Assessment of Hearing-Impaired and Deaf Children with the WISC-III

Jeffery P. Braden and Joseph M. Hannah

Department of Educational Psychology
University of Wisconsin—Madison
Madison, Wisconsin

The WISC-III in its various editions is the most widely used test of intelligence with hearing-impaired and deaf children in North America. In fact, it is probably the most widely used test of intelligence with deaf children in the world. The Wechsler scales are extremely popular with clinicians and researchers interested in measuring the intellectual abilities of deaf children. Furthermore, experts in deafness consistently recommend the Wechsler scales, and in particular, the Performance Scales, as being among the best ways to assess intelligence. Indeed, the Performance Scale of the Wechsler in all its versions has generated more studies on deaf children than any other measure of intelligence (Braden, 1994).

Therefore, it is no surprise that new versions of the Wechsler should attract substantial attention among those who are interested in its use with deaf children. This is true for the WISC-III (Wechsler, 1991). The WISC-III is the first Wechsler scale to include a clinical study of the use of the test with deaf students (see Wechsler, 1991, p. 216). Furthermore, this clinical study was elaborated in detail in a subsequent publication (Maller & Braden, 1993). Therefore, even before its publication, the WISC-III was likely to be a tool for assessing the intelligence of deaf children.

We realize that American Psychological Association style recommends person-first usage (e.g., children who are deaf or hard-of-hearing). However, we use the term “deaf children” to be consistent with the way in which deaf people in North America define themselves. Consequently, we refer to children who have normal hearing as “normal-hearing children.” We hope that our usage does not offend or confuse readers.

WISC-III Clinical Use and Interpretation: Scientist-Practitioner Perspectives
Copyright © 1998 by Academic Press. All rights of reproduction reserved.
This book is intended to address practitioner and scientific perspectives. The professional literature presents both perspectives when describing the use of the Wechsler scales in deaf populations. The first perspective is that of the practitioner. That literature explores the use of the Wechsler scales for assessing deaf children's intelligence, and for differentially diagnosing cognitive deficits within deaf children. For example, substantial literature is available on "profile analysis" with earlier versions of the Wechsler (Braden, 1990). Additionally, there is substantial literature on how the Wechsler Performance Scale can differentially diagnose average, gifted, and mentally retarded deaf children. The primary clinical purpose of the Wechsler is to differentially diagnose the impact of deafness from the possible impact of mental retardation. That is, most deaf children also have significant language delays and deficits in acquired knowledge and other culturally relevant (i.e., crystallized abilities) cognitive skills. Experts frequently argue that these deficits are due not to a lack of intellectual ability, but to lack of opportunity to acquire the dominant language. Therefore, deaf children appear to be retarded if one only examines traditional measures of language development, academic achievement, and the like. However, deaf children as a whole appear to have average intellectual abilities when assessed using language-reduced measures of cognitive ability (Braden, 1994; Vernon, 1967). Therefore, the primary use of the Wechsler scales is to differentially diagnose mental retardation from deafness as a cause of poor academic and language performance in deaf children. Additionally, the Wechsler is occasionally used to diagnose learning difficulties within deaf children. That is, just because a child is deaf does not mean the child does not have additional cognitive deficits that might interfere with learning. This tradition has historically been pursued with the Wechsler through the use of profile analysis, in which it has been found that the Coding subtest is especially sensitive to academic deficits in deaf children (Braden, 1990; Ensor, 1988).

In contrast, the research or empirical tradition of work with deafness has focused on the adequacy of the Wechsler scales from a technical point of view. In other words, there is a fair amount of research attempting to demonstrate the reliability and validity of the Wechsler scales (again, primarily the Performance Scales) for their use with deaf populations. These studies of previous versions of the Wechsler have been summarized elsewhere (see Braden, 1992, 1994). The most important feature of this research is that the previous versions of the Wechsler have generally been found to have adequate technical characteristics when applied to populations of deaf people. However, this general conclusion can be challenged on some specific issues including the variability of correlations between IQ and measures of achievement.

All the issues addressed in previous versions of the Wechsler remain as important issues for the WISC-III. That is, there is still a strong interest in clinical traditions and in particular how effective the WISC-III will be for differentially diagnosing the impact of hearing impairment from the potential impacts of conjoint cognitive impairments. Secondarily, there will be interest in the degree to which Wechsler subtest profiles will be able to be used to differentially diagnose (or be sensitive to) academic deficits in deaf children.
Likewise, all of the issues about the technical integrity of the WISC-III are critical and must be reestablished. Although it is a good bet that many of the characteristics that were true for previous versions of the Wechsler will be true for this version, findings must be replicated to establish adequate reliability and validity for the WISC-III. Therefore, our chapter will address clinical and research issues that have been critical to the use of Wechsler scales with deaf children. In particular, we will identify those issues we see as critical and will bring to bear available research using the WISC-III to address them. We will often illustrate the issue or describe it by citing research on previous versions of the Wechsler scales.

**ADMINISTRATION ISSUES**

**METHODS OF TRANSLATION**

Wechsler (1991) cautions examiners that deviations from the standard administration procedures can reduce the validity of test results. Given that caution, however, the WISC-III manual states that “some flexibility may be necessary to balance the needs of the particular child with the need to maintain standard procedures” (p. 38). Regarding children who are deaf, the manual suggests that modifications, such as translation of the test into signs or the use of additional visual aids, be noted on the record form. This will allow those who are evaluating the child’s functioning to weigh the impact of any modifications made.

When assessing deaf or hard-of-hearing children whose primary mode of communication is American Sign Language (ASL), Pidgin Signed English (PSE), or a sign system such as Signing Exact English (SEE), psychologists have an ethical responsibility to administer tests in the child’s primary language. Ideally, the psychologist would be skilled in the method of sign language used by the child. When this is not feasible, a certified sign language interpreter may be secured to interpret instructions to the child. Examiners should observe the child in various settings and consult with the child’s parents and teacher prior to assessing a child who is deaf or hard of hearing to identify the child’s communication needs.

The most common procedure for administering the Wechsler scales is to translate the directions from speech to another (visual) form to allow the deaf child to understand the tasks at hand. In some cases, the directions are supplemented with gestures or a written transcription of directions. In other cases, the directions are given concurrently in sign language and speech. The range of methods used to supplement directions include the following:

- careful oral enunciation
- supplementing oral directions with written transcripts of the directions
- finger-spelling directions
- signing directions in English signs
- concurrently saying the directions orally and signing them in an English-type sign dialect
- translating directions into ASL without concurrent voice
• gestures
• supplemented examples/demonstrations of subtests.

