Journal of Psychoeducational Assessment
1990, *8,* 231-258

THEORETICAL FOUNDATIONS OF THE WJ-R MEASURES OF COGNITIVE ABILITY

Richard W. Woodcock
University of Southern California

The WJ-R provides a wide age range and comprehensive set of cognitive measures. A major interpretation feature of the battery is the provision of eight factor scores, each based on two measures. The theoretical basis for the tests is founded in *Gf-Gc* theory, work that is often associated with Raymond Cattell and John Horn, although other scholars think and write about intelligence in a similar vein. This report is a brief review of the theory followed by the results of several factor analytic studies. Nine data sets drawn from the 1977 and 1989 norming and concurrent validity studies have been analyzed. In the concurrent studies, the WJ-R cognitive tests were administered in conjunction with other major batteries including the K-ABC, the SB-IV, the WISC-R, and the WAIS-R. Altogether, 15 sets of exploratory and confirmatory factor analyses that included a total of 68 variables were completed. The results of all studies provide support for the WJ-R eight-factor model of *Gf-Gc* theory. As a byproduct of the procedure, comparative information was observed for the congruence of other cognitive batteries to the *Gf-Gc* theory. It is suggested that the other cognitive batteries often have been underfactored, which has led to misinterpretation of their factorial structures. This report demonstrates the need for factor analytic studies in which the set of variables is not constrained to the limited set of subtests that have been published together as a battery. It is indicated that the set of variables to be included in a factor study must include enough breadth and depth of markers to ensure that the presence of all major factor effects can be identified.

The Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R; Woodcock & Johnson, 1989) is a wide-range comprehensive set of individually administered tests (subsequently referred to as "subtests" to provide uniformity in terminology when we are comparing cognitive batteries) for measuring cognitive ability, scholastic aptitude, and achievement. The WJ-R is composed of two major parts: the Woodcock-Johnson Tests of Cognitive Ability (WJ-R COG) and the Woodcock-Johnson tests of Achievement (WJ-R ACH). The subjects in the normative sample were administered both cognitive and achievement subtests. This allows direct comparisons among a subject's scores in cognitive ability, scholastic aptitude, and achievement with a degree of accuracy not possible when scores from separately normed batteries are compared. The subtests were standardized on a nationally representative sample of 6,359 subjects, from 24 months to 95 years of age. Included in the norming sample are 916 students from institutions of higher education, who produced distinct college and university norms. The interpretation plan for the WJ-R includes a full array of derived scores and profiles for reporting and displaying results. The wide age range and breadth of coverage allow the battery to be used for educational, clinical, or research purposes from the preschool to the geriatric level.

The information provided by the WJ-R is especially appropriate for documenting, and differentiating among, three types of psychoeducational discrepancy (Woodcock, 1984). Figure 1 illustrates the relation among these three types of discrepancy. Type I, an aptitude/achievement discrepancy, reflects the amount of disparity between certain intellectual capabilities of an individual and his or her actual academic performance. When an individual's actual achievement falls below some criterion value (often established by an educational agency) relative to the expected achievement score, the result is identified as an aptitude/achievement discrepancy. Type II, an intra-cognitive discrepancy, is present in individuals who have specific cognitive deficits, such as auditory processing or visual processing deficits. It is a bidirectional comparison; for example, equal interest exists in the subject who has a deficit in auditory processing compared to visual processing and the subject who has a deficit in visual processing compared to auditory processing. Type III, an intra-achievement discrepancy, is present in individuals who have a specific achievement deficit, such as inadequate reading comprehension or spelling skill. It is also a bidirectional comparison.

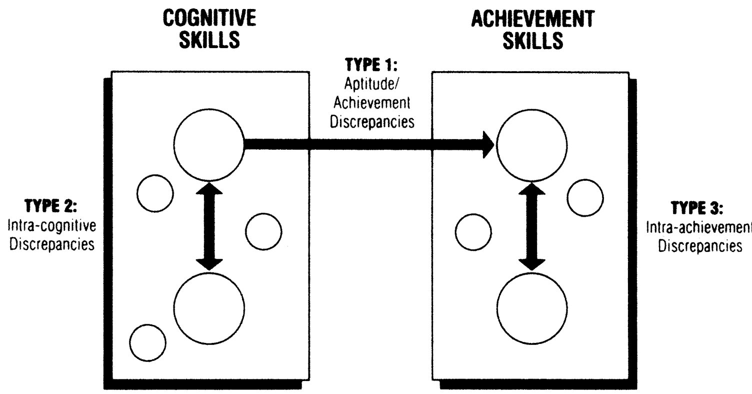

Figure 1
A model of psychoeducational discrepancies.

PURPOSE

The major purpose of this article is to present the theoretical foundation of the WJ-R subtests used to evaluate eight broad cognitive abilities. *Gf-Gc* theory is the conceptualization of intellectual abilities that underlie the WJ-R subtests. The discussion of factor analysis results will be preceded by a statement about certain principles that underlie an approriate factor study. The article concludes with some comments about the interpretation of scores that represent factors and some implications for practice, research, and test development.

There has been no attempt in this report to compare findings to the substantial factor analysis work already accomplished by others. The results reported here represent a descriptive analysis of the data available to the author. More definitive analyses of these data are planned in search of improved confirmatory model fits that will take into account minor loadings of the subtests on other factors. The scope of these analyses is to be broadened by inclusion of data from a number of concurrently administered achievement subtests and by data from a concurrent validity study conducted with 2-year-olds. The final phase of work will address factorial invariance across age.

Gf-Gc THEORY

Effective use of any test battery is enhanced by an understanding of the theory on which the battery is based. The WJ-R COG is a model or operational representation of the *Gf-Gc* (fluid and crystallized abilities) theory of intellectual processing (Cattell, 1963; Horn, 1972, 1976, 1985, 1986, 1988; Horn & Cattell, 1966). *Gf-Gc* theory provides a tool for thinking about and classifying broad cognitive abilities.

A principal contribution of *Gf-Gc* theory of psychoeducational diagnosis is its foundations in the scientific method and data derived from the scientific method. The theory is based on data from many studies that involved both published and unpublished batteries of intellectual tests. Some of these data will be examined in the second half of this report.

Horn (1988) conceptualizes intellectual functioning as a "Milky Way" of human abilities. Just as we do not know precisely how many stars make up the Milky Way, we also do not know precisely how many unique intellectual abilities exist. In the Milky Way, we infer constellations. In the "Milky Way of intellectual abilities," we infer common-factor concepts of constellations that help us describe and understand human intellectual functioning. Factor analysis and structural equation modeling can be used to help us see how a small number of common factors will account for the covariability among performances on different tests.

This common factor conceptualization of human abilities provides a basis for asking scientific questions and collecting data to help answer those questions. To the extent that a battery faithfully represents the different ways people think, it should not be surprising to see the subtests from that battery fit a model based on *Gf-Gc* theory. The combined cognitive and achievement subtests from the WJ-R provide the most comprehensive data set, with enough breadth of measurement and number of markers, to approach an analysis of the entire theory at once.

To date, nine broad intellectual abilities have been identified in the work of Cattell, Horn, and others on *Gf-Gc* theory. The WJ-R measures eight of these nine abilities. (The correct decision speed factor is not represented in the WJ-R and, perhaps, is not represented in any current cognitive test battery.) These nine factors are listed and briefly described in Table 1.

An important aspect of *Gf-Gc* theory is the concept of broad vs. narrow abilities. The narrower abilities that have been established through factor analytic research often are termed "primary mental abilities," "first order factors," or "well-replicated common-factor abilities." Each of the nine broad abilities in the theory could be measured with a variety of tasks, each of which measures a different narrow aspect of the broad ability. For example, *Gc* is the factor that would be measured by a variety of tests such as vocabulary, general information, geology, or even "street wiseness." We would expect scores from the various tests that measure the same broad ability to show a varied pattern of strengths and weaknesses for any given person.

One feature of *Gf-Gc* theory is that it is not based on any particular battery of tests; rather, it has been derived from the statistical and logical analysis of hundreds of data sets that involve several major test batteries and collections of tests. Whenever a collection of cognitive tests with sufficient breadth and depth of variables is appropriately factor analyzed, the results usually fit a model of *Gf-Gc* theory. Thus, the theory is not only empirically based, but eclectic in its data base.

Table 1
Nine Gf-Gc Factors

Factor Name	Symbol	Description
Comprehension-Knowledge	*Gc*	Represents the breadth and depth of knowledge, experience, and sophistication. It includes the comprehension of communication and the types of reasoning based on *previously learned* procedures. *Gc* is not synonymous with overall school achievement, although this is a common misconception. Basic skills tests in reading, writing, and arithmetic usually do not load on this factor.
Fluid Reasoning	*Gf*	Defined as the capability to reason in *novel* situations. Manifested in inductive, deductive, conjunctive, disjunctive, and other forms of reasoning, drawing inferences from relationships, and comprehending implications.
Visual Processing	*Gv*	Involves perceiving and thinking with visual patterns and spatial configurations.
Auditory Processing	*Ga*	Involves the comprehension and synthesis of auditory patterns (but does not require comprehension of language, which is *Gc*). This cognitive ability is important for achievement in music and language development.
Correct Decision Speed	*CDS*	Defined as quickness in providing correct answers to a variety of problems in comprehension, reasoning, and problem solving (correct "snap judgments").
Processing Speed	*Gs*	Involves the ability to perform clerical speed type tasks quickly, particularly when measured under pressure to maintain focused attention.
Short-Term Memory	*Gsm*	Involves the apprehension and use of information within a short period of time. It is often called short-term acquisition-retrieval (SAR) in the literature.
Long-Term Retrieval	*Glr*	Involves the storing of information and the fluency of retrieving it later through association. It is often called tertiary storage-retrieval (TSR) in the literature. *Glr* is not to be confused with the amount of information available, a *Gc* function.
Quantitative Ability	*Gq*	Defined as the capability to comprehend quantitative concepts and relationships and to manipulate numerical symbols.

