1984

A RESPONSE TO SOME QUESTIONS RAISED ABOUT THE *WOODCOCK-JOHNSON* II. EFFICACY OF THE APTITUDE CLUSTERS

Richard W. Woodcock
Measurement/Learning/Consultants

ABSTRACT

A three-type psychoeducational discrepancy model was described and related to the kinds of discrepancy information provided by the *Woodcock-Johnson Psycho-Educational Battery*. The following rationale was presented for using the Scholastic Aptitude clusters rather than the Broad Cognitive Ability cluster in the evaluation of aptitude-achievement discrepancies: The aptitude clusters are not confounded with achievement; they can reduce testing time; they provide differential expectancy information based on norms; and they provide higher clinical validity. The inappropriateness of comparing the norm-referenced discrepancies obtained from the *Woodcock-Johnson* with the results of procedures that do not correct for regression error was explained.

Research reviews to date have often treated the Broad Cognitive Ability score of the *Woodcock-Johnson Psycho-Educational Battery* (WJ; Woodcock & Johnson, 1977) as a measure of expected achievement for comparison with a subject's actual achievement. However, the authors of the WJ have recommended using the set of four scholastic aptitude clusters. As with other broad-based tests of intelligence, the Broad Cognitive Ability score is more appropriately used for predicting intellectual performance *on the average* across life situations requiring various cognitive skills. Cummings and Moscato (1984a) question the value of using the four scholastic aptitude scores rather than the single Broad Cognitive Ability score.[1] This article presents the rationale for using the specific aptitude clusters rather than the Broad Cognitive Ability score to evaluate aptitude-achievement discrepancies.

Background

The decision-making model underlying the WJ differs markedly from models underlying the design of tests such as the *Illinois Test of Psycholinguistic Abilities* (Kirk, McCarthy, and Kirk, 1968) and the *Kaufman Assessment Battery for Children (K-ABC;* Kaufman & Kaufman, 1983). The developers of those tests first identified a model they believe represents important dimensions of cognitive functioning. Next, subtests to measure those dimensions were developed. Finally, validity studies were conducted to determine whether the tests provide useful information. In contrast, the WJ is based on a pragmatic psychoeducational model. The basic question to the design was: "What decisions to practitioners make following a psychoeducational assessment, and what information is needed to make these decisions?"

For the WJ Tests of Cognitive Ability (WJTCA), three major objectives were identified. First and most significant was to develop a method to provide information on an individual's *presently expected achievement* in several areas of school learning. During the development of the WJ, four specialized measures of intelligence, called scholastic aptitudes, were created to provide differentiated estimates of expected achievement in academic areas (reading, mathematics, written language, and knowledge). Subtests were developed that assess various cognitive skills related to achievement but do not require the subject to use those achievement skills.

The second objective was to create a broad-based measure of intellectual ability that could be used in determinations such as giftedness or mental retardation. To develop this heterogenous measure of cognitive ability, subtests were selected that included tasks of lower-order processing (perception and discrimination), higher-order processing (reasoning and comprehension), auditory and visual processing, memory tasks, and controlled learning tasks.

Correspondence and requests for reprints should be addressed to Richard W. Woodcock, Measurement/Learning/Consultants, P. O. Box 161, Tolovana Park, Oregon 97145.

