APPLICATION OF THE CROSS BATTERY APPROACH IN THE ASSESSMENT OF AMERICAN INDIAN CHILDREN: A VIABLE ALTERNATIVE

Gary A. Plank, Ph.D.

Abstract: This article examines current psychometric and testing practices that appear to do a limited job of assessing the intelligence of American Indian individuals. For several reasons, contemporary approaches are found to be inadequate. Unfortunately, these practices are then employed in making educational decisions and placing these same children into Special Education programs. Alternative methods of testing, including the Gf-Gc Cross-Battery Approach, are discussed and evaluated in terms of usefulness in the evaluation of American Indian children and adolescents. The cross-battery method of testing was found to provide more in-depth procedures for bypassing both language and cultural differences among American Indian individuals.

The question of how to appropriately and accurately test the cognitive ability of American Indian children/adolescents, or children limited in English (Limited English Proficient or LEP) is a burning issue both in the Southwest, as well as the rest of the country. Training programs for psychologists, the American Psychological Association (APA), and the National Association of School Psychologists (NASP) require appropriate and culturally sensitive delivery of services (American Psychological Association, 1990; National Association of School Psychologists, 1992, 1997). In particular, when evaluating American Indian individuals, perhaps the primary question should be, "Is it possible to assess the intelligence of American Indian individuals in an unbiased and accurate manner?" While several sources remind us of our responsibility as psychologists and mental health professionals to do just this, few provide us with the pragmatic tools necessary to accomplish this task, including formal training programs (Flanagan, McGrew, & Ortiz, 2000; Ortiz, Ortiz, & Cook-Morales, 1994). The psychologist is warned not to be naïve in these matters because *no* test is either completely unbiased, or "culture free." According to Sattler (1992):

> Probably no test can be created that will entirely eliminate the influence of learning and cultural experiences. The test content and materials, the language in which the questions are phrased, the test directions, the categories for classifying the responses, the scoring criteria, and the validity criteria are all culture bound... in fact, all human experience is affected by the culture, from prenatal development on. (p. 579)

No matter what the promotional ads claim, or what the author fervently proclaims, there is no such test. It would be very convenient if such an assessment tool existed, but it does not. In essence, intelligence is embedded and defined by the specific culture in which it is to be measured (Flanagan & McGrew, 1998; Flanagan, McGrew, & Ortiz, 2000; Sattler, 1992). Cultural influence is very difficult to remove from an intelligence test and unfairly penalizes culturally-different individuals because they do not have the same cultural experience as mainstream individuals (Cervantes, 1988; McShane & Plas, 1982, 1984; Mishra, 1982; Valdes & Figueroa, 1994). For example, Flanagan, McGrew, and Ortiz (2000) report:

> In order for such (test) bias to be controlled, it should be noted that tests of intelligence and cognitive ability developed and normed in the United States will likely measure a lower range of ability in diverse individuals because they apparently fail to sample "cultural content that is part of the cognitive repertoire and processes available to the bicultural individual." Hence, individuals coming from different cultures are often inaccurately assessed and the derived intelligence score is typically suppressed by such tests and poor decisions follow regarding diagnosis of culturally-different children. (p. 299)

A thorough literature review reveals that American Indian children from tribes across North America, score similarly to their Caucasian, Black, and Hispanic counterparts on the Performance Scale from the Wechsler Intelligence Scale for Children (Wechsler, 1991) (WISC, WISC-R, or WISC-III) (McClellan & Walton, 1996; McCullough, Walker, & Diessner, 1985; McShane, 1980; McShane & Plas, 1982, 1984; Naglieri, 1982; St. John, Krichev, & Bauman, 1976; Tanner-Halverson, Burden, & Sabers, 1993; Teeter, Moore, & Peterson, 1982). On the other hand, these same children uniformly score much lower on the Verbal Scale. A twenty-point differential is typically found in the literature by the authors cited above. Interestingly, *all* researchers mentioned above were measuring intelligence by means of the Wechsler scales. However, the question begs to be asked, is this the best tool to be using when measuring the intelligence of American Indian children? There is evidence that the Wechsler scales may not be the best assessment tool for culturally-different children (Flanagan & Miranda, 1995; Lopez, 1997; McLellan & Walton, 1996; Ochoa, Powell, & Robles-Pina, 1996) and may include item bias for American Indian children (Mishra, 1982).