Gf-Gc theory is still evolving, and the previous description presents a report on the present status of the theory. *Gf-Gc* theory is subject to change, probably by the addition of more factors. As new data are gathered and as scholars apply their skills to the accumulation of evidence, it is likely that a more exact, as well as a broader, explanatory model of human cognitive abilities will evolve.

THE WJ-R SUBTESTS THAT MEASURE COGNITIVE ABILITIES

Detailed information on the WJ-R can be found by referring to the examiner's manual for WJ-R COG (Woodcock & Mather, 1989b), to the examiner's manual for the WJ-R ACH (Woodcock & Mather, 1989a), and to the technical manual for the WJ-R (McGrew, Werder, & Woodcock, 1990).

Twenty-nine of the 39 measures included in the WJ-R are included in some of the factor analytic studies to be reported. Excluded from these analyses are most of the reading and written language subtests, measures that have not yet been studied thoroughly within the *Gf-Gc* theoretical framework. Twenty-one of the subtests are from the WJ-R COG, and 8 subtests are from the WJ-R ACH. A brief description of each of those 29 subtests is provided in Table 2.

Table 2
The 29 WJ-R Subtests Included in These Studies

Subtest	Description	Median reliability
Memory for Names	Measures the ability to learn associations between unfamiliar auditory and visual stimuli (an auditory-visual association task). The task requires learning the names of a series of space creatures.	.91
Memory for Sentences	Measures the ability to remember and repeat simple words, phrases, and sentences presented auditorily by a tape player.	.90
Visual Matching	Measures the ability to quickly locate and circle the two identical numbers in a row of six numbers. The task proceeds in difficulty from single-digit numbers to triple-digit numbers and has a 3-minute time limit.	.78
Incomplete Words	An audio tape subtest that measures auditory closure. After hearing a recorded word that has one or more phonemes missing, the subject names the complete word.	.82
Visual Closure	Measures the ability to name a drawing or picture of a simple object that is altered in one of several ways.	.69
Picture Vocabulary	Measures the ability to name familiar and unfamiliar pictured objects.	.86
Analysis-Synthesis	Measures the ability to analyze the components of an incomplete logic puzzle and to determine and name the missing components.	.90
Visual-Auditory Learning	Measures the ability to associate new visual symbols (rebuses) with familiar words in oral language and to translate a series of symbols presented as a reading passage (a visual-auditory association task).	.92
Memory for Words	Measures the ability to repeat lists of unrelated words in the correct sequence. The words are presented by audio tape.	.78
Cross Out	Measures the ability to quickly scan and compare visual information. The subject must mark the five drawings in a row of 20 drawings that are identical to the first drawing in the row. The subject is given a 3-mnute time limit to complete as many rows of items as possible.	.75
Sound Blending	Measures the ability to integrate and then say whole words after hearing parts (syllables and/or phonemes) of the word. An audio tape is used to present word parts in their proper order for each item. This same subtest is called *Blending* in the 1977 WJ.	.87
Picture Recognition	Measures the ability to recognize a subset of previously presented pictures within a larger set of pictures.	.82
Oral Vocabulary	Measures knowledge of word meaning. In Part A: Synonyms, the subject must say a word similar in meaning to the word presented. In Part B: Antonyms, the subject must say a word that is opposite in meaning to the word presented. This same subtest is called *Antonyms-Synonyms* in the 1977 WJ.	.87
Concept Formation	Measures the ability to identify and state the rule for a concept about a set of colored geometric figures when shown instances and non-instances of the concept.	.93

Table 2 (Continued)

Subtest	Description	Median reliability
Delayed Recall—Memory for Names	Measures the ability to recall (after one to eight days) the space creatures presented in *Memory for Names.*	.91
Delayed Recall—Visual-Auditory Learning	Measures the ability to recall (after one to eight days) the symbols (rebuses) presented in *Visual-Auditory Learning.*	.91
Numbers Reversed	Measures the ability to repeat a series of random numbers backward.	.87
Sound Patterns	Measures the ability to indicate whether pairs of complex sound patterns are the same or different. The patterns may differ in pitch, rhythm, or sound content. The sound patterns are presented by an audio tape.	.88
Spatial Relations	Measures the ability to visually match and combine shapes. The subject must select from a series of shapes, the component parts needed to make a given whole shape. (*Spatial Relations* is an untimed subtest in the WJ-R. It is a 3-minute speed subtest in the 1977 WJ).	.82
Listening Comprehension	Measures the ability to listen to a short tape-recorded passage and to verbally supply the single word missing at the end of the passage.	.81
Verbal Analogies	Measures the ability to complete phrases with words that indicate appropriate analogies. Although the vocabulary remains relatively simple, the relationships among the words become increasingly complex. This same subtest is called *Analogies* in the 1977 WJ.	.90
Calculation	Measures the subject's skill in performing mathematical calculations.	.93
Applied Problems	Measures the subject's skill in analyzing and solving practical problems in mathematics.	.91
Science	Measures the subject's knowledge in various areas of the biological and physical sciences.	.87
Social Studies	Measures the subject's knowledge of history, geography, government, economics, and other aspects of social studies.	.87
Humanities	Measures the subject's knowledge in various areas of art, music, and literature.	.87
Word Attack	Measures the subject's ability in applying phonic and structural analysis skills to the pronunciation of phonically regular nonsense words.	.91
Quantitative Concepts	Measures the subject's knowledge of mathematical concepts and vocabulary. No calculation skills are required.	.86
Writing Fluency	Measures the subject's skill in formulating and writing simple sentences quickly. This subtest has a 7-minute time limit.	.76

For purposes of test interpretation, information from various combinations of subtests is reported as clusters. Each cluster consists of two or more subtests that measure different aspects of the broader ability represented by the cluster. The principle of cluster interpretation has been utilized to minimize the danger in generalizing from the score for a single narrow ability (such as measured by Memory for Names) to a broad multifaceted ability (such as long-term retrieval). Thus, for the purposes of interpreting performance on the WJ-R, long-term retrieval is treated as a function of two narrower aspects of that ability (Memory for Names and Visual-

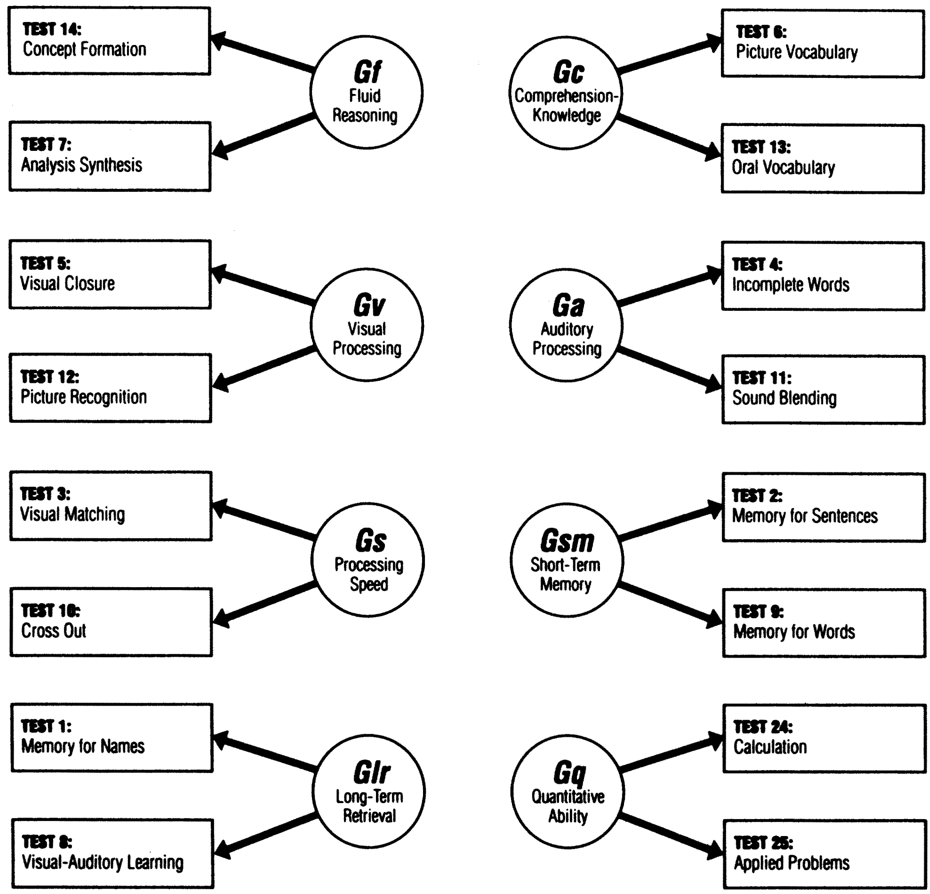

Figure 2
Hypothesized model of the cognitive factor structure underlying the WJ-R (16 primary measures). (Curved double-headed arrows among the latent variables and for the residual variance of the subtests have been omitted.)

