

# Clinical Interpretation of the Woodcock-Johnson Tests of Cognitive Ability



Kevin S. McGrew

### Reading Aptitude

Antonyms-Synonyms  
.42

Visual-Auditory Lrng.  
.23

Analogies  
.18

Blending  
.17

### Math Aptitude

Antonyms-Synonyms  
.39

Analysis-Synthesis  
.27

Visual Matching  
.26

Concept Formation  
.08

### Written Language Aptitude

Quantitative Concepts  
.47

Antonyms-Synonyms  
.29

Visual Matching  
.15

Numbers Reversed  
.08

### Knowledge Aptitude

Antonyms-Synonyms  
.46

Quantitative Concepts  
.28

Analogies  
.16

Memory for Sentences  
.10

Figure 7.1. Subtest composition and weighting of WJTCA Scholastic Aptitude clusters.



## THE CLINICAL UTILITY OF THE SCHOLASTIC APTITUDE CLUSTERS p-212

Woodcock (1984b) notes that the WJTC Scholastic Aptitude clusters provide for higher clinical validity by providing the capability to distinguish more accurately between Type I (viz., aptitude-achievement) and Type II (viz., intracognitive) discrepancies. Woodcock notes that most broad-based aptitude measures often confound these discrepancies, a situation that can often result in inappropriate diagnostic conclusions and interventions. This clouding of Type I and II discrepancies is the result of broad-based ability measures including subtests that contribute to the global score, however, some of these subtests may have little ac-

tual relationship to the aptitude of concern. This point is best demonstrated by two hypothesized scenarios.

Table 7.13 presents the impact of a hypothetical student who

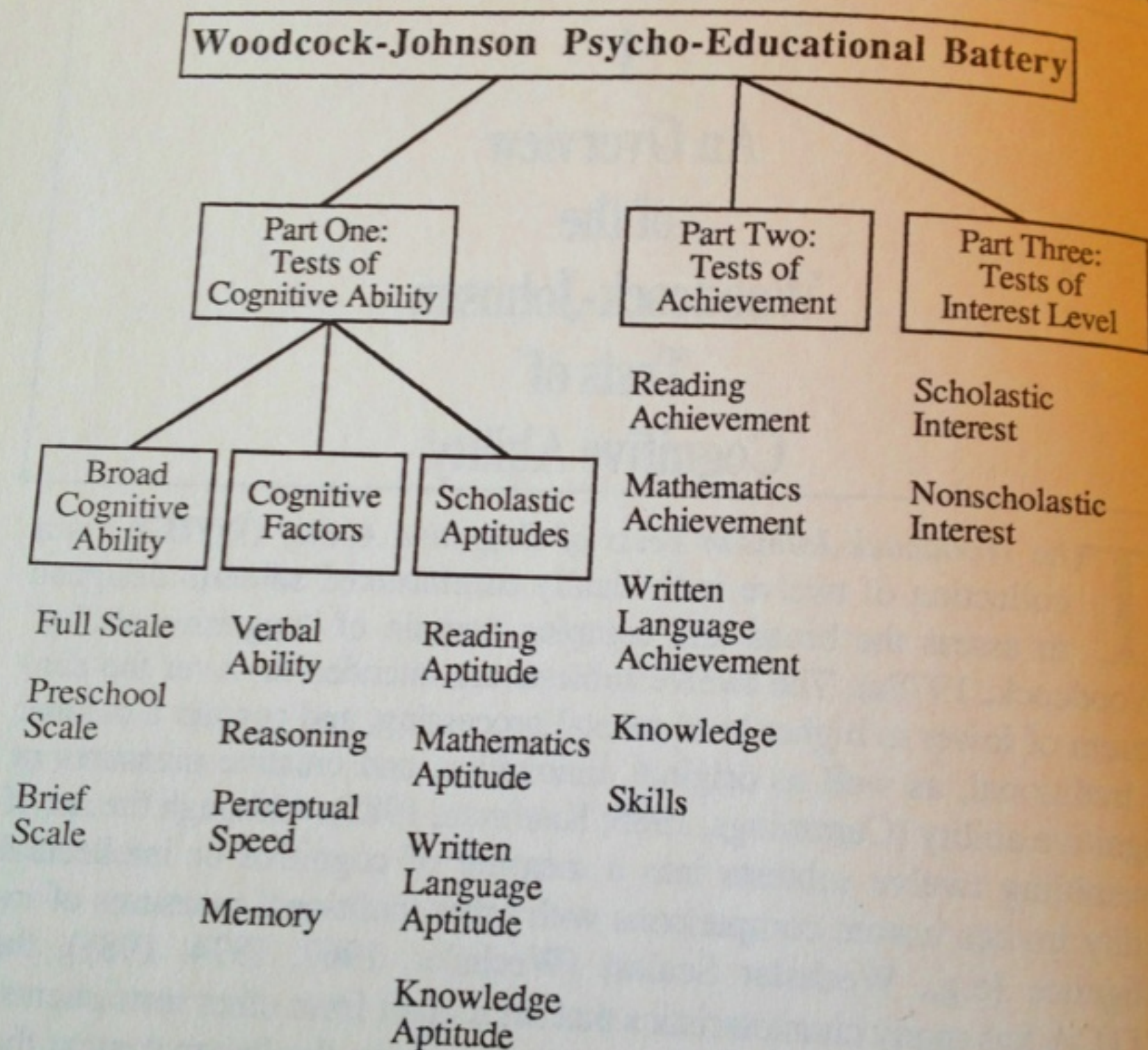
## SUMMARY COMMENTS REGARDING THE CLINICAL UTILITY OF THE SCHOLASTIC APTITUDE CLUSTERS *p 216*