As an intern in 1994, I was confounded by the intelligence results I was collecting from the American Indian children I assessed using the Wechsler scales. These children's Full Scale scores were typically low and their Verbal Scale scores were sharply lower, normally about 18-20 points, than their Performance Scale scores, just as previous research suggested (McCullough, Walker, & Diessner, 1985; McShane, 1980; McShane & Plas, 1982). Neuropsychologists in the community commonly believed such children were all learning disabled. However, it seemed improbably that *all* of the American Indian children I tested were learning disabled, or cognitively impaired (mentally retarded). My supervisors could not shed light on the situation, thus I turned to Michael Gerner, Ph.D., a school psychologist from Flagstaff, Arizona who has specialized in conducting cognitive assessments with American Indian children and LEP children of all ethnicities. Dr. Gerner and I worked together on a reservation several years earlier. Dr. Gerner suggested several causes for the pattern of intelligence scores I observed among the American Indian children I had tested (M. E. Gerner, personal communication, November 15, 1994). First, he suggested that these were typical patterns found among the American Indian populations, as well as among other LEP children of all nationalities. He suggested that differences of 20 points between scales were common. He advised using tests other than the Wechsler Scales in order to better assess these children. The WISC/WAIS (Wechsler Adult Intelligence Scale) (Wechsler, 1981) were supplemented and sometimes replaced with other tests, like the Stanford-Binet-Fourth Edition (Thorndike, Hagen, & Sattler, 1986), Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983), and Kaufman Adult Intelligence Test (Kaufman & Kaufman, 1993). These tests were *sometimes* found to be superior to the WISC because there were measures on them that were less verbally and culturally influenced, but there continued to be weaknesses with each separate measurement tool and none seemed to be thoroughly comprehensive.

Further consultation, training, and intense discussions with Dr. Gerner over the past six years have proved to be fruitful in terms of improved assessment of American Indian individuals. The discussions began to center around the question of the adequacy of using *any* one intelligence battery to assess these children (M. E. Gerner, personal communication, January 20, 1997). It became abundantly clear that the answer to this query was an emphatic, "no." No single test battery was sufficient for this type of testing because of complicating issues of language and cultural difference (McGrew & Flanagan, 1998). Our discussions, more and more frequently, began to revolve around the possible use of the Gf-Gc Cross-Battery Approach and its applicability for use with American Indians. In 1996, the

Gf-Gc Cross-Battery Approach began to emerge as a means to operationalize the Cattell-Horn-Carroll Gf-Gc theory (Flanagan & McGrew, 1997). In essence, this is a model of intelligence that has been statistically proven to be valid and has proven, through research, that there are several different types of intelligence, not one or two (Flanagan & McGrew, 1997; McGhee, 1993; McGrew, 1997; Woodcock, 1990).

The movement really began with Woodcock's seminal research involving numerous subtests from seven major intelligence tests (Woodcock, 1990). He analyzed all subtests from the major batteries in order to provide a framework to understand the complexity of human cognitive ability. He did not find a single overall G, or broad intelligence, and he did not discover a verbal and performance factor. Rather, he discovered a wide array of human abilities. This discovery, in turn, has led numerous researchers to begin viewing intelligence much differently than previously held notions and has led the way for the Gf-Gc Cross-Battery Assessment (Flanagan & McGrew, 1997).

As opposed to the common conception of the theory, it is not a two-factor theory involving only fluid intelligence (Gf) and crystallized intelligence (Gc). In addition, it is not a dichotomous method of viewing intelligence. Conversely, the theory views intelligence as being multi-faceted and complex and involving numerous abilities (Horn, 1991; McGrew, 1997; McGrew, Flanagan, Keith, & Vanderwood, 1997; Woodcock, 1990). An involved discourse in the theory under-girding the cross-battery approach will not be attempted in this article, but the interested reader is encouraged to review the Cattell-Horn-Carroll Gf-Gc Theory and Cattell-Horn-Carroll Integrated Model (Flanagan, McGrew, & Ortiz, 2000; Horn, 1991, 1994; Horn & Noll, 1997). Briefly, the Cattell-Horn-Carroll Integrated Model outlines the various types of human cognition that are currently quantifiable. In turn, the cross-battery approach provides an outline of which subtests from various intelligence tests are able to best measure the various types of intelligence.