Auditory Learning) combined into a single composite score, the Long-Term Retrieval cluster score. Among the WJ-R COG clusters are measures of broad cognitive ability and seven cognitive factors. An eighth cognitive factor (quantitative ability) is obtained by reference to the Broad Mathematics cluster in the WJ-R ACH.

Two subtests for each of the eight factors from the WJ-R have been designated as primary measures. It is the cluster score for each pair of subtests that is used to represent a person's broad ability for that factor as measured by the WJ-R. Seven of the factors are measured within WJ-R COG using subtests 1 through 14. The eighth factor, quantitative ability, is measured by 2 subtests from the WJ-R ACH. Several other subtests from the cognitive and achievement portions of the WJ-R provide supplementary information about the eight factors. Each of the WJ-R subtests is a narrower measure of the broad ability with which it is identified. In certain cases, a WJ-R subtest (e.g., Verbal Analogies) is a mixed or factorially complex measure that taps two broad abilities (in this case, *Gf* and *Gc*).

The seven cognitive factors measured in the WJ-R COG may be evaluated for significant differences by completing the intra-cognitive discrepancy procedure. In this procedure, the score from each of the seven cognitive factors is evaluated against the average performance on the other six factors. The discrepancy is evaluated by a disrepancy percentile rank or in standard error of estimate units. The quantitative ability factor, drawn from the WJ-R ACH, has not been included in the intra-cognitive discrepancy procedure.

Figure 2 presents the hypothesized model of the cognitive factor structure that underlies the 16 primary measures of the WJ-R. This model is a restricted conceptualization of the factor structure for the full WJ-R, which is evaluated by the studies reported in this article. Figure 2 presents a very parsimonious and highly restricted model in which each factor is uniquely defined by two tests. For purposes of visual simplicity, arrows that represent latent factor correlations have been omitted. Similarly, arrows that represent the residual variance of the subtests have been omitted.

SOME COMMENTS ABOUT FACTOR ANALYSIS

All but two of the factor analysis studies reported here are based on confirmatory procedures rather than exploratory. Exploratory factor analysis procedures have their primary value when the factorial structure is unknown and the researcher is attempting to ascertain what that structure may be. Confirmatory factor analysis requires prior knowledge about the expected factorial structure. The purpose of a confirmatory factor analysis is to test hypotheses about an exploratory model believed to underlie the data.

There are three charactertistics of a factor analytic study that make the results more meaningful. First, the data should come from a representative sample of the general population, not from a selected clinical group. Clinical groups have been assembled in the first place because of some unusual behavior or relationships among behaviors. Results from studies with such groups have limited generalizability to the description of cognitive abilities in the population or of the factorial structure of a battery. Second, a breadth of human cognitive abilities should be represented in the factor analytic study, at least to the extent that the various abilities are required to perform any of the tests in the battery. Third, there must be a sufficient number (generally three or more) of reasonably clean measures, or markers, for each of the factors present so that the factor can be identified clearly.

A serious problem exists with many of the factor analytic studies that have been reported on the major cognitive batteries. The variables in those studies routinely have been restricted to only those subtests included within a battery itself. Any single cognitive battery, with the possible exception of the WJ-R, probably does not include enough markers for each embedded factor to allow an appropriate description of the factorial structure of that battery. As a result, factors that are present in a battery are not differentiated or perhaps not even detected. Inappropriate conclusions then may be drawn about the factorial structure of the battery and about the construct validity of the individual subtests.

WJ/WJ-R FACTOR ANALYTIC STUDIES

The following factor analytic results present a range of validity evidence with regard to the model portrayed in Figure 2. Special emphasis is placed on documenting the factorial composition of the 16 tests designated as the primary measures of eight cognitive factors in the WJ-R. Some of this evidence is provided by analyses of the data sets from norming in both 1977 and 1989. Other evidence is provided by analyses of the 1977 and 1989 concurrent validity studies. Analyses have been conducted on nine data sets:

1. *WJ-R norming sample.* Three data sets drawn from the norming sample have been analyzed. The subjects comprise a sample of kindergarten to adult subjects representative of the general population of the United States. The *WJ-R technical manual* (McGrew et al., 1990) includes a detailed description of the 1989 norming sample. The effect of age has been partialled out of each data set through a multiple regression procedure. The resulting residuals were correlated and analyzed by exploratory and confirmatory factor analysis procedures. The three data subsets are as follows:

   A. A sample of 2,261 subjects who were administered the 16 primary measures for the eight factors. The analyses completed on these data include confirmatory and oblique rotation factor analyses.

   B. A sample of 3,063 subjects who were administered the 16 primary factor measures plus 5 supplementary measures from the cognitive battery and 6 supplementary measures from the achievement battery, for 27 measures in all. These data were analyzed by a confirmatory factor analysis.

   C. A sample of 1,425 subjects who were administered the two delayed recall tests in addition to the 27 tests cited above. (These subjects are a subset of the 1B sample.) These data were analyzed by two independently conducted confirmatory factor analyses and by an oblique rotation factor analysis.

2. *WJ norming sample.* A 4,261-subject subset was drawn from the Woodcock-Johnson (1977) norming sample. The subjects comprise a sample of grade one through adult representative of the general population of the United States. (See Woodcock, 1978, for a detailed description of the 1977 norming sample.) The effect for age has been partialled out of the scores by a multiple regression procedure. These data were analyzed by a confirmatory factor analysis.

3. *WJ-R concurrent validity samples.* Further details about the following three samples are provided in McGrew et al. (1990).

   A. Third-grade subjects ($N = 89$) from schools in the Dallas, Texas area were administered selected tests from the WJ-R and the complete WISC-R. This study was directed by Jeanmarie Scarr at DLM (the publisher of the WJ-R). The data were analyzed by two independently conducted confirmatory factor analyses.

   B. Age 9 subjects ($N = 70$) drawn from schools in the Dallas-Fort Worth area were administered the WJ-R, the WISC-R, the K-ABC, and the fourth edition of the Stanford-Binet (SB-IV). This study was conducted under the direction of Sue McCullough and Michael Wiebe at Texas Woman's University. The data were analyzed by two independently conducted confirmatory factor analyses.

C. Age 17 subjects ($N = 53$) drawn from schools in the Dallas-Fort Worth area were administered the WJ-R, the WAIS-R, and the SB-IV. This study also was conducted under the direction of McCullough and Wiebe. The data were analyzed by two independently conducted confirmatory factor analyses.

4. *WJ concurrent validity samples*. Further details about the following two samples are provided in Woodcock (1978).

A. Subjects ($N = 167$) drawn from grades 3 and 5 in Anoka County, Minnesota were administered the 1977 WJ and the WISC-R. Standard scores based on age were analyzed. These data were analyzed by a confirmatory factor analysis.

B. Subjects ($N = 73$) from grade 12 in Anoka County, Minnesota were administered the 1977 WJ and the WAIS. These data were analyzed by a confirmatory factor analysis.

The presentation of results will proceed by first reporting a confirmatory analysis followed by an explanatory oblique analysis of data set 1A (the 16 primary factor measures). This is followed by the results obtained when the 16 primary measures are included in the context of other WJ cognitive and achievement subtests and in the context of variables from the K-ABC, SB-IV, and the Wechsler scales. Confirmatory factor models were estimated with the maximum likelihood fitting function in the LISREL computer program (Jöreskog & Sörbom, 1988).

A more detailed and extended set of exploratory and confirmatory factor studies conducted on WJ-R norming data is reported in McGrew et al. (1990). Those studies include 16-, 21-, 27-, and 29-variable models based on samples of kindergarten to adult subjects. Certain analyses are based on samples across the entire age range, other analyses are based on six age levels. In addition, comparisons to other alternative models (e.g., single *g* factor; verbal/nonverbal; hierarchical *g*) are reported.

Sixteen Variable Models

The factor loading and fit statistics presented in Table 3 are the results obtained from the confirmatory factor analyses of the 16-variable kindergarten through adult sample. The factor loadings for the WJ-R subtests on the respective eight factors are high and positive with the exception of a moderate loading on *Gv* for Picture Recognition. Loadings less than .20 have been omitted from Table 3. Inspection of the fit statistics indicates that the model presented in Table 3 provides a very good fit to the data. These findings provide support for the use of the WJ-R cognitive cluster scores as good indicators of eight *Gf-Gc* factors.

In addition to the confirmatory factor analysis, a traditional exploratory factor analysis of the 16 variables was completed for comparison. The principal axes factoring procedure (R^2 in the diagonals, with iterations) was used to extract eight factors, followed by an oblique rotation of the factors. Table 4 presents the factor pattern matrix for the eight-factor exploratory oblique solution.