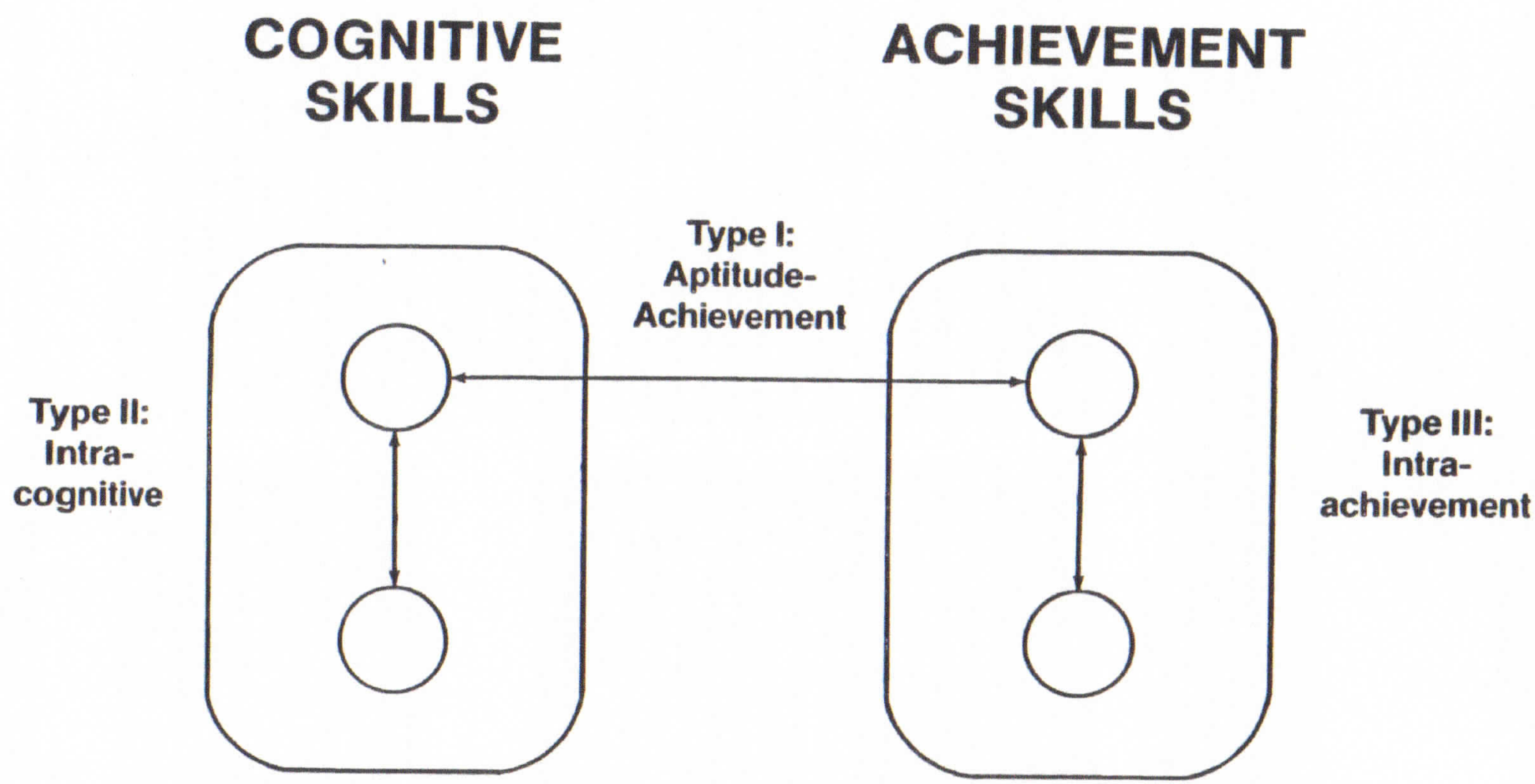

FIGURE 1
Three types of discrepancy which may be evaluated during a psychoeducational assessment.

The third objective was to provide information regarding special intellectual abilities that may influence school learning and certain other kinds of performance. Four special cognitive ability clusters were developed: verbal ability, reasoning, visual-perceptual speed, and memory. This development was aided by, but not limited to, the application of cluster and factor analysis procedures to the final set of 12 cognitive subtests.

Types of discrepancy. The WJ was designed to provide information regarding three types of discrepancy. The psychoeducational discrepancy model underlying the WJ is illustrated in Figure 1. This model distinguishes among three types of discrepancy that may be evaluated during the course of a psychoeducational assessment: Type I, an aptitude-achievement discrepancy; Type II, an intra-cognitive discrepancy; and Type III, an intra-achievement discrepancy.

The Achievement-Aptitude Profile of the WJ provides three ways to evaluate the presence of a Type I (aptitude-achievement) discrepancy. First, the actual achievement grade score may be compared with the "expected" grade score; second, it can be noted whether the subject falls in the lower five percent of the population in respect to size of discrepancy; and third, the subject's Relative Performance Index may be evaluated. This index reports the subject's accuracy on achievement tasks expected to be performed with 90 percent success by others with that aptitude score and the same number of years in school. If an individual's actual achievement falls below some criterion value (established by the school system) relative to the expected achievement score, the individual is identified as having an aptitude-achievement discrepancy.

Whether using the WJ or any other test, the determination of a true Type I discrepancy is predicated on using the most relevant set of cognitive skills for a specific area of achievement and using cognitive tests that do not require the performance of that kind of achievement.