The Cattell-Horn-Carroll model is described by Carroll (1993) as being, "...a true hierarchical model covering all major domains of intellectual functioning...among available models it appears to offer the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities" (p. 62). This line of research has discovered instead of one or two types of intelligence (e.g., verbal and performance), that there are ten distinct types of intelligence (Flanagan, Genshaft, & Harrison, 1997; Horn, 1991; McGrew, Flanagan, Keith, & Vanderwood, 1997; Woodcock, 1990).

The ten areas identified by the Cattell-Horn-Carroll model and supported by Woodcock's research are Crystallized Intelligence (Gc), Fluid Intelligence (Gf), Short-term Acquisition and Retrieval (Gsm), Visual Intelligence (Gv), Auditory Intelligence (Ga), Long-term Storage and Retrieval (Glr), Cognitive Processing Speed (Gs), Correct Decision Speed (Gt or CDS), Quantitative Knowledge (Gq), and Orthographic Knowledge

(Grw) (Carroll, 1993). Again, an in-depth description of these broad abilities will not be outlined here, but is provided by Carroll (1997) or Horn (1994). However, succinctly, crystallized intelligence (Gc) measures a person's stored knowledge of culture and the ability to effectively apply this knowledge (Flanagan, McGrew, & Ortiz, 2000). This type of skill is primarily verbal or language-based. Fluid intelligence (Gf) as described by Flanagan et al. (2000) "refers to mental operations that an individual may use when faced with a relatively novel task that cannot be performed automatically" (p. 30). Quantitative knowledge (Gq) is the store of acquired mathematical information and ability to manipulate numeric symbols. Grw is reading/writing ability and represents the comprehension of writing and reading. Short-term memory (Gsm) is expressed by an individual's ability to acquire and hold information for brief periods of time, just a few seconds. This is distinguished from long-term storage and retrieval (Glr). Glr requires the fluent retrieval of previously learned information to be recalled and related from long-term storage (Flanagan et al., 2000). Flanagan et al. (2000) conceptualize visual processing or (Gv) as "the ability to generate, perceive, analyze, synthesize, store, retrieve, manipulate, transform, and think with visual pattern and stimuli" (p. 42). Auditory processing (Ga) "are cognitive abilities that depend on sound as input and on the functioning of our hearing apparatus" (Stankov, 1994, p. 157). The mental speed or quickness is also referred to as attentive speediness and is labeled processing speed (Gs). It is "the ability to fluently and automatically perform cognitive tasks, especially when under pressure to maintain focused attention and concentration" (Flanagan, et al., 2000). Similar to Gs, Gt measures the quickness in reaction time. The distinction between Gs and Gt is the time required: While Gs requires two or three minutes of sustained attention, Gt is measured in seconds (Woodcock, 1990). Two of the broad areas are covered in the typical academic achievement test, those being reading/writing (Grw) and quantitative reasoning (Gq). Again, for a full review of these categories and specific subtests that measure each one, the reader is referred to Flanagan and McGrew (1997).

Currently, the only broad area that is not measured in any typically used battery is correct decision speed or Gt. That leaves seven broad cognitive areas that could be assessed. Therefore, intelligence cannot be viewed as one global score or two Wechsler Scales (e.g., verbal or performance): Rather, it is diverse and represents numerous different ways for the individual to process information.

American Indian children perform as well on measures of Gv and Gs as their counterparts across the United States (McCullough, Walker, & Diessner, 1985; McShane & Plas, 1982: Naglieri, 1982). This is the case as long as there are not further processing deficits that may suppress one of these scores, such as a visual-motor integration deficit. Hence, it is often imperative that the psychologist isolate these types of intelligence, Gv and Gs, apart from other variables in order to most accurately measure the ability of these children. The cross-battery approach allows for just this type of isolation of skills.