To summarize, the observation that the Antonyms-Synonyms subtest may lower all four Scholastic Aptitude cluster scores (due to its relatively high weighting) is exactly what is desired. The purpose of the Scholastic



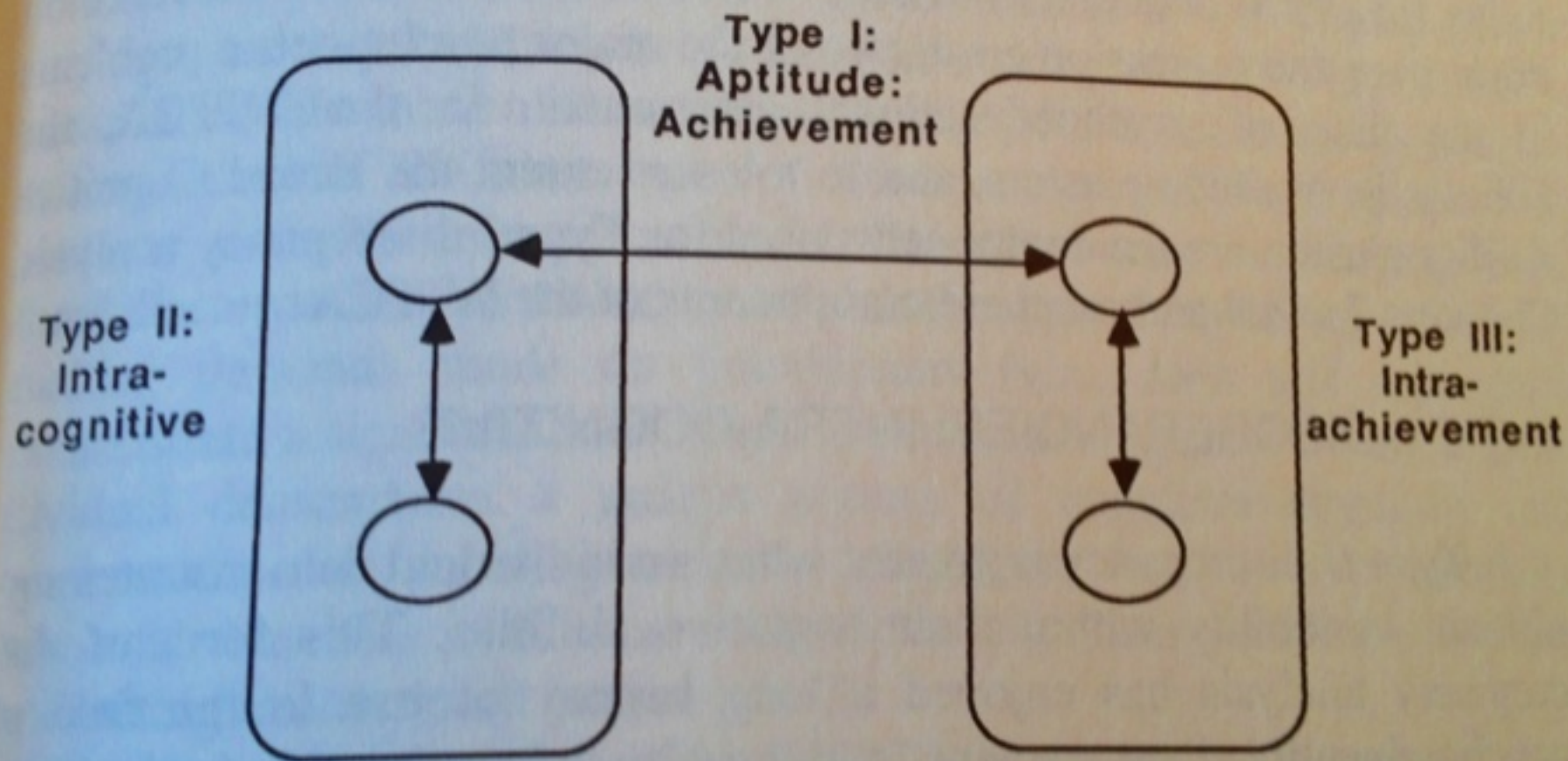
Aptitude clusters is to provide the best predictors of *current* achievement. If a subject obtains a low Antonyms-Synonyms score because of a legitimate language problem, then this person's current achievement expectancies should also be lowered. This expectancy information will be more accurately communicated by the narrower WJTCA Scholastic Aptitude clusters than by any broad-based scores from the WJTCA or other tests. The inability to provide the most *appropriate* services for the individual in the second example because of systems that only acknowledge Type I discrepancies *reflects a problem with the systems and not the measurement instrument*. Examination of the Type I and II discrepancies in the second scenario indicates that this individual's reading expectancies should be lowered and that he or she may benefit from language training. The use of broad-based measures (e.g., WJTCA Broad Cognitive Ability cluster), or aptitude measures that demonstrate little relationship to achievement (e.g., the Wechsler Performance Scale), may result in well intentioned but misdirected intervention efforts and/or the formation of expectations that could prove damaging.