The most popular of these adaptations has historically been concurrent presentation of task directions in voice and sign (Braden, 1992, 1994).

The impact of administrative adaptations varies. Some adaptations produce higher scores in deaf children than do others. Literature reviews find concurrent administration of Wechsler subtests using voice and sign directions as yielding higher PIQs than any of the other methods (Braden, 1992, 1994), although any individual study may fail to achieve statistical significance for the comparison (e.g., Sullivan & Montoya, in press). Generally, oral administration (and oral administration supplemented with written directions) produces lower Performance IQs (PIQs). Most experts encourage examiners to use concurrent voice and sign administrations, although some experts (e.g., Ray, 1982) also recommend supplemental examples and demonstrations.

Often experts recommend simultaneous administration of directions in voice and signs because this method produces higher scores. This is incorrect reasoning from a research perspective, as one cannot assume higher scores are necessarily “better” at estimating deaf children’s intelligence. In other words, it would probably also be true that if we gave deaf children the answers to questions before giving them the test they would get higher scores. Yet few would argue that such an adaptation would be appropriate for assessing deaf children’s intelligence. Therefore, the finding that deaf children get higher IQs when given tests concurrently in voice and signs is insufficient to support its use. Advocates of concurrent voice and sign methods draw on another argument to support the use of that administration method. That argument is that deaf children should be assessed in their native communication or dominant communication method. Because most deaf children in North America are instructed using a simultaneous approach (concurrent presentation of information in voice and signs), it would make sense that they should also be assessed in a concurrent voice and sign approach. However, this may be changing. The bilingual/bicultural movement in deafness is encouraging educators to teach deaf children using ASL before introducing English. In situations where deaf children are fully immersed in an ASL environment, it would make the most sense to administer the WISC-III using ASL directions. That is, the psychologist would administer the test without using concurrent voice, as ASL cannot be concurrently voiced in English.

A final note on adaptations for administration is the use of interpreters. There has been very little systematic research on the impact of interpreters in the assessment situation. However, the available research (e.g., Sullivan, 1978; Sullivan & Montoya, in press; Sullivan & Schulte, 1992) suggests that interpreter use generally yields scores equal to those that are produced when the test is administered by a psychologist who is fluent in sign and voice communication. That is, when the psychologist gives the directions orally and an interpreter translates those oral directions into sign, deaf children generally do about as well as deaf children who
get the tests administered by a psychologist fluent in sign language. However, the impact of interpreters on assessment outcomes is not well understood, especially with respect to ASL. Most previous studies use signed English systems rather than ASL. Preliminary evidence suggests that deaf psychologists who administer the WISC-III to normal-hearing children can use an interpreter and get reasonably accurate results (Kostrubala, 1996). Likewise, the accuracy of an ASL version of the WISC-III was demonstrated using blind back-translation (Mailer, 1996). These results are encouraging in suggesting that careful, comprehensive procedures for translation can provide accurate ASL versions of the Wechsler scales. However, with one exception (Mailer, 1996), none of the previous research has explicitly defined the procedures used to translate the Wechsler scales from English to ASL. The process should follow procedures for translating tests between oral languages, and include blind back-translation accuracy checks. The process for effective translation is provided in Figure 9.1 (from Kostrubala, 1997).

NONLINGUISTIC ADAPTATIONS

Sattler (1992) summarizes many nonlinguistic modifications of the WISC-III recommended for use for children who are deaf. On the Verbal Scale, these include typing the questions for the Information, Comprehension, Similarities, and Vocabulary subtests, and allowing written or typed responses if the child cannot respond orally. He warns that visually administering the Arithmetic and Digit Span subtests may be difficult due to the nature of these tasks. Modifications for the Performance Scale are presented in Appendix D (Sattler, 1992; pp. 896–899). These include pantomime and various visual adaptations to subtest instructions. Examiners are warned, however, that the use of pantomimed instructions or visual aids may produce lower PIQs than those obtained from sign language administration (Braden, 1985a; Sullivan, 1982). Sattler (1992) too cautions that scores obtained using these modifications should be viewed only as estimates of what the child's score would be had standardized procedures been followed. Table 9.1 summarizes the various modifications to the Wechsler subtests for administration to deaf children. Note that few of these recommendations are recent; the trend in the past 10–20 years has been to translate the test directions into the child's primary language rather than emphasize nonlinguistic modifications to subtest administration.

TECHNICAL ADEQUACY OF THE WISC-III WITH DEAF CHILDREN

RELIABILITY

The reliability of the Wechsler scales has not been as widely studied as other characteristics. However, some previous studies suggest that the internal consistency of the Wechsler scales is similar in deaf and hearing samples. That is, the
Form committee of researcher and 3 ASL users

Translate the test

Recruit 3 new ASL users for new committee

New committee independently blind back-translates the test

Researcher reviews test to determine 67% or greater agreement on one item

Review blind back translator's feedback

Retain translation

Pilot test with deaf participants

Pilot Test Successful

Field Test
### TABLE 9.1 Nonlinguistic Modifications to Wechsler Performance Scale Subtests for Deaf Children

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture Completion (Neuhaus, 1967)</td>
<td>Three sample pictures are drawn on separate cards, a complete picture on one side and the same picture with a missing detail on the other. Each card is presented by showing the incomplete side first, then the complete side while indicating the detail.</td>
</tr>
<tr>
<td>Coding (Murphy, 1957)</td>
<td>Examiner models completion of sample items using Coding protocol from previous version of the Wechsler. A card with the words “Do this quickly” is laid beside the child’s protocol.</td>
</tr>
<tr>
<td>Picture Arrangement (Reed, 1970)</td>
<td>Two sets of three cards are used, one with the numbers 1, 2, and 3, and one with the letters, A, B, and C, as well as the sample item. Cards are presented out of sequence and the examiner models arranging them in order then motions for the child to do the same.</td>
</tr>
<tr>
<td>Block Design (Murphy, 1957)</td>
<td>Four blocks are laid out in a line each with the red side on top, followed by the white, and red/white sides. Designs are then presented as is a card with the words “Make one like this.”</td>
</tr>
<tr>
<td>Object Assembly (Neuhaus, 1967)</td>
<td>Three pictures are drawn on cards representing the sample item and first two test items. Pieces are arranged behind the shield and when exposed, the corresponding picture is laid next to the pieces as a model.</td>
</tr>
<tr>
<td>Mazes (Sullivan, 1978)</td>
<td>Examiner models completion of sample items using Mazes protocol from previous version of the Wechsler. A card with the words “Start in the middle. Find your way out” is laid beside the child’s protocol.</td>
</tr>
</tbody>
</table>