Research reveals that the WISC-III does an adequate job of testing two broad cognitive areas, those being Comprehension-Knowledge (Gc) and Visual Processing (Gv) with some of the subtests (McGrew & Flanagan, 1998). On the other hand, it largely ignores, or does not adequately test the other five currently measurable areas of intelligence identified in the Cattell-Horn-Carroll model (Ga, Glr, Gsm, Gs, and Gf). Unfortunately, LEP children generally do not perform well on measures of Gc, for several reasons. These causes include the large influence of language in Gc measurements, different cultural experience, and limited school experience. In such cases, the psychologist is measuring cultural experience or language, not intelligence or ability. Therefore, the school psychologist/clinical psychologist who is using such tests is simply measuring English language proficiency or cultural exposure and these have little to do with intelligence for many American Indian children.

Hence, the Verbal Scale or a Full Scale Score from the WISC-III would exemplify a lowered and unfair measure of a LEP child's ability. This is nothing revolutionary and many school psychologists/clinical psychologists long ago discarded the practice of using the Verbal Scale and Full Scale scores with American Indian individuals. School psychologists began to use the Performance Scale instead, believing that the Performance Scale represented an improved measure of the American Indian child's intelligence because the Verbal Scale was thus discarded and Gc minimized (McCullough, Walker, & Diessner, 1985; McShane & Plas, 1982; Sattler, 1992). Unfortunately, once again, this practice has been found wanting. Research has suggested that the Performance Scale, in whole, does not cleanly measure one type of ability (e.g., Gv), but two, because the Picture Completion and Picture Arrangement subtests on the Performance Scale have been found to be associated with Gc (McGrew, 1997). Therefore, Picture Completion and Picture Arrangement, and not just on the Verbal Scale, of the WISC are solely or partially intertwined with formal learning or experience (Gc). Consequently, the Performance Scale *does not* yield a pure visual processing score (Gv). Research has found that the best two Gv subtests on this scale are Object Assembly and Block Design (Flanagan & McGrew, 1998; McGrew, 1999; Woodcock, 1990). Potentially, the diagnostician using the WISC Performance Scale is measuring not only nonverbal skills, but has lowered the intelligence score by including subtests that are associated with Gc. Therefore, the Performance Scale on the Wechsler scales will consistently represent a lower ability score for many American Indian individuals.

This problem is not confined to the Wechsler scales. Most commonly used batteries have the identical problem (e.g., the Stanford-

Binet and K-ABC). They have not adequately isolated and separated the various forms of intelligence. Why is this so paramount? Because we must ensure that we clearly and distinctly identify a child's ability and not confuse this ability with a cultural or linguistic difference. Otherwise, we run the risk of suppressing these children's intelligence scores. In addition, if the psychologist lets English Language Proficiency (Gc) remain in the intelligence formula for LEP children, the score becomes lowered, biased, and indefensible. In such cases, the ability score will be suppressed and if a discrepancy model is being employed, the discrepancy will likewise also be suppressed. Discrepancy in this case refers to the difference between intelligence and achievement, which is the process that many psychologists use to diagnose learning disabilities. Finally, poor educational decision-making and planning will take place based upon the faulty testing.

Similar to the Wechsler scales, most of our contemporary assessment tools gauge two or three cognitive areas, typically Gc, Gv, and short-term memory (Gsm) (Genshaft & Gerner, 1998). As previously noted, the Cattell-Horn-Carroll Theory identifies *ten* broad and 69 narrow areas. In order to adequately test a broad area, one must collect scores from a minimum of two qualitatively different, narrow abilities. This is the difficulty in such cases as the Digit Span from the Wechsler scales. There is only one subtest on the WISC/WAIS for Gsm, that being Digit Span. Digit Span represents one narrow ability and cannot constitute a broad area in isolation. When only using one narrow ability score, little reliability is garnered. Again, the psychologist is advised to have at least *two* narrow ability scores in order to make judgments on the child. For that reason the WISC yields valid scores for Gc and Gv, because there are two or more subtests representing these broad areas. The Stanford-Binet produces scores on three, or perhaps four broad areas, depending on what source you consult (Woodcock, 1990). Sattler (1988) claims three broad areas are assessed, while Woodcock asserts four. As mentioned above, one broad area is not tested on any major battery, that being Correct Decision Speed (CDS or Gt). The only battery that attempts to investigate the remaining seven broad areas of intelligence is the Woodcock-Johnson-Revised Test of Cognitive Ability (WJ-R) and the recently published WJ-III.