The combination of both scenarios should demonstrate the advantage of using aptitude measures comprised of subtests most directly related to the curriculum area of concern. Because of their differential weighting system, the WJTCA Scholastic Aptitude clusters should provide some of the best curriculum-specific expectancy information available in the field of psychoeducational assessment. In contrast, broad-based measures contain subtests that introduce extraneous "noise" into expectancy formulation (i.e., some of the abilities measured by certain subtests demonstrate little relationship to the academic area of concern). The use of broad-based measures can result in the masking of Type I discrepancies (viz., first scenario), or may suggest Type I discrepancies where none really exist (viz., second scenario). Although the Broad Cognitive Ability and Scholastic Aptitude clusters may demonstrate similar correlations with achievement, the broad-based measures may frequently confuse Type I and Type II discrepancies.



**Figure 1.1.** Assessment structure of the *Woodcock-Johnson Psycho-Educational Battery*. (From Hessler, G. (1982). *Use and interpretation of the Woodcock-Johnson Psycho-Educational Battery*. Allen, Texas: DLM/Teaching Resources, p. 7. Copyright 1982 by DLM/Teaching Resources. Reprinted by permission.)



**COGNITIVE  
SKILLS****ACHIEVEMENT  
SKILLS**

**Figure 1.3.** The underlying WJ decision-making discrepancy model. (From Woodcock, R. (1984b). A response to some questions raised about the Woodcock-Johnson: Efficacy of the aptitude clusters. *School Psychology Review*, 13, 355–362. Reprinted by permission.)