Wechsler Performance Scale subtest scores and PIQ are equally precise in deaf and hearing samples (Braden, 1994). Additional research on the stability of previous versions of the Wechsler suggest good stability over 1 year to multiyear intervals (Braden, Maller, & Paquin, 1993). However, carefully controlled test–retest research is not available on either previous versions of the Wechsler or the WISC-III. Therefore, the stability of WISC-III IQs and subtests is unknown. We will discuss comparisons between the WISC-R and WISC-III with deaf children in the following (Validity) section.

Estimates of internal consistency suggest that the WISC-III is equally accurate for deaf and normally hearing children (Maller & Braden, 1993). Unfortunately, there have been no replications of internal consistency studies with the WISC-III using deaf participants. We conclude that the available evidence does not show reliability differences when the WISC-III is used with deaf children, but we also note there is not much evidence to show that reliabilities are the same. The best verdict for the case of whether WISC-III reliability is consistent within deaf children is the Scottish Law verdict of “not proven.” Scientists and practitioners will need additional replications of reliability studies, particularly for temporal stability, before they can adequately judge the reliability of the WISC-III with deaf children.
### TABLE 9.2  Mean WISC-III Scale and Index Scores from Studies of Deaf Children

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>VIQ M</th>
<th>PIQ M</th>
<th>FSIQ M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wechsler (1991)</td>
<td>30</td>
<td>81.10 (20.30)</td>
<td>105.80 (20.80)</td>
<td>92.20 (19.80)</td>
</tr>
<tr>
<td>Braden, et al. (1994)</td>
<td>19</td>
<td>81.63 (18.74)</td>
<td>102.32 (11.03)</td>
<td>90.68 (12.54)</td>
</tr>
<tr>
<td>Slate &amp; Fawcett (1995)</td>
<td>47</td>
<td>88.00 (18.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sullivan &amp; Montoya (1997)</td>
<td>106</td>
<td>75.35 (17.55)</td>
<td>100.63 (19.48)</td>
<td>86.22 (17.37)</td>
</tr>
</tbody>
</table>

*Values enclosed in parentheses represent standard deviations. VIQ = Verbal Intelligence Quotient; PIQ = Performance Intelligence Quotient; FSIQ = Full Scale Intelligence Quotient.*

### VALIDITY

The WISC-III manual (Wechsler, 1991, p. 216) presents one study regarding the validity of the WISC-III with a sample of deaf children. The sample, as described by Maller and Braden (1993), consisted of 30 children who had severe to profound hearing losses. Etiology of hearing loss was attributed to unknown or suspected genetic causes for the majority of the sample. The median age at which hearing loss was diagnosed was 2 years. Twenty-seven children had parents with normal hearing, and 3 had parents who are deaf. Administration procedures were modified by using ASL or PSE, depending on the child’s primary mode of communication. The results indicated mean scale and index scores in the expected direction (e.g., PIQ > VIQ). These results provide tentative evidence that the WISC-III, much like its predecessors, is a useful measure with deaf children (Wechsler, 1991).

Since the publication of the WISC-III, other studies have investigated its validity with samples of deaf children. First, we examine the mean Scale and Index scores reported in various studies. Table 9.2 presents Scale and Index means and standard deviations.

The data in Table 9.2 replicate previous studies of the Wechsler Scales by noting that deaf children consistently demonstrate lower scores on subtests from the Verbal Scale (i.e., VIQ, VC, FFD) than scores drawn from Performance Scale subtests (i.e., PIQ, PO, PS). The exception to this generalization is the Processing Speed (PS) Index score. Research with previous Wechsler tests found a pattern of lower Coding scores (Braden, 1990), thus raising the question of lower speed of information-processing abilities. The current research is inconsistent in defining whether lower scores are limited to the Coding subtest (e.g., Maller & Braden, 1993; Sullivan & Montoya, in press) or generalize to the PS factor (Slate & Fawcett, 1995).
The second issue we address is the variability among WISC-III subtest scores reported in these studies. Figure 9.2 presents the means and standard deviations of WISC-III subtests from those studies and depicts the mean scores for each study as a graphic psychometric profile.

There are two striking features of Figure 9.2: (a) the mean Performance Scale scores are higher than the mean Verbal Scale scores, and (b) there are substantial differences in average scores among studies. That is, all studies that included both Verbal and Performance Scales reported higher Performance Scale subtest means, but among the Performance subtests, the most striking variability is between studies, not between subtests. Thus, it is difficult to identify a "characteristic" psychometric profile for deaf children.

The factor structure of the WISC-III was examined in two studies. Sullivan and Montoya (in press) included all of the WISC-III subtests, whereas Slate and Fawcett (1995) only included Performance Scale subtests. Sullivan and Montoya found two factors, which clearly corresponded to the Verbal and Performance Scales. This replicates previous factor analyses of the WISC-R (Sullivan & Schulte, 1992). In contrast, Slate and Fawcett found Performance subtest loadings consistent with PO and PS factors. It is not clear whether the differences in these studies are due to differences in factor analytic procedures, the variables included in the analyses (i.e., Performance versus Full Scale subtests), or differences in the participants (e.g., Sullivan & Montoya, in press, use a much larger and less heterogeneous cohort than Slate & Fawcett, 1995). Therefore, psychologists can be reasonably certain that Verbal and Performance factors are replicated within samples.
Finally, we turn our attention to relationships between the WISC-III and other tests. Again, the data are limited, but generally supportive of the WISC-III. Slate and Fawcett (1995) examined the relationships between the WISC-III and WISC-R PIQ and subtests, and the relationship between WISC-III PIQ and subscales of the Wide Range Achievement Test-Revised (WRAT-R). The gap between WISC-R and WISC-III testing was about 3 years, whereas the WISC-III and WRAT-R were administered concurrently. Maller and Braden (1993) examined the relationship between WISC-III IQs and subtests of the Seventh Edition of the Stanford Achievement Test, which was normed on a large national sample of deaf children (SAT-HI). Finally, Braden, Reed, and Kostrubala (1994) reported correlations between concurrent administrations of the WISC-III and the Kaufman Test of Educational Achievement (K-TEA). The results of these studies are summarized in Table 9.3.