A review of Table 4 indicates that seven *Gf-Gc* factors are well defined in the pool of 16 WJ-R subtests. These seven factors are defined by high positive loadings for the two respective WJ-R subtests. The results presented in Table 4 suggest that the *Gv* factor may be defined more weakly than the other seven factors, as evidenced by a single loading for Visual Closure (.903). The Picture Recognition subtest failed to load highly on the *Gv* factor, although it was the second highest loading subtest

Table 3
Sixteen Test Confirmatory Factor Analysis Results for 2,261 Kindergarten to Adult WJ-R Norming Subjects

	LISREL maximum likelihood estimates							
Test	*Glr*	*Gsm*	*Gs*	*Ga*	*Gv*	*Gc*	*Gf*	*Gq*
1. Memory for Names	.672	—	—	—	—	—	—	—
8. Visual-Auditory Learning	.837	—	—	—	—	—	—	—
2. Memory for Sentences	—	.822	—	—	—	.160	—	—
9. Memory for Words	—	.490	—	.235	—	—	—	—
3. Visual Matching	—	—	.880	—	—	—	—	—
10. Cross Out	—	—	.650	—	.200	—	—	—
4. Incomplete Words	—	—	—	.646	—	—	—	—
11. Sound Blending	—	—	—	.752	—	—	—	—
5. Visual Closure	—	—	—	—	.639	—	—	—
12. Picture Recognition	.242	—	—	—	.378	—	—	—
6. Picture Vocabulary	—	—	—	—	—	.771	—	—
13. Oral Vocabulary	—	—	—	—	—	.914	—	—
7. Analysis-Synthesis	—	—	—	—	—	—	.725	—
14. Concept Formation	—	—	—	—	—	—	.746	—
24. Calculation	—	—	—	—	—	—	—	.845
25. Applied Problems	—	—	—	—	—	.259	—	.639

Note.—Chi-square with 71 *df* = 333.93.
Goodness of fit index = .982.
Adjusted goodness of fit index = .965.
Root mean square residual = .019.

Table 4
Sixteen Test Oblique Factor Analysis Results for 2,261 Kindergarten to Adult WJ-R Norming Subjects

	Oblique factor loadings							
Test	*Glr*	*Gsm*	*Gs*	*Ga*	*Gv*	*Gc*	*Gf*	*Gq*
1. Memory for Names	.741	—	—	—	—	—	—	—
8. Visual-Auditory Learning	.658	—	—	—	—	—	—	—
2. Memory for Sentences	—	.616	—	—	—	—	—	—
9. Memory for Words	—	.779	—	—	—	—	—	—
3. Visual Matching	—	—	.679	—	—	—	—	—
10. Cross Out	—	—	.902	—	—	—	—	—
4. Incomplete Words	—	—	—	.478	—	—	—	—
11. Sound Blending	—	—	—	.758	—	—	—	—
5. Visual Closure	—	—	—	—	.903	—	—	—
12. Picture Recognition	.254	—	—	—	—	—	—	—
6. Picture Vocabulary	—	—	—	—	—	.726	—	—
13. Oral Vocabulary	—	—	—	—	—	.590	—	—
7. Analysis-Synthesis	—	—	—	—	—	—	.517	—
14. Concept Formation	—	—	—	—	—	—	.707	—
24. Calculation	—	—	—	—	—	—	—	.743
25. Applied Problems	—	—	—	—	—	.271	.200	.445

Note.—Loadings lower than .200 have been excluded from the table.

on the factor. A comparison of the results of the oblique exploratory and the confirmatory solutions revealed a high degree of similarity. The confirmatory and exploratory analyses both provide empirical support of the presence of eight *Gf-Gc* factors in the WJ-R.

Extended Models

The other data sets to be evaluated include additional WJ subtests and, in certain cases, subtests from other cognitive batteries. A model was fitted to each of these data sets by confirmatory factor analysis procedures. Each of the solutions was obtained independently of the results from the others. To provide a cross-validation on the solutions, four of the major data sets (1C, 3A, 3B, and 3C) were analyzed independently by David Epstein at the University of Southern California.

The results from the 15 sets of solutions for the 9 data sets are consolidated into Table 5. The number of analyses including each subtest is indicated. The median of the factor loadings observed across the several independent analyses is reported for each subtest. Zero factor loadings represent the outcomes where more than half of the fitted models had that variable fixed at zero.

Table 5
Summary of the Factor Analysis Results

Test	Number of analyses	Median loading							
		Glr	*Gsm*	*Gs*	*Ga*	*Gv*	*Gc*	*Gf*	*Gq*
WJ (77)/WJ-R									
Memory for Names	12	.676*	0	0	0	0	0	0	0
Memory For Sentences	15	0	.554*	0	0	0	.335	0	0
Visual Matching	15	0	0	.842*	0	0	0	0	0
Incomplete Words	10	0	0	0	.554*	0	0	0	0
Visual Closure	12	0	0	0	0	.472*	0	0	0
Picture Vocabulary	15	0	0	0	0	0	.751*	0	0
Analysis-Synthesis	15	0	0	0	0	0	0	.586*	0
Visual-Auditory Learning	14	.697*	0	0	0	0	0	0	0
Memory for Words	12	0	.782*	0	0	0	0	0	0
Cross Out	12	0	0	.622*	0	.198	0	0	0
Sound Blending	7	0	0	0	.691*	0	0	0	0
Picture Recognition	12	.248	0	0	0	.378*	0	0	0
Oral Vocabulary	15	0	0	0	0	0	.612*	0	0
Concept Formation	15	0	0	0	0	0	0	.682*	0
Delayed Recall—Memory for Names	3	.834	0	0	0	0	0	0	0
Delayed Recall—Visual-Auditory Learning	3	.350	0	0	0	0	0	.151	0
Numbers Reversed	11	0	.365	0	0	0	0	.370	0
Sound Patterns	4	0	0	0	.260	0	0	.235	0
Spatial Relations (WJ-R)	8	0	0	0	0	.192	0	.401	0
Spatial Relations (WJ)	3	0	0	.609	0	0	0	0	0
Listening Comprehension	8	0	0	0	0	0	.662	0	0
Verbal Analogies	11	0	0	0	0	0	.416	.421	0
Calculation	15	0	0	0	0	0	0	0	.792*
Applied Problems	15	0	0	0	0	0	.170	0	.780*
Science	8	0	0	0	0	0	.796	0	0

Table 5 (Continued)

Test	Number of analyses	Median loading							
		Glr	*Gsm*	*Gs*	*Ga*	*Gv*	*Gc*	*Gf*	*Gq*
Social Studies	8	0	0	0	0	0	.831	0	0
Humanities	8	0	0	0	0	0	.810	0	0
Word Attack	7	0	0	0	.715	0	0	0	0
Quantitative Concepts	11	0	0	0	0	0	.251	0	.618
Writing Fluency	5	0	0	.422	0	0	0	0	0
K-ABC									
Hand Movements	2	0	.257	0	0	0	0	0	.202
Gestalt Closure	2	0	0	0	0	.319	0	0	0
Number Recall	2	0	.718	0	0	0	0	0	0
Triangles	2	0	0	0	0	.672	0	0	0
Word Order	2	0	.555	0	0	0	0	0	0
Matrix Analogies	2	0	0	0	0	.334	0	.254	0
Spatial Memory	2	0	0	.158	0	.162	0	0	0
Photo Series	2	0	0	0	0	.284	0	.278	0
Faces & Places	2	0	0	0	0	0	.616	0	0
Arithmetic	2	0	0	0	0	0	.142	0	.616
Riddles	2	0	0	0	0	0	.698	0	0
SB-IV									
Vocabulary	4	0	0	0	0	0	.810	0	0
Bead Memory	4	0	.340	0	0	.237	0	0	0
Quantitative	4	0	0	0	0	0	0	0	.654
Memory of Sentences	4	0	.494	0	0	0	.473	0	0
Pattern Analysis	4	0	0	0	0	.574	0	0	0
Comprehension	4	0	0	0	0	0	.318	0	0
Absurdities	2	0	0	0	0	0	.384	0	0
Memory for Digits	4	0	.782	0	0	0	0	0	0
Copying	2	0	0	0	0	.470	0	0	0
Memory for Objects	4	0	.324	0	0	0	0	0	0
Matrices	4	0	0	0	0	0	0	.609	0
Number Series	4	0	0	0	0	0	0	0	.676
Paper Folding & Cutting	2	0	0	0	0	.510	0	.122	.326
Verbal Relations	2	0	0	0	0	0	.713	0	0
Equation Building	2	0	0	0	0	0	0	0	.776
WISC-R/WAIS/WAIS-R									
Information	8	0	0	0	0	0	.670	0	0
Similarities	8	0	0	0	0	0	.600	0	0
Arithmetic	8	0	0	0	0	0	0	0	.753
Vocabulary	8	0	0	0	0	0	.809	0	0
Comprehension	8	0	0	0	0	0	.692	0	0
Digit Span	8	0	.692	0	0	0	0	0	0
Picture Completion	8	0	0	0	0	.453	.248	0	0
Picture Arrangement	8	0	0	0	0	.197	.315	0	0
Block Design	8	0	0	0	0	.578	0	.123	0
Object Assembly	8	0	0	0	0	.622	0	0	0
Coding (Digit Symbol)	8	0	0	.582	0	0	0	0	0
Mazes	4	0	0	0	0	.464	0	0	0

*Primary measure of the factor in the WJ-R.