The WJ-R is grounded theoretically in the Cattell-Horn-Carroll Gf-Gc Theory (Horn, 1991; Horn & Noll, 1997) and provides a broad and comprehensive base from which to assess several types of ability. It is the only test battery available which produces ability scores in such broad areas as long-term memory (Glr), auditory processing (Ga), and fluid intelligence (Gf). These are critical domains to measure because each has been found to be highly correlated with various areas of academic achievement (Flanagan & McGrew, 1998). While many test batteries are able to reveal a discrepancy between ability and achievement, they often provide no information concerning the *cause* of the discrepancy. In turn, most tests are also limited in their capacity to provide information regarding remediation for teachers, which should be a primary goal of the psychologist.

The Gf-Gc Cross-Battery Approach suggests that the clinician "cross-batteries" in order to provide the most comprehensive means of assessment (Flanagan, et al., 2000). This does not imply administering several total batteries to the student, but rather, selecting specific *portions* of batteries in order to answer the referral question and more accurately measure intelligence. The evaluator could chose two or three pure Gv subtests from different batteries in order to establish intelligence. In this case, for an American Indian person, Gv or Gs subtests could be selected. For instance, if a school psychologist/clinical psychologist gave the Stanford-Binet to a child with a reading problem, there is little to no information gathered from this test concerning auditory processing (Ga), which has been found to be highly correlated with reading success (Flanagan, et al., 2000). However, if the examiner supplemented the standard battery (the Stanford-Binet) with the Auditory Processing Cluster from the Woodcock-Johnson, then he/she could speak more directly to the possible cause of the reading difficulty (e.g., limited Ga). Additionally, because it has been found that one must collect two narrow subtest scores to constitute a broad category, it is unnecessary to continue testing the same narrow ability with four, five, or more subtests. This appears to be the case with the Wechsler scales. Assuming that the two subtests fall within the same range, only two subtests are required to be given in either Gv or Gc. There is no justifiable cause to include all of the Wechsler Verbal subtests in order to understand an individual's verbal ability. Hence, the number of subtests could be reduced so that overall administration time would not be significantly increased.

This approach allows the examiner to isolate specific abilities and test for a broader spectrum of abilities than the traditional standard battery. If the psychologist is parsimonious with time, he/she is able to carefully select the subtests beforehand, or during testing, which answer the referral question. Therefore, additional time is often not required to complete the testing from a cross-battery approach, or the increased time should amount to no more than ten to fifteen minutes.

At present there is no one test on the market that taps into all the broad categories of ability. Until such a test is produced, the vigilant psychologist will need to conduct a broad-based assessment of his or her own. As advocated by Genshaft and Gerner (1998), "...researchers and clinicians will need to 'cross' batteries in order to conduct complete assessments of cognitive functioning or *selective but in-depth* assessment of particular abilities (e.g., Glr, Gf) that have been found to underlie specific achievements" (p. 25). There are ever increasing possibilities on the horizon, including the new Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2000), which was released the fall of 2000. The Universal Nonverbal

Intelligence Test (UNIT) (McCallum & Bracken, 1998), Leiter International Performance Scale-Revised (Roid & Miller, 1997), Comprehensive Test of Nonverbal Intelligence (CTONI) (Hammill, Pearson, & Wiederholt, 1996), or Test of Nonverbal Intelligence (TONI) (Brown, Sherbenou, & Johnson, 1997) are nonverbal tests which may provide good nonverbal or visual scores for American Indian children, but there is no current research on these tests. Such nonverbal tests may hold some promise for testing culturally-different individuals. However, in the near future, it continues to be unlikely that one test will accurately measure the intelligence of culturally-different individuals. We should also continue to challenge the test developers to devise new and improved tools of measurement that do a better job of assessing culturally different children. In addition, there is a need for the effectiveness of the cross-battery to be fully applied and researched with American Indians.