These findings generally support the use of the WISC-III with deaf children. By finding strong correlations between the WISC-III and WISC-R, Slate and Fawcett (1995) support generalization of research with the WISC-R to the WISC-III. Correlations between the WISC-III and tests of achievement are also moderate to high. However, the finding that VIQs and Indexes correlate higher with achieve-

---

**TABLE 9.3** Correlations of the WISC-III with Other Psychological Tests

<table>
<thead>
<tr>
<th>Study</th>
<th>WISC-III IQ</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VIQ</td>
<td>PIQ</td>
</tr>
<tr>
<td>Braden et al. (1994)</td>
<td>.48</td>
<td>.14</td>
</tr>
<tr>
<td>K-TEA Math</td>
<td>.60</td>
<td>.16</td>
</tr>
<tr>
<td>K-TEA Reading</td>
<td>.54</td>
<td>-.04</td>
</tr>
<tr>
<td>K-TEA Spelling</td>
<td>.58</td>
<td>.11</td>
</tr>
<tr>
<td>Mailer &amp; Braden (1993)</td>
<td>.80</td>
<td>.46</td>
</tr>
<tr>
<td>SAT-HI Total Reading</td>
<td>.85</td>
<td>.54</td>
</tr>
<tr>
<td>SAT-HI Total Math</td>
<td>.83</td>
<td>.63</td>
</tr>
<tr>
<td>Slate &amp; Fawcett (1995)</td>
<td>.93</td>
<td>.41</td>
</tr>
<tr>
<td>WISC-R PIQ</td>
<td>.64</td>
<td>.64</td>
</tr>
<tr>
<td>WRAT-R Reading</td>
<td>.63</td>
<td>.54</td>
</tr>
<tr>
<td>WRAT-R Spelling</td>
<td>.83</td>
<td>.63</td>
</tr>
<tr>
<td>WRAT-R Arithmetic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*K-TEA = Kaufman Test of Educational Achievement; SAT-HI = Stanford Achievement Test (7th edition) using norms for Hearing-Impaired children; WISC-R = Wechsler Intelligence Scale for Children-Revised; WRAT-R = Wide Range Achievement Test-Revised.*

of deaf children, but it is not clear whether additional Index factors (e.g., Processing Speed) are also present.
ment than PIQs and Indexes suggests a dilemma in the assessment of deaf children. That is, the WISC-III Verbal Scale offers greater predictive accuracy, but confounds language knowledge with the estimation of intelligence. We will return to this issue in our concluding section.

ISSUES IN USING
AND INTERPRETING RESULTS

There are many issues that test users and researchers should address when using the WISC-III with deaf children. Many of these issues are controversial, with experts in stark disagreement. We will address seven issues in this section, including the following:

- Special norms based on deaf children
- VIQ versus PIQ
- Differential item functioning (i.e., item bias) in Verbal subtests
- The mean IQ of deaf children of deaf parents versus deaf children of hearing parents
- Differences in IQ among deaf children in different educational settings
- Differences among other subgroups of deaf children
- Interpretation of Performance Scale profiles

Each of these issues has a history with previous versions of the Wechsler, and we believe these issues will continue to challenge scientists and practitioners who use the WISC-III with deaf children.

SPECIAL NORMS

Many experts in deafness (e.g., Sullivan & Vernon, 1979; Vernon & Brown, 1964) strongly recommend the use of special norms based on deaf children when assessing deaf children. These norms, often called "deaf norms" because they are based on deaf children, are available for at least five intelligence tests (Braden, 1992), including the WISC-R Performance Scale (Anderson & Sisco, 1977). Proponents of deaf norms do not clearly articulate the reasons why such norms are preferable to those based on normal-hearing children, but at least three reasons might be advanced: (a) norms based on normal-hearing children exclude deaf children, and therefore reduce the validity of those norms for deaf children, whereas norms based on deaf children are valid for use with deaf children; (b) deaf norms would better reflect how a deaf child compares to other deaf peers; and (c) deaf norms provide data demonstrating the reliability and validity of a test with deaf children.

The value of these arguments varies. The first reason, that deaf norms are intrinsically more valid than norms that exclude deaf children, is the most popular. However, it is the least legitimate. Believing that representation (or lack thereof) in
a norm group affects the validity of a test for the group is a classic example of the “normative fallacy” (Jensen, 1980, Chap. 9). The validity of a test for a particular group is independent of its norm group. For example, the primary British measure of height was normed on one male Saxon ruler (i.e., a “foot” is literally the length of King Charlemagne’s foot), yet it is a valid measure of height for women, non-Britons, and others not included in the norm group. Conversely, a test of intelligence composed of Swahili vocabulary would be invalid for English-speaking undergraduates, even if the test was normed on English-speaking undergraduates. Therefore, the case for deaf norms cannot be justified by arguing that norms based on deaf children are intrinsically better than norms excluding deaf children. The case for deaf norms must rest on other criteria.

Another criterion often invoked to justify deaf norms is the argument that one should compare deaf children to other deaf children, not to hearing children, to best describe their cognitive abilities. This argument presumes that deaf children have different cognitive abilities than hearing children; otherwise, why use different norms? Ironically, the evidence with respect to PIQ overwhelmingly shows the distribution of PIQ in deaf and hearing populations to have similar parameters (Braden, 1994). This in itself argues against the value of deaf norms (Braden, 1985b). However, even if there were differences between deaf and hearing children, we would argue against the exclusive use of deaf norms. Our reasoning is based on clinical practice and scientific theory. Clinically, special deaf norms would obscure differences between deaf and normal-hearing samples. These differences are useful for identifying additional disabling conditions (see the discussion of profile analysis later in this section). Scientifically, the use of norms to obscure between-group differences may encourage the “egalitarian fallacy” (Jensen, 1980, Chap. 9), in which differences between groups are attributed solely to some form of test bias. That is, there is a temptation to think that, by referencing scores only to deaf children (and thus eliminating the differences in scores between deaf and hearing children), one has successfully removed the bias inherent in the test (see Vonderhaar & Chambers, 1975, for an example of this argument). This is flawed reasoning that serves neither the child nor the research community (see Braden, 1994, for an extensive discussion of this issue).