For the purpose of comparing the factor measurement capability of the 68 variables included in Table 5, criteria for classifying the tests were devised. These criteria are presented in Table 6. A subtest is classified as a *strong* measure of a factor when its loading on that factor is greater than .50 and it has no loading on a second factor greater than one-half of its loading on the primary factor. (If the loading on the second factor were one-half the loading on the primary factor, the amount of variance accounted for by the second factor would be 25% of the primary factor.) A subtest is classified as *moderate* when its loading on the primary factor is less than .50 and there is no significant loading on a second factor. A subtest is classified as *moderate* when the loading on a second factor falls between five-tenths and seven-tenths of the primary loading (25 to 50% of the variance accounted for by the primary factor). A subtest is classified as *mixed,* regardless of the loading on the primary factor, when there is a loading on a second factor greater than seven-tenths (50% of the variance) of the loading on the primary factor.

Table 6
Criteria for Classifying Tests

Rating	Primary factor loading	Secondary factor loading
Strong	Greater than .500	Less than one-half of primary loading
Moderate	Less than .500	Less than one-half of primary loading
Moderate	(Any)	Between one-half and seven-tenths of primary loading
Mixed	(Any)	Greater than seven-tenths of primary loading

Table 7 presents a qualitative interpretation of the information, by factor, from Table 5. The subtests reported as measuring each factor are classified according to the criteria in Table 6.

Table 7
Measures of Eight Gf-Gc Factors Included in Six Cognitive Batteries

	Quality of measure			
Cognitive battery	Strong loading	Moderate loading	Mixed loading	Number of clean measures
Long-Term Retrieval *(Glr)*				
WJ-R (1989)	Mem Names* V-A Lrng* DR-M Nam	DR-VALrng	—	4
K-ABC	—	—	—	0
SB-IV	—	—	—	0
WISC-R	—	—	—	0
WAIS-R	—	—	—	0
WJ (1977)	V-A Lrng	—	—	1
Short-Term Memory *(Gsm)*				
WJ-R (1989)	Mem Words*	Mem Sents*	Nums Rev *(Gf)*	2
K-ABC	Num Recl Word Ord	—	Hand Mov *(Gq)*	2

Table 7 (Continued)

Cognitive battery	Quality of measure: Strong loading	Moderate loading	Mixed loading	Number of clean measures
SB-IV	Mem Digs	Mem Objs	Bead Mem *(Gv)* Mem Sents (Gc)	2
WISC-R	Digit Sp	—	—	1
WAIS-R	Digit Sp	—	—	1
WJ (1977)	—	Mem Sents	Nums Rev *(Gf)*	1
Processing Speed *(Gs)*				
WJ-R (1989)	Vis Match* Cross Out*	Wrtg Flu	—	3
K-ABC	—	—	—	0
SB-IV	—	—	—	0
WISC-R	Coding	—	—	1
WAIS-R	Dig Symb	—	—	1
WJ (1977)	Spa Rels[b] Vis Match	—	—	2
Auditory Processing *(Ga)*				
WJ-R (1989)	Inc Words* Snd Blndg* Word Atk	—	Snd Patts *(Gf)*	3
K-ABC	—	—	—	0
SB-IV	—	—	—	0
WISC-R	—	—	—	0
WAIS-R	—	—	—	0
WJ (1977)	Blending Word Atk	—	—	2
Visual Processing *(Gv)*				
WJ-R (1989)	—	Vis Clos* Pict Recg*	—	2
K-ABC	Triangles	Gest Clo	Photo Ser *(Gf)* Mat Anlgs *(Gf)* Spa Mem *(Gs)*	2
SB-IV	Patt Anl	Copying Paper F&C	—	3
WISC-R	Blk Des Obj Assm	Pict Cmpl Mazes	—	4
WAIS-R	Blk Des Obj Assm	Pict Cmpl	—	3
Comprehension-Knowledge *(Gc)*				
WJ-R (1989)	Pict Voc* Orl Voc* List Comp Science Social St Humanities	—	—	6
K-ABC	Faces & P Riddles	—	—	2
SB-IV	Vocab Vrb Rels	Absurd Comp	—	4

Table 7 (Continued)

Cognitive battery	Quality of measure: Strong loading	Moderate loading	Mixed loading	Number of clean measures
WISC-R	Inform Simils Vocab Comp	Pict Arrg	—	5
WAIS-R	Inform Simils Vocab Comp	Pict Arrg	—	5
WJ (1977)	Pict Voc Ant-Syn Science Social St Humanities	—	—	5
Fluid Reasoning *(Gf)*				
WJ-R (1989)	Anl-Synth* Con Form*	Spa Rels	Vrb Anlgs *(Gc)*	3
K-ABC	—	—	—	0
SB-IV	Matrices	—	—	1
WISC-R	—	—	—	0
WAIS-R	—	—	—	0
WJ (1977)	Anl-Synth Con Form	—	Anlgs *(Gc)*	2
Quantitative Ability *(Gq)*				
WJ-R (1989)	Calc* App Probs* Quant Con	—	—	3
K-ABC	Arith	—	—	1
SB-IV	Quant Num Sers Eq Bldg	—	—	3
WISC-R	Arith	—	—	1
WAIS-R	Arith	—	—	1
WJ (1977)	Quant Con Calc App Probs	—	—	3

*Primary measure of the factor in the WJ-R.
[a]Sum of strong plus moderate loading measures.
[b]Spatial Relations was a speed test in the 1977 WJ.

The first section in Table 7 includes information on measures of the long-term retrieval (*Glr*) factor. The WJ-R includes four measures of long-term retrieval with both primary measures classified as strong. The 1977 WJ includes one strong measure of long-term retrieval. No other cognitive battery appears to measure this important ability. The last column in Table 7 reports the number of *clean* (strong plus moderate) measures available in each battery—four subtests for the WJ-R, one subtest for the 1977 WJ.

The next section of Table 7 presents information about the measures of short-term memory (*Gsm*) included in the six cognitive batteries. The measurement of this factor appears relatively weak in the WJ-R because one primary subtest (Memory for Words) is classified as strong, but the other primary subtest (Memory for Sentences) has only a moderate loading on that factor. This is because Memory for Sentences also loads on the *Gc* factor, though not highly enough to cause it to be classified as mixed. Note that Table 7 shows that these two WJ-R subtests are most associated, factorwise, with the subtests from the other cognitive batteries that also would be considered measures of short-term memory (K-ABC: Number Recall and Word Order; SB-IV: Memory for Digits; Wechsler Scales: Digit Span). The WJ-R contains two clean measures of *Gsm*; the WJ has one. The K-ABC and SB-IV each have two clean measures and the Wechsler scales have one.

Note, in Table 7, that Numbers Reversed in the WJ-R and WJ are classified as mixed subtests with significant loadings on the *Gf* factor. The related tasks in the K-ABC, SB-IV and Wechsler scales are classified as a strong measure and do not show a significant loading on *Gf*. This is because they are a combination of a numbers forward task plus a numbers reversed task. A numbers reversed task appears to require both short-term memory and fluid reasoning. A numbers forward task, however, is a purer measure of short-term memory than numbers reversed. A subtest based on a combination of both numbers forward and numbers reversed will appear stronger on short-term memory due to the influence of numbers forward.

The next section of Table 7 presents information on the processing speed (*Gs*) subtests. The WJ-R includes two strong measures of *Gs*. These are seen in Table 7 to be most associated, factorwise, with the Coding or Digit Symbol subtests of the Wechsler scales, two speed subtests. The WJ-R contains three clean measures of *Gs* (including Writing Fluency from the WJ-R ACH), the Wechsler scales contain one, and the 1977 WJ contains two. (Note that the Spatial Relations subtest in the 1977 WJ was a speed test.)

Table 7 next presents information on the measures of auditory processing (*Ga*) included in the cognitive batteries. The WJ-R and the WJ are the only cognitive batteries that provide measures of the *Ga* factor, a significant cognitive ability required for school success. Both primary measures of auditory processing in the WJ-R are classified as strong.

The next section in Table 7 summarizes the information about the measures of the visual processing (*Gv*) factor in the cognitive batteries. Visual processing appears to be relatively weak in the WJ-R. Both of the primary measures (Visual Closure and Picture Recognition) have only a moderate loading on that factor. These subtests are most associated, factorwise, with subtests such as Triangles from the K-ABC, Pattern Analysis from the SB-IV, and Object Assembly or Block Design from the Wechsler scales. The 1977 WJ has no measure of the *Gv* factor.

The next section of Table 7 summarizes the results for the subtests associated with the comprehension-knowledge (*Gc*) factor. All six batteries provide two or more strong measures of the *Gc* factor. Both of the two primary measures of *Gc* in the WJ-R (Picture Vocabulary and Oral Vocabulary) are classified as strong measures. One other test from the WJ-R COG (Listening Comprehension) and three tests from the WJ-R ACH (Science, Social Studies, and Humanities) are also strong measures of the *Gc* factor. Including the three knowledge tests, there are six *Gc* measures available in the WJ-R. These subtests are most associated, factorwise, with subtests such as Faces and Places from the K-ABC, Vocabulary from the SB-IV, and Information from the Wechsler scales.

Table 7 next summarizes the results for the fluid reasoning (*Gf*) factor. The two primary measures of *Gf* in the WJ-R are classified as strong. The surprising feature of this table is the apparent lack of *Gf* measures in most other cognitive batteries. Matrices from the SB-IV, according to these analyses, are the only clean measure of *Gf* available in batteries other than the WJ-R and WJ. Subtests from other batteries that sometimes are assumed to measure reasoning appear to be measures of visual processing. For example, Block Design in the Wechsler scales is a task that requires the copying of a pattern with blocks. It does not seem unreasonable, then, that this subtest should load on *Gv* rather than *Gf*. In comparison, tasks such as SB-IV Matrices or the WJ-R Concept Formation also present visual stimuli, but the subject must transform this visual information into a different product.