The cross-battery is not a panacea and the psychometric field will inevitably continue to evolve. Eventually, there may actually be one battery that “does it all,” but this laudable goal may also be elusive since research will continue to refine cognitive abilities that can be assessed and test development will hasten to “catch up.” In the meantime, we must make do with our current resources. Presently, for this psychologist, the cross-battery approach has proven to be the best method to accurately assess American Indian children and adolescents. It is an invaluable process that improves testing and provides a more thorough evaluation of the student, it has not significantly added time to the testing process, and it allows for an increased capacity to address referral questions. While the cross-battery has been effective for testing American Indian and ESL students, this approach applies to all individuals, and is not confined to culturally and linguistically different students.

The original question posed by this article concerning whether or not it was possible to accurately test the intelligence of American Indian children can be answered. It is the firm belief of this psychologist that the answer is “yes.” Are we able to provide accurate scores by using one single test battery that was developed and normed with Caucasian children? “No,” and frequently this method of testing is found lacking for such children. Over the past year, I have been requested to conduct a second opinion or reevaluation for eight Intellectually Disabled (mentally retarded) American Indian children. Original intelligence scores from the WISC-III and WAIS-R ranged from 58 to 69; the criterion is 70 for classification with this disorder. All eight children and adolescents had previously been tested with the WISC or WAIS, using either the Full Scale or Performance Scale. The reevaluations employed either the cross-battery approach, or pure measures of Gv/Gs to establish intelligence (Visual Processing from the Woodcock-Johnson or the UNIT). Reevaluations of these children produced average intelligence scores (i.e., between 90 and 105). *These scores indicate that the children were not mentally retarded, nor was their* *intelligence impaired*. Clearly these children had been misdiagnosed, misplaced, and further, inappropriately educated. Two had been mislabeled and placed in a "self-contained" class for several years. In essence, these children's civil rights had been violated and this process is legally indefensible. It is the obligation of every school psychologist/clinical psychologist who works with American Indian children and adolescents to utilize the most accurate and up-to-date means available to measure intelligence and make special education decisions concerning placement and remediation. Otherwise, poor decisions and inaccurate diagnoses regarding the children we serve will inevitably be made. In turn, our decisions are tantamount to damaging the children we aspire to aid.

Gary A. Plank, Ph.D.
3360 E. Linda Vista Dr.
Flagstaff, AZ 86004

References

American Psychological Association. (1990). *Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations,* Washington, DC: Author.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1997). *Test of Nonverbal Intelligence-Third Edition.* Austin, TX: Pro-Ed.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D.P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issue* (pp. 122-130). New York: Guilford.

Cervantes, H. T. (1988). Nondiscriminatory assessment and informal data gathering: The case of Gonzaldo L. In R. L. Jones (Ed.), *Psychoeducational assessment of minority group children: A casebook* (pp. 239-256). Berkeley, CA: Cobb & Henry.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 314-325). New York: Guilford.

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory: Joint confirmatory factor analyses of the WJ-R and KAIT in a non-White sample. *Journal of School Psychology, 36,* 151-182.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligences scales and Gf-Gc theory: A contemporary approach to interpretation*. Needham Heights, MA: Allyn & Bacon.

Flanagan, D. P., & Miranda, A. H. (1995). Working with culturally different families. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 1039-1060). Washington, DC: National Association of School Psychologists.

Genshaft, J. L., & Gerner, M. E. (1998). Gf-Gc Cross-Battery Assessment: Implications for school psychologists. *Communiqué',* 26(8), 24-27.

Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (1996). *Comprehensive Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson Technical Manual* (pp. 197-232). Chicago: Riverside.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 443-451). New York: Macmillan.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities. Gf-Gc theory. In D. Flanagan, J. L Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

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

Kaufman, A. S., & Kaufman, N. L. (1993). *The Kaufman Adolescent and Adult Intelligence Test.* Circle Pines, MN: American Guidance Service.