Finally, one might argue in favor of deaf norms because the data from those norms would delineate the test’s technical adequacy and validity for use with deaf children. In our opinion, this is a logical argument—if data are collected and retained at the item level to allow for reliability, differential item functioning, and other studies. However, few studies that develop special deaf norms collect or retain such data, and none of the studies have conormed achievement batteries (to allow for large-scale criterion validity studies). Also, a single large-scale norm study is more expensive (and statistically less powerful) than a series of independent, small-scale research efforts. Therefore, we question the scientific and practical value of special norms for deaf children. Given the available evidence, we would echo the need for more information on the reliability and validity of the WISC-III with deaf children, but we suspect development of special deaf norms
for the WISC-III would do little to enhance effective assessment, and could possibly erode effective interpretation of test outcomes.

**VIQ VERSUS PIQ**

The PIQ of Wechsler Scales, as well as other language-reduced tests of intellectual abilities, have traditionally been administered to children who are deaf and hard of hearing. As mentioned previously, deaf children as a group have been found to score in the average range on PIQ. Additionally, Braden (1994) reports that the distribution of IQ in deaf people, as assessed by language-reduced measures, is almost identical to the distribution in normal-hearing people. This allows for the use of the PIQ as a means (though certainly not the sole means) to differentially diagnose the effects of deafness from the effects of cognitive delays (Braden, 1992; Mailer & Braden, 1993).

The use of the Wechsler Verbal Scales with deaf and hard-of-hearing children has received relatively limited study (Mailer & Braden, 1993). A meta-analysis of available studies reported a mean VIQ of approximately 1 SD below the mean of the standardization sample (Braden, 1992). Because VIQ is dependent on the acquisition of spoken English as well as on incidental knowledge gained through language, the Verbal Scale cannot necessarily be viewed as a valid measure of a deaf person's cognitive ability (Mailer & Braden, 1993). Sattler (1992) suggests that the Verbal and Performance Scales can be compared to evaluate the level of a child's verbal language deficit, but that the Verbal Scale should not be used in the computation of IQ.

Sullivan and Schulte (1992) argue that because most deaf children attend school with normal-hearing peers, and engage in tasks that require language-related abilities, the use of the Verbal Scale may enhance one's ability to predict academic performance. This is concordant with Maller and Braden's (1993) finding that VIQ correlated with achievement significantly higher than did PIQ and achievement using the WISC-III. Maller and Braden (1993) caution examiners to carefully present Verbal Scale results, because low Verbal scores may be misconstrued as evidence of low cognitive ability.

**DIFFERENTIAL ITEM FUNCTIONING**

The debate between those who condemn the use of Verbal subtests with deaf children, and those who argue its use may be valuable for more accurate prediction of success, has taken a new twist. Maller (1996) examined differential item functioning (DIF) with Verbal Scale items in deaf samples. That is, rather than simply asking whether deaf children were more or less likely to get the items right or wrong, or asking whether the score on items was better predictive of achievement outcomes than other (nonverbal) item scores, she asked the question, "Do the items function the same way in deaf and hearing groups when deaf and hearing children are equated for overall scaled score?" If this question is not answered...
in the affirmative, it provides evidence that Verbal subtests are biased measures of ability. That is, the Verbal subtests are inaccurate because deaf children and hearing children who have the same overall score are more or less likely to get specific items right or wrong. Evidence of DIF suggests that items sample verbal knowledge differently within deaf and normal-hearing children. This is a statistical definition of item bias, because DIF implies the item does not measure the same construct with equal accuracy in both groups.

Maller's research found that Verbal subtests of the WISC-III (e.g., Vocabulary, Information) had numerous items demonstrating DIF. That is, she provided evidence to show that the Verbal Scales of the Wechsler do not measure the same construct with equal accuracy in deaf and hearing samples. Therefore, rather than rejecting Verbal subtests on the basis of a philosophical argument, or accepting them on the basis of an empirical statement, psychologists might be wise to avoid the use of Verbal subtests with deaf children because they are less accurate for measuring intelligence. However, the degree to which this inaccuracy, or DIF, actually influences the viability of the Verbal Scale as a predictor of success is not well established. Although Verbal subtests may show evidence of item bias, they still predict academic achievement better than Performance Scale scores. Therefore, additional research on the WISC-III is needed to clarify the degree to which DIF may contribute to inappropriate conclusions from a clinical, not just a research, perspective.

DEAF CHILDREN OF DEAF PARENTS
VERSUS DEAF CHILDREN OF HEARING PARENTS

The meta-analysis reported by Braden (1994) indicated that deaf children who have two deaf parents (DP) have higher mean PIQs than deaf children of hearing parents. DP (who comprise about 4% of all deaf children) also display consistently higher scores on measures of academic achievement than deaf children of hearing parents. Several hypotheses have been offered to explain these findings. Most of these hypotheses invoke the linguistic and emotional support that DP enjoy relative to their other deaf peers. That is, DP are raised in homes where deafness is accepted, sign language is used naturally and consistently by all family members for all communication, and where their parents understand, accept, and participate in the deaf community. In contrast, deaf children with normal-hearing parents experience disadvantages on all of these dimensions (i.e., their parents are less likely to understand and accept deafness, consistently use sign language for communication, and understand and participate in the deaf community).

Although the relative superiority of DP to other deaf children on cognitive and academic tests is widely noted in the deafness literature, it is often overlooked that DP have higher mean PIQs than normal-hearing children (Braden, 1987, 1994). This finding challenges family-based explanations for higher DP scores, because normal-hearing children presumably enjoy all of the same advantages (i.e., acceptance, native language use, and participation in the broader community). This has led some to hypothesize that DP have higher PIQs because of genetic, not social,
factors (e.g., Braden, 1987, 1994; Kusche, Greenberg, & Garfield, 1983; cf. Conrad & Weiskrantz, 1981). Although there is evidence (Paquin, 1992) to support kinship correlations, and higher PIQs among deaf parents, scholars do not yet agree on a common explanation for the above-average PIQs of DP.