The last section of Table 7 summarizes the results for the quantitative ability (*Gq*) factor. Both primary measures of *Gq* included in the WJ-R are classified as strong. These WJ-R subtests are most associated, factorwise, with the subtests called Arithmetic in the K-ABC or Wechsler scales and the SB-IV Quantitative subtest.

Factorial Congruence Across Batteries

A unique value of the data available for analysis in this report is the number of cognitive batteries included. Each of these batteries provides some system of factor scores, all with different names. What is the congruence with *Gf-Gc* theory among the several factor interpretation schemes? Remarkably high, as can be observed in Table 8. This congruence of posited factors from other batteries with the broad

Table 8
Congruence of Factor Scores Provided by Six Cognitive Batteries

Gf-Gc factor	Cognitive battery				
notation	WJ-R (1989)	K-ABC	SB-IV	Wechsler scales	WJ (1977)
Glr or *TSR*	Long-Term Retrieval	—	—	—	—
Gsm or *SAR*	Short-Term Memory	Sequential Processing	Short-Term Memory	—	Memory
Gs	Processing Speed	—	—	—	Perceptual Speed
Ga	Auditory Processing	—	—	—	—
Gv	Visual Processing	Simultaneous Processing	—	Perceptual Organization	—
Gc	Comprehension-Knowledge	—	Verbal Reasoning	Verbal Comprehesion	Verbal Ability, Oral Language
Gf	Fluid Reasoning	—	—	—	Reasoning, Broad Reasoning
Gq	Quantitative Ability	—	Quantitative Reasoning	—	—

abilities of *Gf-Gc* theory provides further construct validity evidence for the WJ-R cognitive clusters.

Another examination of Table 7 will show that the set of three K-ABC subtests shown there as measuring *Gsm* is identical to the set of three subtests identified as Sequential Processing in the K-ABC. For the SB-IV, the set of four subtests shown in Table 7 as measuring *Gsm* is identical to the set of four subtests identified as Short-Term Memory in that battery. Also, the two 1977 WJ *Gsm* subtests comprise the Memory cluster in that battery.

In respect to the processing speed factor, Table 7 shows that the 1977 WJ Spatial Relations and Visual Matching subtests load on that factor. These two subtests comprise the Perceptual Speed cluster of that battery. No battery other than the WJ-R or WJ provides a score for the processing speed factor.

Table 7 shows information with regard to subtest loadings on the *Gv* factor. Note that the set of five K-ABC subtests identified there as measuring *Gv* is identical to the set of five subtests identified as Simultaneous Processing in the K-ABC. (The K-ABC Magic Window and Face Recognition subtests were not included in these analyses.) Note also that, except for the absence of Picture Arrangement (which does have some loading on *Gv* and may have relevance for the WISC-R, but not the WAIS-R), the set of Wechsler subtests shown in Table 7 as measuring *Gv* is identical to the set of subtests that comprise the Perceptual Organization factor (Kaufman, 1975).

Table 7 reports the information with regard to *Gc* subtests. The four highest loading SB-IV subtests associated with this factor are the same four subtests identified as a Verbal Reasoning factor in that battery. In the Wechsler scales, the four strong measures of *Gc* are those that comprise the Verbal Comprehension factor (Kaufman, 1975). For the 1977 WJ, the two measures of *Gc* (used in conjunction with a *Gf* suppressor variable) comprised the Verbal Ability cluster. An alternative WJ cluster, provided later, termed Oral Language, consisted of these two cognitive battery subtests (1 and 8) plus the Analogies subtest, which has an almost equal weighting on *Gc* and *Gf*.

Table 7 reports the subtests that load primarily on *Gf*. No other battery except the WJ-R or the 1977 WJ provides a score for the fluid reasoning factor, an important cognitive ability. The subtests from the 1977 WJ (used in conjunction with a *Gc* suppressor variable) comprised the Reasoning cluster score. An alternative cluster, provided later, termed Broad Reasoning, consisted of the three subtests without the addition of a suppressor variable.

Table 7 reports information on subtests that load on *Gq*. The set of three SB-IV subtests shown there as primarily measuring *Gq* is identical to the set of three subtests identified as Quantitative Reasoning in that battery.

Note that Table 8 does not include the SB-IV Visual/Abstract Reasoning composite. This cluster would have been included in Table 8 as a measure of *Gv* except for its inclusion of Matrices (a measure of *Gf*). It appears that the four-factor interpretation of the SB-IV as provided in the published battery is more closely supported by these results than the three-factor interpretation proposed subsequently by Sattler (1988).

Of interest in these findings is the lack of any combination of loadings that supports the Wechsler Freedom from Distractibility factor. As discussed earlier, many factor analysis studies have been restricted to only those subtests included within a cognitive battery itself. It is suggested that the three subtests that comprise the Freedom from Distractibility factor (Digit Span, Arithmetic, and Coding) are each clean measures, in their own right, of three separate factors (*Gsm, Gq,* and *Gs*). Freedom from Distractibility may or may not provide a clinically useful composite score, but there is no suggestion here that a common factor underlies the three subtests.

General Intelligence

> The practice of intellectual assessment of children is currently marked by controversy. However, much of that controversy could be set aside if intelligence tests were viewed appropriately. Intelligence tests are simply samples of behaviors. And different intelligence tests sample different behaviors. For that reason, it is wrong to speak of a person's IQ. Instead, we can refer only to a person's IQ on a specific test. . . . Because the behavior samples are different for different tests, one must always ask, *"IQ on what test?"* (Salvia & Ysseldyke, 1988, p. 158)

All six cognitive batteries provide a broad measure of intellectual ability based on average subtest performance. Users of cognitive batteries often assume that these broad-based scores are measures of general intelligence. Not so; they are simply an average of whatever has been chosen by the test author to be included in that battery. The broad-based score from the K-ABC (Mental Processing Composite)

Table 9
Factorial Composition of the Broad Measure of Intelligence Provided by Six Cognitive Batteries

	WJ-R BCA		K-ABC MPC		SB-IV CSAS		WISC-R FSIQ		WAIS-R FSIQ		WJ(1977) BCA	
Factor	Sum	%	Sum	%	Sum	%	Sum	%	Sum	%	Sum	%
Long-Term Retrieval (*Glr*)	2	14	0	0	0	0	0	0	0	0	1	8
Short-Term Memory (*Gsm*)	2	14	2½	31	3	20	0	0	1	9	1½	12
Processing Speed (*Gs*)	2	14	½	6	0	0	1	10	1	9	2	17
Auditory Processing (*Ga*)	2	14	0	0	0	0	0	0	0	0	1	8
Visual Processing (*Gv*)	2	14	3½	44	3½	23	3	30	3	27	0	0
Comprehension-Knowledge (*Gc*)	2	14	0	0	4½	30	5	50	5	45	2½	21
Fluid Reasoning (*Gf*)	2	14	1	12	1	7	0	0	0	0	3	25
Quantitative Ability (*Gq*)	0	0	½	6	3	20	1	10	1	9	1	8
Total battery	14	100	8	100	15	100	10	100	11	100	12	100
Applicable age range	(5 to 90+)		(6 to 12-5)[a]		(12 to 13)[b]		(6 to 16-6)		(16 to 74)		(5 to 80+)	
Achievement (*Gc* + *Gq*)	2	14	½	6	7½	50	6	60	6	55	3½	29

[a]Fewer tests and a different factorial mix will be observed below age 6.
[b]Fewer tests and different factorial mixes will be observed below age 12 and above age 13.

is based on a different weighted mix of cognitive abilities than is the broad-based score from the SB-IV (Test Composite), or a Wechsler (Full Scale IQ), or the WJ-R (Broad Cognitive Ability). The interpretation of these broad-based scores is complicated further by the fact that, in some batteries, different combinations of subtests are involved in the calculation of a particular score at one age than at another age.

Table 9 summarizes the factorial composition of the broad-based measure of intelligence provided by each of the batteries. For each battery, the number of tests that measure each cognitive factor is reported, as well as the approximate percent those tests contribute to the total broad score. (Tests with "mixed" loadings are counted as one-half test toward the total of a factor's contribution to the broad score.)

For the WJ-R Broad Cognitive Ability score (Extended Scale), there are two subtests that measure each of seven factors. (The two measures of quantitative ability are not included in the calculation of the WJ-R Broad Cognitive Ability score.) The two measures for each factor contribute approximately 14% each toward the total score. The applicable age range for this factorial composite is from age 5 to more than 90 years.

In the K-ABC, there appear to be five factors (*Gsm, Gs, Gv, Gf,* and *Gq*) that contribute to the composite score (MPC) aross the age range 6 to 12 years. The proportional contribution of these five factors varies from 6% each for processing speed and quantitative ability to 44% for visual processing.

For the SB-IV, there appear to be five factors (*Gsm, Gv, Gc, Gf,* and *Gq*) that contribute to the composite score (CSAS). The proportional contribution of these five factors varies from 7% for fluid reasoning to 30% for comprehension-knowledge. The number of tests and their mix reported here for the SB-IV are only valid for 12- and 13-year-olds. Below 12 years of age and above 13 years of age, fewer tests and different factorial mixes will be represented in the SB-IV broad score of intelligence. To address this problem, Sattler (1988, p. 259) provides a table of suggested subtest combinations for factor scores at various age ranges.