A Type II (intra-cognitive) discrepancy is present within individuals who have specific cognitive deficits such as delayed language or visual-perceptual disorders. In the WJTCA a Type II discrepancy is detected by evaluating patterns among the cognitive clusters on the Percentile Rank Profile and among the cognitive subtests on the Subtest Profile. Hessler (1982) and McGrew (1984) have presented detailed analyses of each subtest's content and have suggested procedures for evaluating intra-cognitive discrepancies in the WJTCA. (While the comprehensive nature of the WJTCA may direct attention to the presence of a cognitive deficit, supplemental information and tests may be needed to adequately diagnose and verify Type II discrepancies.)

A Type III (intra-achievement) discrepancy is determined by using the achievement Percentile Rank Profile, the Instructional Implications Profile, and the achievement Subtest Profile. Hessler (1982) has also provided information to aid in evaluating intra-achievement discrepancies. Identification of strengths and weaknesses in achievement is generally for the purpose of instructional planning, although at least one state requires an intra-achievement discrepancy for placement in learning disability (LD) programs (State of Louisiana, Department of Education, 1981).

The WJ aptitude clusters. The four scholastic aptitude clusters of the WJ are used to determine the presence of an aptitude-achievement discrepancy. Table 1 identifies the subtests contained in each of the four clusters. The relative weight of each subtest in the cluster is also included in the table. Note the composition of the aptitude clusters and how they differ. For example, the Reading Aptitude cluster includes Visual-Auditory Learning, Blending, Antonyms-Synonyms, and Analogies. Antonyms-Synonyms and Analogies both measure certain language processing skills. Visual-Auditory Learning is a controlled learning task in which unfamiliar (pictorial/abstract) symbols are used to represent a reading vocabulary taught to the subject. Blending is an oral auditory-processing task closely related to certain word-attack processes in reading.

In contrast, the Mathematics Aptitude cluster contains two controlled learning tasks highly similar to certain processes in mathematics. The Analysis-Synthesis task is a miniature mathematics system with a set of equivalencies that must be used to solve problems

TABLE 1
Composition of the WJTCA Scholastic Aptitude Clusters

	Subtest	Rdg Apt	Math Apt	Wr Lang Apt	Know Apt
1	Picture Vocabulary	—	—	—	—
2	Spatial Relations	—	—	—	—
3	Memory for Sentences	—	—	—	.10
4	Visual-Auditory Learning	.23	—	—	—
5	Blending	.17	—	—	—
6	Quantitative Concepts	—	—	.47	.28
7	Visual Matching	—	.26	.15	—
8	Antonyms-Synonyms	.42	.39	.29	.46
9	Analysis-Synthesis	—	.27	—	—
10	Numbers Reversed	—	—	.08	—
11	Concept Formation	—	.08	—	—
12	Analogies	.18	—	—	.16

in number combination (number values are represented with colored squares instead of numerals). Concept Formation is a rule-learning procedure, closely related to certain operations in logic and set theory. Visual Matching measures a subject's ability to rapidly identify the two identical numerals in a row of numerals. The remaining subtest, Antonyms-Synonyms, appears in all four scholastic aptitude clusters with varying weights.

Rationale for the Aptitude Clusters

Cummings and Moscato (1984a) demonstrate comparable correlations between Broad Cognitive Ability and each of the four areas of achievement and between the four scholastic aptitude measures and their respective areas of achievement. Based on these correlations, they are correct in suggesting that the single Broad Cognitive Ability score could serve equally well as a predictor for all four academic areas. However, four reasons exist for considering the scholastic aptitude clusters advantageous: (a) they are not confounded with achievement; (b) they can reduce testing time; (c) they provide differential expectancy information based on norms; and (d) they provide higher clinical validity.

Avoidance of confounding with achievement. The Broad Cognitive Ability scale contains certain subtests judged to require achievement skill. The most obvious is the Quantitative Concepts subtest, which measures knowledge of mathematical concepts and vocabulary. Although no calculation skills are required, this subtest measures one aspect of mathematics achievement and consequently is an inappropriate component of a measure of expected achievement in mathematics. Several of the Analogies subtest items have mathematics content, so that subtest was also precluded from the Mathematics Aptitude cluster. The Broad Cognitive Ability score includes both of these subtests. If that score is used as an index of expected achievement in mathematics, it is confounded with mathematics achievement. (The WISC-R has the same confounding problem because it contains an Arithmetic subtest.) Each of the subtests contained in a WJ scholastic aptitude cluster is related to that specific area of achievement but does not require the subject to perform those achievement skills. Stepwise multiple regression analyses (based on 3142 subjects) were used to determine which of the 12 subtests were to be included in each aptitude cluster (Woodcock, 1978). As a result, redundancy is minimized within each set of aptitude subtests and the validity is maximized.