COMPARISONS BETWEEN DEAF CHILDREN IN DIFFERENT EDUCATIONAL PROGRAMS

Since the early part of this century, psychologists who work with deaf children have attempted to compare deaf children who attend programs in residential settings to their counterparts who attend commuter schools (i.e., programs in which children return home every night and return to school the following morning). Early comparisons between deaf children in these educational groups found that deaf children who attended residential schools generally had lower IQs. However, the interpretation of this finding has been a matter of dispute. Some (e.g., Raviv, Sharan, & Strauss, 1973) interpret this finding as evidence that residential school environments are less cognitively stimulating for deaf children than the more customary home and school environments provided in commuter settings. In contrast, others interpret the difference in IQs between residential and day programs as evidence of selection bias (that is, children with lower IQs are generally sent to residential schools, whereas children with higher IQs are more successful and can be retained in commuter schools). Research to address this question was conducted using the WISC-R. In two longitudinal studies (Braden & Paquin, 1985; Braden et al., 1993), it was found that deaf children who were in residential programs had lower IQs than deaf children in commuter programs. However, over 3- and 6-year periods, the IQs of deaf children in residential programs steadily rise, whereas those of deaf children in commuter programs are stable. Therefore, contrary to popular belief, the residential school appeared to have a positive impact on IQ. This study should be replicated in other samples using other instruments such as the WISC-III. As it stands, the study suggests long-term residential placement facilitates development of higher PIQ. The available data on the WISC-III suggest there are no differences in FSIQ, PIQ, nor VIQ between students attending residential versus mainstream settings (Sullivan & Montoya, in press).

COMPARISONS AMONG OTHER SUBGROUPS OF DEAF CHILDREN

Contrasts between younger versus older deaf children, moderate to severe versus profoundly deaf children, and children using signs versus those using speech showed no differences in WISC-III PIQ, VIQ, nor FSIQ (Sullivan & Montoya, in press). Contrasts between boys versus girls are less consistent. Whereas Slate and Fawcett (1996) reported substantial gender differences (based on a small, nonrepresentative sample), Sullivan and Montoya (in press) report no gender differences (based on a larger, more representative sample). We were not able to find any
study investigating ethnic group differences on the WISC-III within samples of deaf children, but we note that previous research with the WISC-R shows ethnic group differences quite similar to those reported in samples of normal-hearing children (Braden, 1994). Currently, there is little research investigating the similarities and differences among subgroups of deaf children on the WISC-III; clearly, psychologists need more research to draw firm conclusions regarding differences among groups of deaf children.

PROFILE ANALYSIS

The issue of whether and how to interpret Wechsler subtest profiles has received somewhat limited attention in the deafness literature. Whereas WISC-III profile analysis has been vigorously challenged in the literature regarding normal-hearing children (e.g., Glutting, McDermott, & Konold, 1997), it has received little attention in the literature regarding deaf children. We were unable to find any studies of profile interpretation using the WISC-III with deaf children, and we found no studies attempting profile analysis with any Wechsler version of the Verbal Scales.

Researchers have consistently noted that Wechsler PS profiles discriminate between deaf children who have relatively few academic difficulties, and deaf children with significant academic impairments (i.e., poor achievement in comparison to other deaf children). A review of this literature (Braden, 1990) shows that scores on the WISC-R Coding subtest are particularly sensitive to academic and neurological deficits in deaf children. That is, deaf children who have academic deficiencies, learning disabilities, and the like often exhibit low Coding scores relative to their other PS subtest scores. Although other interpretations of profile constellations have been suggested (e.g., Vonderhaar & Chambers, 1975, suggested depressed Picture Arrangement scores indicate reduced social awareness and judgment), the only characteristic of profile analysis that consistently emerges in deaf children is a link between depressed Coding subtest scores and psychoeducational difficulties (Braden, 1990). Also, there may be a sufficiently large number of such children that they depress the mean Coding score in large, representative samples of deaf children.

We note that the data in Figure 2 fail to show any clear or consistent profiles on the WISC-III PS. However, there appears to be consistency among studies in showing substantially lower Vocabulary scores relative to other Verbal Scale subtests. We cannot draw conclusions about the clinical sensitivity of WISC-III subtest profiles until researchers generate clinical studies that use WISC-III PS subtests to compare academically "successful" and "unsuccessful" deaf children.

WISC-III users should consider three additional factors with respect to profile analysis with deaf children. First, the use of special norms based on deaf children reduces or eliminates depressed Coding scores in many children, and thus may reduce the sensitivity of the instrument to psychoeducational difficulties (Braden, 1990). Second, none of the research on deaf children has used the large samples
and sophisticated analysis techniques that are used by profile analysis studies in studies of normal-hearing children (e.g., Glutting et al., 1997). Third, the available factor analyses of the WISC-III PS with deaf children are not consistent in supporting the separation of PS subtests into PO and PS factors. That is, the Coding and Symbol Search subtests may form a factor independent of other PS subtests, but they may not. We suspect that PSI (rather than Coding alone) may be sensitive to psychoeducational deficits in deaf children (as it is in hearing children), but we have no direct evidence to support our prediction at this time.

GUIDELINES FOR USING THE WISC-III WITH DEAF CHILDREN

Research on the WISC-III and other Wechsler scales suggests practical guidelines for scientists and practitioners who use the WISC-III with deaf children. We suggest guidelines for researchers (scientists) and practitioners in the final section of this chapter. Although we encourage the successful combination of scientist and practitioner perspectives, we realize that researchers and practitioners have different interests and orientations to the use of the WISC-III with deaf children.

RESEARCH GUIDELINES

Researchers should consider three distinct issues when using the WISC-III for research with deaf children. These issues are as follows:

- Assessment of test bias
- Selection of an achievement metric for validity studies
- Test translation and accommodation

TEST BIAS

Assessment of test bias must move beyond studies of differences between groups (e.g., deaf and normal-hearing children). Sadly, the literature describing the Wechsler with deaf children shows a marked lack of sophistication with respect to test bias. Most researchers simply compare means for deaf samples to normative means, and conclude there is bias when those means are different. This does not demonstrate test bias, because it invokes the “egalitarian fallacy” (see Jensen, 1980, Chap. 9; Reynolds, 1995). Instead, researchers must strive to examine genuine forms of test bias. These should include DIF studies (see Maller, 1996), direct comparisons of factor structures between deaf and hearing samples, and differential reliability and validity studies. Unfortunately, researchers have difficulty recruiting the large samples required for DIF and factor analytic comparisons. Failing such large-sample studies, small studies that report WISC-III PS reliabilities, stabilities, and correlations with other measures are useful in building
a body of evidence regarding the degree to which the WISC-III may function differently in deaf children (i.e., the degree to which the WISC-III exhibits bias when used with deaf populations).