For the WISC-R, there appear to be four factors (*Gs, Gv, Gc,* and *Gq*) that contribute to the Full Scale IQ across ages 6 to 16 years. The proportional contribution of these four factors varies from 10% each for processing speed and quantitative ability to 50% for comprehension-knowledge.

For the WAIS-R, there appear to be five factors (*Gsm, Gs, Gv, Gc,* and *Gq*) that contribute to the Full Scale IQ across ages 16 to 74 years. The proportional contribution of these five factors varies from 9% each for short-term memory, processing speed, and quantitative ability to 45% for comprehension-knowledge.

For the 1977 WJ, there appear to be seven factors (*Glr, Gsm, Gs, Ga, Gc, Gf,* and *Gq*) that contribute to the Broad Cognitive Ability score across ages 5 to over 80 years. The proportional contribution of these seven factors varies from 8% each for long-term retrieval, auditory processing, and quantitative ability to 25% for fluid reasoning.

A common misconception about the Woodcock-Johnson cognitive tests has been that they are based in large part upon measures of academic achievement (Sattler, 1988; Shinn, Algozzine, Marston, & Ysseldyke, 1982; Thompson & Brassard, 1984a, 1984b). Measures of *Gc* and *Gq* are the two cognitive factors most influenced by school learning. The bottom row of Table 13 reports the number of *Gc* and *Gq* subtests in each battery and their percentage contribution to the broad score of intelligence.

The WJ-R BCA includes two measures of achievement, as defined above, which contribute 14% to the broad score. The K-ABC Mental Processing Composite has almost no contribution from achievement, as intended by the authors of that battery. The 1977 WJ has about 29% of the Broad Cognitive Ability score contributed by measures of achievement. Note that the SB-IV and the Wechsler scales have 50% or more of their broad scores of intelligence contributed by measures of the *Gc* and *Gq* achievement factors.

INTERPRETING FACTOR SCORES

If a clinician is to interpret the results from a battery of cognitive tests in respect to a subject's strengths and weaknesses among the factors measured in a battery, it is necessary to observe four principles similar to those followed by factor analysts when they interpret the results of an analysis:

1. The clinician must be knowledgeable about the factorial composition of each subtest in the battery. "Teachers, administrators, counselors, and diagnostic specialists must go beyond test names and scores to look at the kind or kinds of behaviors sampled on the test" (Salvia & Ysseldyke, 1988, p. 149). This information aids in selecting tests to meet specific diagnostic needs and, subsequently, in interpreting an individual's test performance.

2. The clinician must have information from two or more clean measures of a cognitive factor before generalizing about a subject's ability in that cognitive area. The tests should represent qualitatively different aspects of that factor. For example, two measures of vocabulary do not generalize to a person's *Gc* ability as well as a measure of vocabulary plus a measure of general information.

3. The factor measures must be clean or any attempt to explain performance on a subtest will be complicated, if not impossible. Mixed-factor subtests usually do not measure some unique composite ability. For example, the WJ-R Spatial Relations subtests is a mixed *Gv* and *Gf* measure, but it should not be interpreted as a measure of visual reasoning. There is no way to determine whether a subject's poor performance on a mixed factor subtest is due to low ability in A, low ability in B, or low ability in both A and B. Furthermore, a mixed factor subtest may require all of the independent abilities (A and B) to perform the task, or it may be a subtest in which different persons apply different abilities (A or B) to their solution.

4. The clinician may find it helpful to "cross" batteries to obtain a set of measures required for a particular assessment. (Table 7 provides a guide for such selection of cross-battery tests). This does present a problem from the lack of common norms. A partial solution may be to approach assessment more from the viewpoint of a "clinical evaluation" than from obtaining "measurements."

Table 10 summarizes the information on the number of clean measures for each factor included in the six cognitive batteries. These analyses indicate that the WJ-R measures all eight of its factors with two or more clean measures across an age range of 5 to more than 90 years. The 1977 WJ measures five of its seven factors with two or more clean measures across an age range of 5 to more than 80 years. In comparison, the Wechsler scales measure only two of their four or five factors with two or more clean measures across their respective age ranges.

Table 10
Clean Measures for Each Factor Included in Six Cognitive Batteries

	Cognitive battery					
Factor	WJ-R	K-ABC	SB-IV	WISC-R	WAIS-R	WJ
Long-Term Retrieval *(Glr)*	4	0	0	0	0	1
Short-Term Memory *(Gsm)*	2	3	1	1	1	1
Processing Speed *(Gs)*	3	0	0	1	1	2
Auditory Processing *(Ga)*	3	0	0	0	0	2
Visual Processing *(Gv)*	2	2	3	4	3	0
Comprehension-Knowledge *(Gc)*	6	2	4	5	5	5
Fluid Reasoning *(Gf)*	3	0	1	0	0	2
Quantitative Ability *(Gq)*	3	1	3	1	1	3
Total battery						
Number of factors with two or more clean measures	8	3	4	2	2	5
Applicable age range	(5 to 90+)	(4 to 12-5)	(7 to 13)	(6 to 16-6)	(16 to 74)	(5 to 80+)

SUMMARY AND CONCLUSIONS

The results from these analyses conducted on nine data sets demonstrate that the WJ-R is a strong instrument for measuring *Gf-Gc* broad abilities. The conclusions with regard to the factorial composition of the WJ-R subtests are similar whether derived from Woodcock-Johnson norming data sets or from studies that include subtests from other cognitive batteries. When other batteries are included, the various WJ-R subtests are observed to be most highly associated, factorwise, with the subtests from the other batteries that are measuring comparable abilities. Furthermore, the combined set of subtests from the WJ-R and the other cognitive batteries load appropriately onto a set of factors defined by *Gf-Gc* theory.

For each of eight factors, at least two clean measures are provided in the WJ-R. There may be a relative weakness, however, in the measurement of visual processing and short-term memory because some of the primary measures for these two factors are classified as moderate rather than strong. One feature of the WJ-R is the provision for evaluating the presence of intra-cognitive discrepancies. These results provide evidence that the clinician may assume that different cognitive abilities are being compared.

The Broad Cognitive Ability cluster for the WJ-R provides a balanced mix of information from seven of the eight cognitive factors. The quantitative ability factor is not represented in the Broad Cognitive Ability score.

The principles that underlie the interpretation of factor analysis results have parallel implications for the clinical interpretation of performance on subtests of intelligence. These implications include the need for two or more clean measures for any factor to be assessed and the exercising of extreme caution in attempting to interpret mixed measures. It has been demonstrated that the necessary clean measures are included in the WJ-R for each factor. Factorially mixed WJ-R subtests have been identified, both in this report and in the published test materials.

The results of these studies demonstrate the need for factor analytic studies in which the set of variables is not constrained to the limited set of subtests that have been published as a battery. It is recommended that future factor studies ensure that a breadth of cognitive abilities be repesented in the set of included measures and that an adequate number of subtests that represent each factor is provided so that the presence of all factor effects can be identified. It appears that cognitive test batteries usually have been underfactored. For example, the analyses that included the WISC-R (with the optional Digit Span subtest) suggest that five factors (*Gsm, Gs, Gv, Gc,* and *Gq*) are all measured by that battery, not just three as is commonly believed. For the purpose of test development and validation, it is recommended that that work includes validating the factorial loadings of subtests by including strong measures from other batteries as needed for markers.

In the long term, some of the most important questions for the field to address concern implications for intervention. *Gf-Gc* theory, in general, provides a useful reference with regard to the variety of human cognitive abilities. If a more exact picture of the abilities measured by the subtests in various cognitive batteries is obtained, advances in intervention planning should follow.

The most important outcome of this report may be the stimulation of new research, or even the reanalysis of old data sets, in the light of *Gf-Gc* theory with the recognition that an adequate breadth and depth of markers are required to identify the factors present. Information from such studies may further validate these findings, or it may lead to a more accurate interpretation. Any replication or extension of this work by others will result in a better understanding and application of test results from all cognitive batteries, not just the WJ-R.

RESPONSE TO CONFERENCE COMMENTS

My compliments are extended to Professors Ysseldyke and Reschley for their thorough and thoughtful reviews. Their efforts are especially noteworthy because they received my paper only 2 weeks before the conference.

In addition to the points raised by these distinguished reviewers, several conference participants have introduced other comments and raised additional questions that I would like to address. This response will address three broad topics: (1) the evolution of *Gf-Gc* theory, (2) criticisms concerning a lack of theory underlying the 1977 WJ, and (3) the *g* word.

The Evolution of Gf-Gc Theory

Several presenters and discussants at this conference have interpreted *Gf-Gc* as a two-factor theory that consists of fluid and cyrstallized intelligences. This is a common misconception of the theory, even though it has not been the view of the two major proponents, Horn and Cattell, for 25 years. The following are some of the developmental milestones in *Gf-Gc* theory, beginning with its two-factor origin to its present definition of 9 or 10 broad abilities that characterize what most lay people and professionals call "intelligence."