Economy of testing time. Results from the stepwise multiple regression analyses indicated no significant improvement when using aptitude clusters consisting of more than three or four subtests. The addition of other subtests requires more testing time and may actually reduce the quality of the prediction because extraneous information becomes incorporated into the expected achievement score. An average of five minutes is required for administering each WJ subtest. If the examiner's only desire is to obtain an expected achievement score in reading, just four subtests need be administered (20 minutes testing time); complete administration of the WJTCA requires approximately 60 minutes.

Differential expectancy information based on norms. Even if the Broad Cognitive Ability cluster were used as the single measure of aptitude, it still would be desirable to provide a different set of interpretation tables for each area of achievement. The distribution of discrepancies for mathematics is not the same as the distribution of discrepancies for reading or for the other areas of achievement. Since the WJ cognitive tests have a common norm base with the WJ achievement tests, *discrepancy norms* are provided for each area of achievement. The subject's aptitude-achievement discrepancy is based on the average achievement of others in the norming sample having the same aptitude score and grade or years in school. Since expected achievement is a function of both aptitude and exposure to instruction, it is suggested that practitioners use number of years in school if the subject has been retained (Hessler, 1982). In addition, the subject's discrepancy is compared with the *range of discrepancy* demonstrated by the middle 90 percent of such others in the norming sample. No other test has these features. With a test like the WISC-R or the K-ABC, the user is limited to using a point predictor (the deviation-IQ score). Furthermore, this score provides only a single estimate of expectation for all areas of achievement.

Some writers have indicated that the four scholastic aptitude clusters lack adequate specificity to provide differentiated expectancy scores because the Antonyms-Synonyms subtest appears in each of the four clusters. (Six other subtests are only used once and three subtests are used twice.) Ysseldyke, Algozzine, and Shinn (1981) state that "a poor performance on (Antonyms-Synonyms) would result in poor Reading Aptitude, poor Mathematics Aptitude, poor Written Language Aptitude, and poor Knowledge Aptitude" (p. 249). However, they do not point out that this effect will also occur if a poor score is obtained on the Vocabulary subtest of the WISC-R, especially if the Verbal Scale is used. In either instance, poor oral vocabulary limits school achievement. Historically, measures of vocabulary have provided powerful predictors of achievement across all areas of achievement at all levels. The impact, empirically determined, of poor oral vocabulary on the different scores is exactly the expected achievement needed by the practitioner.

Specificity does exist in the WJ aptitude clusters via differing subtest inclusions and weights, and the separately calculated discrepancy norms for each aptitude cluster. In contrast, the user of the WISC—R uses the same set of subtests with the same weighting for every prediction. Unless one believes that cognitive ability is a single unitary trait with the same predictive relationship to various psychoeducational capabilities, a set of differential predictors will maximize the information available from a broad-based intelligence battery.

Cummings and Moscato (1984a) and Thompson and Brassard (1984) note that the sheer number of scores and the concomitant scoring procedures place a heavy load on examiners. A microcomputer scoring program (Hauger, 1984) is available that will significantly reduce both scoring time and the problem of clerical error.

Clinical implications. The mix of cognitive skills included in each of the four scholastic aptitude clusters represents the best match with those achievement skills that could be obtained from the WJ cognitive subtests. For individuals with specific deficits, the aptitude clusters are not interchangeable with each other or with Broad Cognitive Ability, even though comparable statistical correlations are observed in group data.

A broad-based measure of aptitude such as the WJ Broad Cognitive Ability cluster or the WISC-R Full Scale includes cognitive tasks that will influence the score but have little relationship to the aptitude of concern. For example, a person may be significantly weak in the cognitive skills most relevant to an area of achievement. If this person performs highly on the other subtests in a broad-based measure his or her *aptitude* for that area of achievement is overestimated by the broad-based score. Although a significant discrepancy between aptitude and achievement may be observed, it is really a false positive selection error; the diagnostic picture has been confused by the extraneous information included in the broad-based aptitude score. Conversely, if a person is relatively weak in certain cognitive skills not relevant to the achievement area, the broad-based score will underestimate his or her aptitude for that specific area of achievement. As a result, the examiner will fail to detect the presence of a significant aptitude-achievement discrepancy and a false negative selection error will be made.