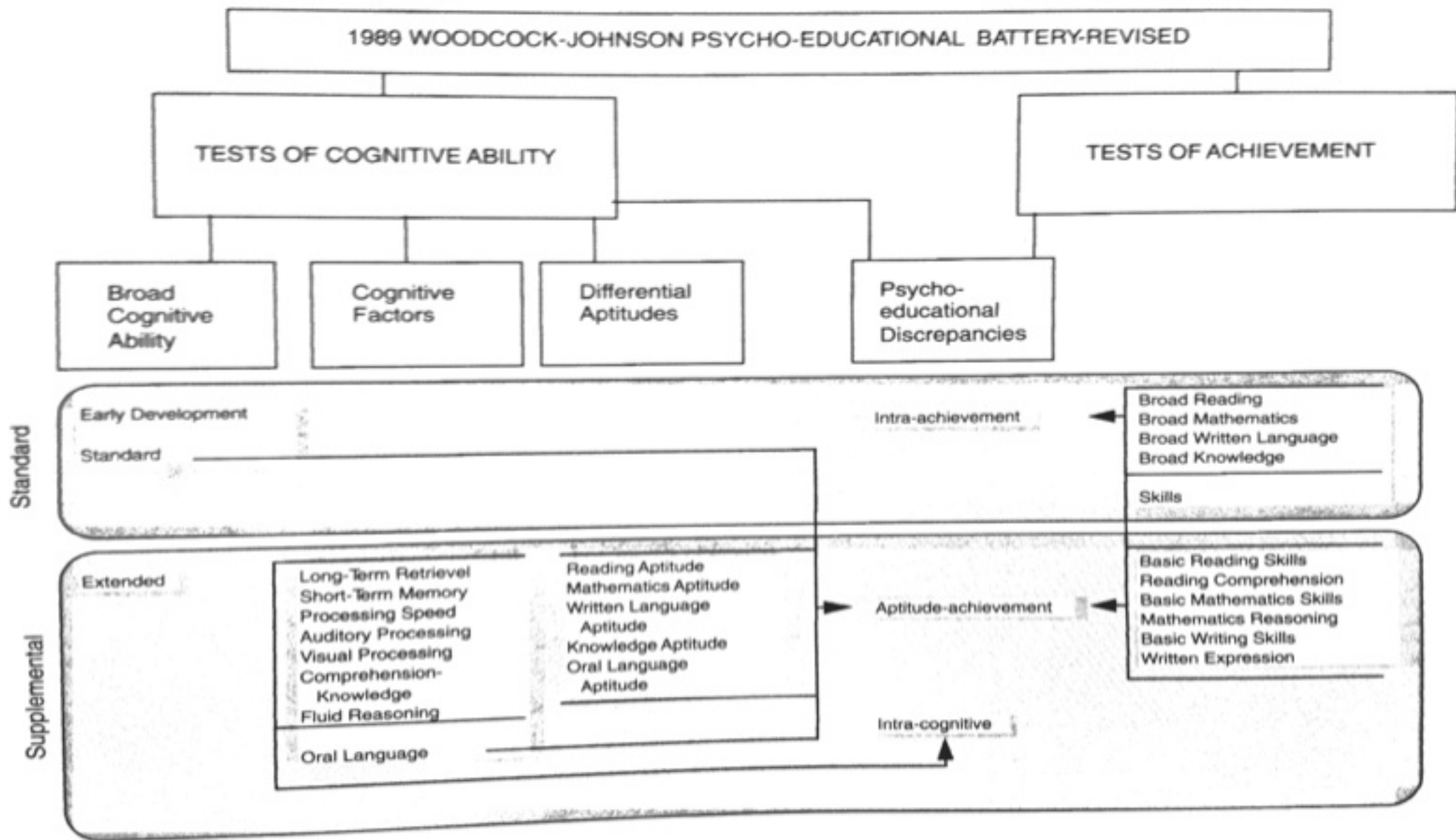


CLINICAL  
INTERPRETATION  
OF THE  
WOODCOCK-JOHNSON  
TESTS OF COGNITIVE  
ABILITY—REVISED



KEVIN S. MCGREW

A LONGWOOD PROFESSIONAL BOOK



**FIGURE 1-2** Assessment structure of Woodcock-Johnson Psycho-Educational Battery—Revised



## THE WJTCA-R DIFFERENTIAL APTITUDE CLUSTERS

p. 143

The previous four chapters dealt extensively with information related to Type II or intra-cognitive interpretation of the WJTCA-R. In contrast, the WJTCA-R differential aptitude clusters (i.e., Scholastic and Oral Language Aptitude clusters) are the cornerstone for Type I or aptitude/achievement discrepancy analysis in the WJ-R pragmatic decision-making model. In many respects the Scholastic Aptitude clusters represented the philosophical heart and soul of the 1977 WJ battery (McGrew, 1986). However, a review of the literature suggests that the true value of the Scholastic Aptitude clusters was not appreciated in the original WJ battery. This chapter puts the WJTCA-R differential aptitude clusters in proper perspective.

