the same *Gf-Gc* abilities (see Table 8-5). The most likely explanation of the full scale score difference may lie in those WJTCA-R tests that measure *Gf-Gc* abilities not measured by the K-ABC (i.e., long-term retrieval, auditory processing, and comprehension-knowledge). The information presented in Table 8-5 may help practitioners better understand full-scale score differences between the WJTCA-R and the other intelligence batteries.

Given the prominence that the achievement content hypothesis played in generating significant misconceptions about the original WJTCA (McGrew, 1986), a misconception this author hears being generalized to the WJTCA-R, it is important to examine the achievement-related content of the major intelligence batteries. As summarized in Table 8-5, when achievement-related content is defined as tests measuring quantitative (*Gq*) and comprehension-knowledge (*Gc*) abilities (i.e., crystallized intelligence), the original WJTCA contained 29% of such content. The SB-IV and Wechsler scales consist of 50% to 60% achievement content. The achievement content criticism of the original WJTCA was inaccurate (McGrew, 1986).

More importantly, the WJTCA-R consists of tests that only reflect 14% achievement content. *The WJTCA-R has approximately three to four times less achievement content than the SB-IV and Wechsler scales.* The only intelligence battery included in Table 8-5 that has a lower degree of this type of content is the K-ABC (i.e., 6%).

THE TOTAL DOES NOT EQUAL THE AVERAGE OF THE PARTS

Things Don't Add Up

Traditionally, most intelligence batteries (e.g., Wechslers, K-ABC) report subtest scores in a standard score metric with a mean of 10 and a standard deviation of 3. The SB-IV is idiosyncratic (Kamphaus, 1993), as individual test scores are reported on a standard score scale with a mean of 50 and a standard deviation of 8. The WJTCA-R is different from all the major intelligence batteries by providing individual test scores based on the same metric used to report full-scale or composite scores (i.e., mean = 100; SD = 15). There has been an unintended outcome of using the same scale for the individual WJTCA-R tests and the composite cluster scores. Clinicians are becoming confused by a statistical phenomenon that has always been present

in the interpretation of intelligence test scores but which up until now has been hidden.

The issue is that a composite standard score (e.g., Broad Cognitive Ability cluster; Wechsler Full Scale) is not the simple average of the individual standard scores that are comprised in the composite. For example, a student tested at grade 6.5 obtained standard scores (mean = 100; SD= 15) that ranged from 73 to 88 on the seven WJTCA-R Standard tests. Although the average of the seven standard scores was 81, the Broad Cognitive Ability cluster standard score was 67. This reflects a 14-point standard score difference between the average and the obtained composite cluster standard score. Why? Is there something wrong with the WJTCA-R norm tables? It has been this author's observation that this phenomenon has generated questions among a number of practitioners. However, this observation is not unique to the WJTCA-R.

A hypothetical WISC-R case study is presented in Table 8-6. This example assumes that an individual obtained scaled scores of 4 (2 standard deviations below the mean) on all 10 individual WISC-R tests. When this performance is converted to the standard score scale with a mean of 100 and standard deviation of 15, the individual test scores are 70. The average of these standard scores is 70. However, the obtained WISC-R Full Scale score is 59, a score that is 11 points lower than the average of the subtest standard scores.

An Explanation

The phenomenon of a total score not equaling the average of the individual tests is not unique to the WJTCA-R. This apparent aberration will be observed for most all tests that provide composite scores. However, this discrepant finding is typically masked in all other intelligence batteries by the use of different standard score metrics for reporting the individual and composite scores. Why does this occur?

This phenomenon is most noticeable the farther away from the mean a person consistently scores on tests that make up a composite. To be significantly below or above the mean on a number of different tests is a less frequent or rarer occurrence than being low or high on just one or two of a number of tests. The composite score will reflect this "rareness" by being lower (in the case of consistently low performance on a number of tests) or higher (in the case of consistently high performance on a number of tests) than the average, as the purpose of the standard score (or percentile rank) is to "indicate the likelihood in a population of obtaining a particular score or combination of scores" (McGrew et al., 1991, p. 42).

This phenomenon is a function of the intercorrelations and number of tests comprised in the composite score (Paik & Nebenzahl, 1987). The lower the intercorrelations between the tests that make up a composite score, the more extreme this average/obtained score discrepancy will be (McGrew et al., 1991). This phenomenon is particularly noticeable for the WJTCA-R

TABLE 8-6
WISC-R E

Subtest
Information
Similarities
Arithmetic
Vocabulary
Comprehension
Picture Completion
Picture Arranging
Block Design
Object Assembly
Coding
Average Standard Score
Full Scale Score
Difference

Standard scores are made distinct and never phenomenon. Although phenomenon difference. As presented measure a degrees of phenomenon dramatic. equal the individual purpose of measure of. Research phenomenon Broad Cognition. The same is usually

**TABLE 8-6 The Total Does Not Equal the Average of the Parts:
WISC-R Example**

Subtest	Scaled Score (Mean = 10, SD = 3)	Standard Score (Mean = 100, SD = 15)
Information	4 (-SD)	70 (-2 SD)
Similarities	4 "	70 "
Arithmetic	4 "	70 "
Vocabulary	4 "	70 "
Comprehension	4 "	70 "
Picture Completion	4 "	70 "
Picture Arrangement	4 "	70 "
Block Design	4 "	70 "
Object Assembly	4 "	70 "
Coding	4 "	70 "
Average Standard Score	= 70	
Full Scale Standard Score	= 59	
Difference	= 11	

Standard or Extended Broad Cognitive Ability clusters since these composites are made up of 7 to 14 tests that have been designed to be measures of distinct abilities (i.e., low intercorrelations). The reader should consult Paik and Nebenzahl (1987) for a detailed treatment of the statistical basis of this phenomenon.

Although the example presented in Table 8-6 indicates that the same phenomenon occurs for the WISC-R, the magnitude of the observed score difference will be less with intelligence batteries that are more homogeneous. As presented in Table 8-5 and Chapter 4, most other intelligence batteries measure a smaller range of *Gf-Gc* abilities and, thus, typically show higher degrees of intercorrelations among the individual tests. As a result, this phenomenon will occur in other intelligence batteries, but it will be less dramatic. The only situation where a composite intelligence test score will equal the average of the individual tests for all individuals is when all the individual tests are perfectly correlated. Such a situation would defeat the purpose of having intelligence batteries that contain individual tests that measure different abilities.

Researchers and clinicians need to be aware that this score discrepancy phenomenon will be seen frequently in the WJTCA-R, particularly for the Broad Cognitive Ability cluster standard scores. Clinicians should be reassured that there is nothing "wrong" with the WJTCA-R norm tables or scores. The same phenomenon occurs with all intelligence test batteries, although it is usually masked by different individual test and composite score metrics.