Many diagnostic procedures fail to distinguish between, and in fact confound, Type I and II discrepancies. As a result, diagnostic classification and educational programming may miss the real problem. For example, a language disordered or language delayed child truly has a lower expectancy for reading achievement than a child who is not language handicapped. This child should not be identified as having an aptitude-achievement discrepancy if reading achievement is *as high as can be expected* considering the language deficit. Such a child has an intra-cognitive discrepancy, not an aptitude-achievement discrepancy, and his or her first programming needs are more likely to be language training than remedial reading.

Research on the Aptitude Clusters

Algozzine, Ysseldyke, and Shinn (1982) have compared the results of the WJ aptitude-achievement procedure with a second procedure, common in the field, involving computation of a standard score difference between the WISC-R and the *Peabody Individual Achievement Test* (Dunn & Markwardt, 1970). The authors state that "the school identification decisions, based on application of a severe discrepancy on the Woodcock-Johnson Psycho-Educational Battery, did not correlate with decisions *based on application of the federal definition* (italics added)" (p. 299). Two points must be brought to the attention of the reader of that article. First, Table 4 and the discussion in the Algozzine et al. study should not be perceived as an evaluation of the validity of decisions based on the WJ compared with the federal guidelines but simply as the difference between two alternate procedures for estimating discrepancy. Second, their results indicated that the correlation between the two procedures studied was essentially zero. This finding is not surprising, since they did not correct their discrepancy estimates for regression error and consequently their classifications of students as LD versus non-LD are inconsistent across the range of intelligence of their sample.

As described earlier, the WJ discrepancies are norms. Discrepancy estimates obtained from other tests that do not provide discrepancy norms must be corrected for regression error. The correction for regression, based on the correlation between the aptitude and achievement tests, provides an estimate of what would be obtained if discrepancy norms were available for a pair of tests and certain statistical assumptions could be met. If this procedure is not performed, the obtained discrepancies not only are inaccurate but also cannot be compared with the results of a procedure that is free of regression error.

Failure to correct for regression error is probably the most serious procedural error committed today by practitioners and researchers. The result is to overidentify high aptitude students as LD. Low aptitude students are underidentified as LD and overidentified as slow-learners, or even as mentally retarded. The crucial point is that unless the aptitude score is at or near 100, *the expected achievement score for an individual is not the same as his or her obtained aptitude score.* Since aptitude and achievement tests do not correlate perfectly, the aptitude score must be adjusted for regression by some formula such as the following:

$$Y' = r_{xy}\,(X-100)+100$$

where Y' is the expected score after correction for regression, r_{xy} is the correlation between the two measures, and X is the aptitude test score. The aptitude and achievement scores are assumed to be standard scores with a mean of 100 and the same standard deviation. (More complex corrections for regression have been suggested in the literature, but a discussion of these is beyond the scope of this article.)

If a student has a deviation-IQ of 130 and the correlation between an aptitude test and an achievement test is .65, the adjusted aptitude score (expected achievement) is 120. If the school system's required discrepancy for classification is 15 points, then the criterion reading achievement score is 105 (120 minus 15), not 115 (130 minus 15). Conversely, if a student has a deviation-IQ of 70, the correction for regression operates in the other direction. In that case, the expected achievement score is

80, not 70, resulting in a criterion reading score of 65 (80 minus 15), not 55 (70 minus 15). If the regression correction is not made, a presumed 15-point discrepancy at an IQ of 100 is indeed 15 points, but it is 5 points at an IQ of 130 and 25 points at an IQ of 70.

The standard deviation of actual achievement scores about the expected score for a group of individuals having the same aptitude score is called the standard error of estimate and is calculated by

$$SE_{Est}=15\sqrt{1-r_{xy}^2}$$

It is the SE_{Est}, not the standard deviation (SD), that should be used to evaluate the educational significance of a discrepancy. The higher the correlation between a measure of aptitude and a measure of achievement, the smaller the SE_{Est}. Aptitude tests that have higher correlations with achievement, and thus smaller SE_{Est} values, require a smaller difference between the two scores to be educationally significant—that is, to identify the same number of subjects as having a significant aptitude-achievement discrepancy. Furthermore, fewer false positive and false negative selection errors occur when using aptitude tests that have higher correlations with achievement.