Lopez, E. C. (1997). The cognitive assessment of limited English proficient and bilingual children. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 506-516). New York: Guilford.

McCallum, R. S., & Bracken, B. A. (1998). *Universal Nonverbal Intelligence Test.* Chicago: Riverside.

McCullough, C. S., Walker, J. L., & Diessner, R. (1985). The use of Wechsler scales in the assessment of Native Americans in the Columbia River Basin. *Psychology in the Schools, 22*, 23-38.

McGhee, R. (1993). Fluid and crystallized intelligence: Confirmatory factor analysis of the Differential Abilities Scale, Detroit Tests of Learning Aptitude-3, and Woodcock-Johnson Psycho-Educational Battery-Revised. *Journal of Psychoeducational Assessment Monograph Series: WJ-R Monograph,* 20-38.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L., Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 159-180). New York: Guilford.

McGrew, K. S. (1999). The Wechsler freedom-from-distractibility index: A tale of three subtests. *Communiqué, 27*(8), 24.

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross battery assessment.* Boston: Allyn & Bacon.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretations intelligence test batteries in the schools. *School Psychology Review, 26,* 189-210.

McLellan, M. J., & Walton, M. J. (1996, August). Concurrent validation of the Leiter-R and WISC-III with Navajo children. Paper presented at the meeting of the American Psychological Association, Toronto, Ontario, Canada.

McShane, D. (1980). A review of scores of American Indian children on the Wechsler intelligence scales. *White Cloud Journal, 1,* 3-10.

McShane, D., & Plas, J. M. (1982). Wechsler scale performance patterns of American Indian children. *Psychology in the Schools, 19,* 8-17.

McShane, D., & Plas, J. M. (1984). The cognitive functioning of American Indian children: Moving from the WISC to the WISC-R. *The School of Psychology Review, 13,* 61-73.

Mishra, P. (1982). The WISC-R and evidence of item bias for Native American Navajos. *Psychology in the Schools, 19,* 458-464.

Naglieri, J. A. (1982). Does the WISC-R measure verbal intelligence for non-English speaking children? *Psychology in the Schools, 19,* 478-479.

National Association of School Psychologists. (1992). *Standards for the provision of school psychological services.* Silver Spring, MD: Author.

National Association of School Psychologists. (1997). *Principles for professional ethics.* Bethesda, MD: Author.

Ochoa, S. H., Powell, M. P., & Robles-Pina, R. (1996). School psychologist's assessment practices with bilingual and limited-English-proficient students. *Journal of Psychoeducational Assessment, 14,* 250-275.

Ortiz, S. O., Ortiz, O. G., & Cook-Morales, V. J. (1994). *Preliminary analysis: Survey of California school psychologists listed in the CASP Multilingual Directory.* Paper presented at the CASP Multicultural Affairs Committee Meeting and Workshop at the annual conference of the California Association of School Psychologists, Long Beach, CA.

Roid, G. H., & Miller, J. (1997). *The Leiter International Performance Scale-Revised Edition*. Wood Dale, IL: Stoelting.

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

Sattler, J. M. (1992). *Assessment of children* (Rev. and updated 3rd edition). San Diego, CA: Sattler.

Stankov, L. (1994). Auditory abilities. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 157-162). New York: Macmillan.

St. John, J., Krichev, A., & Bauman, E. (1976). Northwestern Ontario Indian children and the WISC. *Psychology in the Schools, 13*, 407-411.

Tanner-Halverson, P., Burden, T., & Sabers, D. (1993). WISC-III normative data for Tohono O'Odham Native American children (Monograph). *Journal of Psychoeducational Assessment,* 125-133.

Teeter, A., Moore, C. L., & Peterson, J. D. (1982). WISC-R verbal and performance abilities of Native American students referred for school learning problems. *Psychology in the Schools, 19,* 39-44.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition.* Chicago: Riverside.

Valdes, G., & Figueroa, R. (1994). *Bilingualism and testing: A special case of bias.* Norwood, NJ: Ablex Publishing.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2000). *Woodcock-Johnson Psycho-Educational Battery-Third Edition*. Chicago: Riverside.