**SELECTION OF ACHIEVEMENT METRIC**

One of the issues that has dogged previous versions of the Wechsler is the relationship between Wechsler scores and scores on achievement tests within samples of deaf children. Many studies report that these correlations are lower for deaf children than they are for hearing children. Three possible explanations for this phenomenon have been provided. First is the nonverbal nature of the Performance Scales (the most widely used part of the Wechsler to estimate deaf children's intelligence) overlaps less with the linguistic and verbal requirements of achievement tests. Therefore, the correlations between PIQ and academic achievement are low primarily because they sample distinct domains. A second argument that has been advanced is that the Wechsler is less effective at measuring the cognitive abilities that are needed for success in deaf individuals (Hirshoren, Hurley, & Kavale, 1977). A third argument is that the metric that is used for describing deaf children's achievement is inappropriate. That is, studies correlating IQ and achievement in hearing children typically convert IQ and achievement to age-based standard scores and then calculate correlations. In contrast, studies with deaf children typically calculate IQ but convert achievement scores into grade equivalents. The difference in metric is quite likely to lead to lower estimates of IQ–achievement correlations (Braden, Wollack, & Allan, 1995; cf. Kishor, 1995a, 1995b). This is frankly less an issue of the WISC-III and more an issue of how achievement metrics are selected to validate the WISC-III. However, readers of research using the WISC-III should carefully consider the metric that is used in the study when evaluating the outcomes. If the metric uses some form of age-adjusted score (where achievement measures are preferably normed on other deaf children), the correlation of that study is more likely to be comparable to studies of hearing children than when a grade equivalent or non-age-adjusted score is used to describe deaf children's achievement. Thus, it is likely that existing literature significantly underestimates the correlation between IQ and achievement that actually exists between Wechsler scales and tests of achievement because of inappropriate selection (see Braden et al., 1995, for more details).

**TEST TRANSLATION AND ACCOMMODATION**

Two factors may invalidate assessment results: construct underrepresentation, and construct-irrelevant variance (Messick, 1995). Construct underrepresentation means the test samples the domain of interest too narrowly; that is, the test excludes important samples of behavior to represent the domain of interest. Construct-irrelevant variance means the test includes task characteristics that are unrelated to the relevant domain. This irrelevant sampling increases or decreases the difficulty
of the test for a person or group. Researchers should consider the degree to which each of these factors may affect the use of the WISC-III with deaf children.

Construct underrepresentation is an issue when deciding whether to include Verbal Scale subtests with deaf children. One might argue that intelligence is broadly defined, and therefore one must include verbal abilities in any broad representation of intelligence. Conversely, construct-irrelevant variance argues against using Verbal Scale subtests with deaf children. Use of verbal items increases the difficulty of the task for deaf people by including a task characteristic (knowledge of English) that is unrelated to the domain of interest (i.e., depressed English language skills—not depressed intelligence—may lower deaf children's scores). We agree with other experts that there is more risk in introducing construct-irrelevant variance than there is risk for underrepresenting cognitive domains, but we note that the argument for exclusive use of nonverbal or performance tests with deaf children may excessively narrow the breadth of the domain sampled (i.e., PIQ, PO indexes, and other Performance Scale scores may not adequately reflect broad intellectual abilities).

Test administration modifications can also introduce construct-irrelevant test variance. Generally, the literature focuses on the degree to which test modifications create "a level playing field" for deaf children. For example, by administering the WISC-III in ASL, the deaf child has the opportunity to respond to items with the same degree of understanding that hearing children enjoy. However, it is also possible that test modifications may decrease task difficulty for deaf people. For example, the ASL sign for "conflagration" is the same as the sign for "fight." If a psychologist signs "fight" while asking the child to define "conflagration," that psychologist has made the item significantly easier for that deaf child than for a hearing child.

This discussion highlights the issues surrounding testing modifications. We are aware of only two studies that used appropriate translation to blind back-translation techniques to ensure that sign translations did not significantly alter the nature of Wechsler Scale content (Kostrubala, 1997; Maller, 1996). There are no formal studies to evaluate the degree to which other recommended test modifications (e.g., supplemental instruction, practice items) reduce construct-irrelevant variance. Rather, it is implicitly assumed that any accommodation that increases deaf children's scores is "good." This assumption may be warranted in many cases (e.g., oral administration of the WISC-III to a profoundly deaf ASL user would certainly introduce construct-irrelevant difficulty), but researchers must invoke appropriate objective techniques to ensure that modifications do not inappropriately introduce construct-irrelevant variance (e.g., make the items excessively easy or change the nature of the task).

**PRACTITIONER GUIDELINES**

Practitioners vary in their exposure to and experience with deaf children. Practitioners who administer the WISC-III should remember that there are many test
administration issues to consider beyond the communication method. Ideally, practitioners will gain expertise in methods of sign language communication and in understanding the unique, and varied, characteristics deaf children may display. Deaf children do not form a homogeneous group; etiologies vary, as does age of onset, severity of hearing loss, presence of additional disabilities, degree of exposure to sign language, family characteristics, and so on. Psychologists may pursue numerous avenues to acquire expertise in working with deaf children. Resources are listed annually in the April issue of the *American Annals of the Deaf*. The National Information Center on Deafness at Gallaudet University in Washington, D.C., is another resource (internet address: http://www.gallaudet.edu:80/~nicd/). For psychologists who do not have expertise in deafness, we offer the following recommendations:

- **Obtain the services of a qualified psychologist who has deafness expertise.** Psychologists may contact state residential schools for the deaf, state departments of education, or review resources in the April issue of the *American Annals of the Deaf* to identify local resources.