Correcting for regression is a simple procedure requiring an easy calculation or a prepared table of corrected scores. It is essential for practitioners to conduct such a correlation before determining whether a discrepancy is present, unless the pair of tests being used provides discrepancy norms.

Summary

The design of the WJ is based on a psychoeducational decision-making model. The major design objective of the battery was to provide information about, and distinguish among, three types of discrepancy: Type I, aptitude-achievement; Type II, intra-cognitive; and Type III, intra-achievement. Further objectives of the WJ cognitive ability tests were to provide a broad-based measure of intelligence for use in determinations such as giftedness and mental retardation and to provide information about specialized areas of intellectual functioning, including verbal ability, reasoning, visual-perceptual speed, and memory.

Some reviewers have questioned the efficacy of using the scholastic aptitude scores rather than the Broad Cognitive Ability score for obtaining estimates of expected achievement. Based on the correlations observed in group data, about the same results can be expected in groups of nonhandicapped subjects when making standard score comparisons with either measure. However, when the WJ is used with clinical populations, certain characteristics and interpretive features of the scholastic aptitude clusters may become quite important with respect to obtaining valid estimates of expected achievement and distinguishing between aptitude-achievement and intra-cognitive discrepancies. Four reasons exist for considering use of the scholastic aptitude clusters superior to the Broad Cognitive Ability cluster for estimating the presence of an aptitude-achievement discrepancy: the aptitude clusters are not confounded with achievement; their use can reduce testing time; they provide differential expectancy information based on norms; and they provide higher clinical validity.

The WJ provides discrepancy norms to aid in the determination of an aptitude-achievement discrepancy. Discrepancy estimates that are not based on norms must be corrected for regression error. This article provided a brief description of a procedure to correct for regression error. If such a procedure is not performed, the obtained discrepancies not only are inaccurate for psychoeducational purposes but also cannot be compared with the results of the WJ procedure, which is without regression error. □

Footnote

[1]Cummings and Moscato (1984b) have suggested that "the developers of the WJPED. . .construct an additional table which would use the Broad Cognitive scale as the predictor" (p. 47). Such information is available from the author or the publisher in conjunction with a table prepared by Woodcock (1982) for correcting regression effects if using standard scores from either the Scholastic Aptitude clusters or the Broad Cognitive Ability cluster.

References

Algozzine, B., Ysseldyke, J., & Shinn, M. (1982). Identifying children with learning disabilities: When is a discrepancy severe? *Journal of School Psychology, 20,* 299-305.

Cummings, J. A., & Moscato, E. M. (1984a). Research on the Woodcock-Johnson Psycho-Educational Battery: Implications for practice and future investigations. *School Psychology Review, 13,* 33-40.

Cummings, J. A., & Moscato, E. M. (1984b). Reply to Thompson and Brassard. *School Psychology Review, 13,* 45-48.

Dunn, L. M., & Markwardt, F. C. (1970). *Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Hauger, J. (1984). *Compuscore: Woodcock-Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.

Hessler, G. L. (1982). *Use and Interpretation of the Woodcock-Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children.* Circle Pines, MN: American Guidance Service.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1968). *Illinois Test of Psycholinguistic Abilities, Revised Edition.* Urbana: University of Illinois Press.

McGrew, K. S. (1984). Normative based guides for subtest profile interpretation of the Woodcock-Johnson Tests of Cognitive Ability. *Journal of Psychoeducational Assessment, 2,* 141-148.

State of Louisiana (1981). *Pupil Appraisal Handbook* (Bulletin 1508). Baton Rouge, LA: Department of Education.

Thompson, P. L., & Brassard, M. R. (1984). Cummings and Moscato soft on Woodcock-Johnson. *School Psychology Review, 13,* 41-44.

Woodcock, R. W. (1978). *Development and standardation of the Woodcock-Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W. (1982). *Woodcock-Johnson Psycho-Educational Battery: Standard Scores.* Unpublished manuscript.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery.* Allen, TX: DLM Teaching Resources.

Ysseldyke, J., Algozzine, B., & Shinn, M. (1981). Validity of the Woodcock-Johnson Psycho-Educational Battery for learning disabled youngsters. *Learning Disabilities Quarterly, 4,* 244-249.