The first mention of fluid and crystallized intelligences is in a paper presented by Cattell to a convention of the American Psychological Association (Cattell, 1941). Horn's doctoral dissertation (Horn, 1965) started the expansion of this theory. Horn analyzed a broad sample of tests that measure primary mental abilities and, in addition to *Gf* and *Gc*, identified four factors that he labeled *Gv, SAR, TSR,* and

Gs. Horn later published an article in the *Psychological Review* (Horn, 1968) that foreshadowed a conceptual shift in *Gf-Gc* theory to a theory of multiple intelligences. In that article, Horn discussed further refinements that concerned *Gv, Gs, Glr,* and possible *Ga* and tactile functions. (At that time, he referred to these additional abilities as "functions" rather than intelligences.)

Later, Hakstian and Cattell (1978) reported deriving six common factors from measures of 20 primary abilities. The factors included crystallized intelligence (*Gc*), fluid intelligence (*Gf*), visualization capacity (*Gv*), general perceptual speed (*Gps*), memory capacity (*Gm*), and general retrieval capacity (*Gr*). They also obtained three third-order factors. In respect to the third-order factors they stated, "Because this stratum has been so rarely explored . . . it is probably not worthwhile to attempt to discuss further theory until more studies have been accumulated" (p. 668).

In 1981, Carroll and Horn stated:

> the evidence of multiple factor research indicates that on the order of 80% of intellectual ability can be measured and predicted in terms of as few as 30 basic processes. . . . There is considerable evidence to suggest that most of the variation represented in primary processes can be organized in terms of as few as eight or nine second-order abilities . . . there is general agreement that a major portion of the lawful . . . variation in what is regarded, by lay person and ability specialist alike, as intelligence can be measured and predicted with only the second-order factors (Carroll & Horn, 1981, p. 1016).

By 1985, *Gf-Gc* theory became more clearly distinguished, in the professional literature, from its dichotomous beginnings:

> You see that the current theory is notably different from the theory of a similar nature that was put forth some 40 years ago by Raymond Cattell. . . . research results have accumulated to indicate early formulations must be reformulated. . . . the available evidence leaves virtually no doubt about a conclusion that there are several distinct factors among performances, all of which are said to indicate intelligence (Horn, 1985, pp. 273-274).

In 1985, Horn presented a figure that represented a model of *Gf-Gc* theory based on 10 abilities, which included 2 sensory detectors (Horn, 1985). This figure, or variations thereof, has since appeared several times in the literature. Horn's most recent discussion of *Gf-Gc* theory is a chapter in the second edition of the *Handbook of multivariate experimental psychology* (Horn, 1988). The model, as described in that chapter, has been expanded to include a quantitative ability (*Gq*) factor and suggests the possibility of at least one broad English-language factor separate from *Gc*. This reference is recommended reading for anyone who wishes a thorough discussion of *Gf-Gc* theory in its current form.

Gf-Gc Theory and the 1977 WJ

A frequent criticism of the 1977 WJ has been the lack of an articulated theoretical base for the battery. This criticism is based on the relatively recent trend of defining cognitive functioning in terms of theoretical constructs. The trend has my wholehearted support. This criticism of the 1977 WJ is deflated, however, by historical perspectives of *Gf-Gc* theory, including its form in the early 1970s when the WJ was being developed.

While I was designing the 1977 WJ, I reviewed the major theories of intelligence extant in the early 1970s. These included *Gf-Gc* theory, which I perceived at that time to be a two-factor theory. Although Horn's research from the 1960s was published in a book by Cattell (1971), the concept of multiple intelligence still was not articulated clearly or widely acknowledged in the professional literature. In fact, my original proposal was to develop a battery of tests that followed Newland's process-product conceptualization of *Gf-Gc* theory (referenced by Ysseldyke in his review). Factor analyses of the early data studies (studies 4A and 4B), however, did not support that dichotomization of cognitive abilities.

Articulating a theoretical base for the 1977 WJ was omitted intentionally because no theory, *at that time,* provided an adequate description of cognitive functioning. Not articulating a theoretical base is not the same thing as *lacking* a theoretical base. Remember that *Gf-Gc* theory has its foundation in the scientific method and that its data are derived from studies with both published and unpublished batteries of intellectual tests, including the WJ. The evolution of *Gf-Gc* into an articulated multi-factor theory of intelligence and the development phase and standardization phases of the 1977 WJ occurred *concurrently in time.*

Proponents of the view that tests should be based on a theoretical model are assuming that acceptable and useful models are available. I do not believe there was a theory of intelligence available in the early 1970s that adequately specified a model for the set of cognitive tests needed for broad psychoeducational applications. Does this mean that I should have followed some theory of intelligence, *any theory,* no matter how inadequate or unresearched that theory might have been? That is like advising two people to get married because they are desperate, even though they are not compatible!

The g Word

During this conference, three approaches for obtaining measures of general intelligence have been mentioned:

1. Calculate the simple average of a set of subtest scores, as performed with the cognitive batteries (except the 1977 WJ) reported in Table 9.

2. Calculate the average of subtest scores weighted according to the results of a first-principal-components analysis. This is the procedure espoused by Arthur Jensen and was the procedure followed in preparing the Broad Cognitive Ability scores for the 1977 WJ (Woodcock, 1978).

3. Calculate a third- or higher-order factor score, termed "hierarchical *g*" by some. No currently published battery provides a measure of general intelligence based on this procedure.

There are two broad sets of issues to consider: one is technical and the other is practical. First, the technical issue. *Any one of the three approaches produces a broad score that is a function of the variety and mix of tests chosen by the test author.* This is true even of the hierarchical *g* approach. Some comments voiced during the conference appear to be based on the assumption that hierarchical *g* scores would have consistent content implications across different collections of tests. This is true only to the extent that the collections of tests are similar. There is nothing esoteric embedded in a matrix of cognitive-ability correlations that would not be there if the same matrix were based on geological observations! Computers are obliging and will compute a broad score from any set of correlations, based on any one of the three approaches, if properly asked to do so.

I am not questioning that fuller explanatory models that describe the structure and organization of cognitive abilities cannot be obtained through hierarchical approaches. Carroll (1989) presently is engaged in a definitive effort toward that goal. His work is based on reanalyses of 461 data sets, which represent the greatest breadth of cognitive variables studied to date. His results may provide us with benchmarks against which to compare various *g*s obtainable from current test batteries.

Now the practical issue. There is no question, technically, that broad scores of intelligence can be calculated, but what do such broad scores mean in practice? At best, they predict the average outcome in a variety of life situations that require a variety of cognitive abilities. They do not allow us to predict anything specific. By analogy, similar computations could be applied to the data provided by composite sets of reading, writing, and arithmetic tests. An achievement quotient, or AQ, could be determined by following any of the three computational approaches that have been applied to intelligence test data. No matter how the AQ is determined, will it be of much solace to the referring teacher to hear that the student has an AQ of 90 or 122? Most teachers would respond, "So what?" Unfortunately, that has not been a frequent enough response to reporting the single-number IQ as a representation of intelligence. Perhaps this dedication to the single-number IQ is one cause behind the move, by some, to depose the use of intelligence tests altogether. Perhaps it is, also, one cause behind the complaint that useful intervention information is not provided by intelligence tests. The time is past due for more test developers and users to appreciate intelligence as a many-splendored thing, not a single trait analogous to "mental height."

REFERENCES

Carroll, J.B. (1989). Factor analysis since Spearman: Where do we stand? What do we know? In R. Kanfer, P. L. Ackman, & R. Cudek (Eds.), *Abilities, motivation, and methodology* (pp. 43-67). Hillsdale, NJ: Erlbaum.

Carroll, J. B., & Horn, J. L. (1981). Scientific basis of ability testing. *American Pyschologist, 36,* 1012-1020.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38,* 592. (Abstract)

Cattell, R. B. (1963). Theory for fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1-22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology, 70,* 657-669.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic and development study of structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242-259.

Horn, J. L. (1972). State, trait and change dimensions of intelligence. *British Journal of Educational Psychology, 42,* 159-185.

Horn, J. L. (1976). Human abilities: A review of research and theory in the early 1970s. *Annual Review of Psychology, 27,* 437-485.

Horn, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267-300). New York: John Wiley.

Horn, J. L. (1986). Some thoughts about intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 91-96). Norwood, NJ: Ablex.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology, Second edition* (pp. 645-685). New York: Plenum Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57,* 253-270.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology, 43,* 135-147.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1990). *WJ-R technical manual.* Allen, TX: DLM Teaching Resources.

Salvia, J., & Ysseldyke, J. E. (1988). *Assessment in special and remedial education, Fourth edition.* Boston: Houghton Mifflin.

Sattler, J. M. (1988). *Assessment of children, Third edition.* San Diego, CA: Author.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. (1982). A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery. *Journal of Learning Disabilities, 15,* 221-226.

Thompson, P., & Brassard, M. (1984a). Validity of the Woodcock-Johnson Tests of Cognitive Ability: A comparison with the WISC-R in LD and normal elementary students. *Journal of School Psychology, 22,* 201-208.

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

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

Woodcock, R. W. (1984). A response to some questions raised about the Woodcock-Johnson Psycho-Educational Battery: II. Efficacy of the aptitude clusters. *School Psychology Review, 13,* 355-362.

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

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Mather, N. (1989a). WJ-R Tests of Achievement: Examiner's manual. In R. W. Woodcock & M. B. Johnson, *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Mather, N. (1989b). WJ-R Tests of Cognitive Ability—Standard and Supplemental Batteries: Examiner's manual. In R. W. Woodcock & M. B. Johnson, *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.