**TABLE 7-1 Test Composition and Weighting Comparisons of WJTCA and WJTCA-R Differential Aptitude Clusters**

Tests	Gf-Gc Factor	Scholastic Aptitudes									
		Reading		Mathematics		Written Language		Knowledge		Oral Language Aptitude	
		WJTCA	WJTCA-R	WJTCA	WJTCA-R	WJTCA	WJTCA-R	WJTCA	WJTCA-R	WJTCA	WJTCA-R
Oral Vocabulary	(Gc)	.42	.25	.39	.25	.29	.25	.46	—	—	—
Visual Matching	(Gs)	—	.25	.26	.25	.15	.25	—	—	—	—
Sound Blending	(Ga)	.17	.25	—	—	—	.25	—	.25	—	—
Concept Formation	(Gf)	—	—	.08	.25	—	—	—	.25	—	.25
Memory for Sentences	(Gsm)	—	.25	—	—	—	—	.10	.25	—	—
Analysis-Synthesis	(Gf)	—	—	.27	.25	—	—	—	—	—	—
Visual-Auditory Learning	(Glr)	.23	—	—	—	—	.25	—	—	—	—
Numbers Reversed	(Gsm)	—	—	—	—	—	—	—	—	—	.25
Visual Closure	(Gv)	—	—	—	—	.08	—	—	.25	—	—
Picture Recognition	(Gv)	—	—	—	—	—	—	—	—	—	.25
Sound Patterns	(Ga)	—	—	—	—	—	—	—	—	—	.25
Verbal Analogies	(Gf/Gc)	.18	—	—	—	—	—	.16	—	—	—
Quantitative Concepts	(Gq)	—	—	—	—	.47	—	.28	—	—	—

Note: Boxes indicate common tests in respective clusters across WJTCA and WJTCA-R. Test names are from WJTCA-R. Three of the tests have different names in WJTCA (Oral Vocabulary = Antonyms-Synonyms; Sound Blending = Blending; Verbal Analogies = Analogies).



222 By using the lower Reading Aptitude score, a clinician would conclude that this person's problem is not of the aptitude-achievement discrepancy variety warranting remedial academic services, but an intra-cognitive discrepancy possibly warranting language training. The problem this may create for some practitioners is that Type II discrepancies are not often recognized as a legitimate basis for recommending special services.

Frequently, eligibility for services is dictated solely by Type I or aptitude-achievement discrepancy procedures (Mather, 1993). In such rigid systems the person in this example would probably not receive any services due to the lack of an aptitude-achievement discrepancy. This point can be a major concern to practitioners, who may confuse the problem of inflexible and rigid criteria with a problem with the WJTCAR Scholastic Aptitude clusters. The problem in this example is not the WJTCAR Reading Aptitude cluster because it produces a lower score, but with administrative or legal systems that force all decisions into a Type I or aptitude-achievement discrepancy model. *The problem is not with the instrument (which is based on sound psychological principles and research) but with how practitioners are forced to use an instrument.* This point is similar to Kamphaus' (1993) discussion of how composite full scale intelligence test scores often cloud accurate diagnosis when practitioners are forced to use them within inflexible regulatory frameworks that do not allow the use of well-grounded professional judgment.

The pressures surrounding regulation-driven professional practice can result in practitioners preferring broad-based aptitude measures that provide higher aptitude results (since they contain abilities unrelated to the area of concern) in order to "get the kid qualified." Although the motives may be admirable, even this process can do more harm than good. In the



example, using the broad-based scores as the aptitude estimate, or, even more inappropriately, using only the WISC-III Performance Scale under the rationalization that it is not affected by the individual's comprehension-knowledge problem may result in a large enough Type I discrepancy that qualifies the person for services. However, these broad-based aptitude estimates suggest a "gap" between ability and achievement that warrants remedial services. In reality this person may be doing the best he or she currently can based on those abilities most directly related to reading. Remedial reading services may be vended, while in reality this individual needs more help in the development of basic comprehension-knowledge abilities. Thus, programming efforts may be misdirected and inappropriately high expectations formed.

p. 223

### Summary Comments

The purpose of the WJTCA-R differential aptitude clusters is to provide predictions of *current* levels of achievement. If a person obtains low scores on individual tests that measure cognitive abilities related to a specific achievement area and these tests are included in the aptitude cluster, then the person's current achievement expectancies should also be lowered. This expectancy information will be more accurately communicated by the narrower WJTCA-R differential aptitude clusters than by any broad-based scores from the WJTCA-R or other tests. The inability to provide the most appropriate services for an individual because of systems that only acknowledge Type I discrepancies reflects a problem with the systems and not the measurement instrument. After all, if you give a monkey a Stradivarius violin and you get bad music, you don't blame the violin.

The clinical utility example demonstrates the advantage of using aptitude measures comprising tests most directly related to the curriculum area of concern. It is because of their differential content that the WJTCA-R differential aptitude clusters provide the best curriculum-specific expectancy information available in the field of psychoeducational assessment. In contrast, broad-based measures contain tests that introduce extraneous "noise" into expectancy formulation (i.e., some of the abilities measured by certain tests demonstrate little relationship to the academic area of concern). The use of broad-based measures can result in the masking of Type I discrepancies or may suggest Type I discrepancies where none really exist. Broad-based aptitude measures may frequently confound Type I and Type II discrepancies.