- **If an expert psychologist is not available, use a sign language interpreter.** We recommend psychologists use qualified, nationally certified interpreters (i.e., not parents, educational aides, or teachers). The psychologist should meet with the interpreter before the evaluation to review testing and administration procedures. Interpreters are bound by a code of ethics that prohibits revealing the content of sessions with clients; psychologists should remind the interpreter that the content of the intelligence test is included in this confidential mandate. If the psychologist plans to administer the Vocabulary subtest, we recommend that the first administration of verbal items use finger spelling and/or printed cards to present specific items. Later testing-of-limits could readminister failed items using sign translations. We discourage the use of signs for standard administration, because some translations may render the item easier, or harder, than it would be for normal-hearing children. We refer readers to examples of translation problems, and recommended solutions, in Table 9.4. We also recommend that psychologists schedule a posttesting session with the interpreter to review response translations, and to tap the interpreter’s cultural expertise in deafness (e.g., a psychologist may view a deaf child’s affect as excessively labile, whereas the interpreter may perceive it as appropriate within deaf culture).

- **Determine the child’s preferred mode of communication, and administer the WISC-III in that mode.** Observe the child in a natural setting, and consult with the child’s parent(s) and teacher(s) to establish communication preferences.

- **In most cases, interpret PIQ as the best indicator of the child’s underlying cognitive ability.** To the degree that the examiner believes the test administration yielded reliable and valid results, and the deaf child is free from additional disabilities (e.g., coordination problems) that might affect PIQ, the
### TABLE 9.4 Examples of Administration Modifications for Use with Deaf Children

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral directions assume client can listen while looking at materials. Deaf people cannot “listen” and “look” at the same time.</td>
<td>Simplify directions, refer to visual aspects of materials, and pause to allow client to look to materials, then look back.</td>
</tr>
<tr>
<td>The query “Show me where you mean” fails to capitalize on spatial parameters of ASL.</td>
<td>Sign “Where?” with a questioning expression and point towards the picture.</td>
</tr>
<tr>
<td>Because the sign for “dozen” in ASL is the same as “12,” signing “How many are in a dozen?” provides the answer to the client.</td>
<td>Fingerspell the word “dozen” when asking the question.</td>
</tr>
<tr>
<td>The signs for “tribe” and “family,” are variants of the same root (“group”), so the signs signify the similarity between the two words.</td>
<td>Fingerspell the words “tribe” and “family” when asking the question.</td>
</tr>
<tr>
<td>The ASL translation for “If you have 3 and I have 3, how many are there altogether?” literally shows the answer with the fingers.</td>
<td>Translate this item into signed English so that the numbers are presented in serial, not parallel, order.</td>
</tr>
<tr>
<td>The ASL translations of rare English words usually substitute a more common signs (e.g., “conflagration” would be signed “fight,” thus making the ASL version easier).</td>
<td>Fingerspell and/or write the words.</td>
</tr>
</tbody>
</table>

PIQ is the best index of general ability. PIQ is one means of differentiating deafness from mental retardation (i.e., a PIQ >70 generally rules out a diagnosis of mental retardation). However, placement decisions must not be made based solely on the use of PIQ. Other data, from other sources, settings, and respondents, must be included in any assessment.

- If desired, use VIQ to estimate the deaf child’s incidental acquisition of language and knowledge; do not use VIQ to estimate underlying cognitive ability.
- Do not assume special (“deaf”) norms are superior to norms based on normal-hearing children. Special norms may reduce the sensitivity of the WISC-III to differentiating psychoeducational difficulties, and may have few practical benefits. Consider carefully the use of special norms if they become available for the WISC-III.
- Consider administering a motor-free nonverbal test to supplement WISC-III Performance Scale results. Deaf children’s scores may be lower on motor-free tests as compared to the WISC-III Performance Scale, but motor-free nonverbal tests may tap reflective thinking better than performance tasks.
- Compare the deaf child’s performance across tasks that are speeded (e.g., Coding, Block Design) and those that are not (e.g., Picture Completion, motor-free nonverbal tests). Some experts argue that deaf children may be penalized on speeded tasks because they do not understand the time-limited nature of the test; also, motor coordination difficulties are more common.
among deaf children than among normal-hearing peers. Therefore, examiners should consider the possible effects of communication and psychomotor disabilities on deaf children's Performance Scale scores.

CASE EXAMPLE: WISC-III

We have included a fictionalized case study to illustrate and apply the themes in this chapter. This case presents the report a psychologist produced in the context of an adolescent female who is seeking admission to the state residential school for the deaf. The report specifically touches on issues of test administration, selection of norms, the use and interpretation of Verbal Scale subtests, the value of the WISC-III for differential diagnosis, and the integration of intelligence, achievement, and adaptive behavior data to form a coherent presentation. Subsequent to this case study, a multidisciplinary team recommended Teresa be placed at the residential school. A follow-up suggested a dramatic improvement in social and interpersonal relationships, but no significant gains (nor losses) in academic achievement.

**Name:** Teresa  
**Date of birth:** 2-6-82  
**Date of assessment:** 12-6-95  
**Age:** 13 years, 10 months  
**School:** Mainstream middle  
**Grade:** 7  
**Parents:** Tom & Toni  
**Ethnicity:** Caucasian

**REASON FOR REFERRAL**

Teresa was referred for this assessment as part of the intake process at the residential school for the deaf. She has expressed dissatisfaction with her current educational setting, and she and her parents are exploring the possibility of placement in the residential school.

**BACKGROUND INFORMATION**

Teresa lives with her parents and two younger brothers in a small urban area. She is the only deaf member of her family. Teresa was diagnosed with a profound, bilateral hearing loss at the age of 2 years, although the loss is believed to be congenital. The etiology of her deafness is unknown and she has no additional physical concerns.

Teresa communicates using PSL with her family and with a normal-hearing friend who lives in her neighborhood. At school she communicates using ASL with several other deaf students, but her interpreter and support teacher use Signed English.

Academically, Teresa receives language arts instruction in a self-contained classroom with five other children who are deaf. She attends all of her other classes with her peer with the assistance of a sign language interpreter. Her grades have been Bs and Cs thus far this year.
PREVIOUS ASSESSMENTS

Teresa’s most recent psychoeducational evaluation was completed nearly 3 years ago as part of the triennial review process. She obtained a score within the Average range on the Performance Scale of the WISC-III, and scores within the Well Below Average range in Reading and the Below Average range in Mathematics on the Brief Form of the K-TEA.

CURRENT ASSESSMENT TECHNIQUES

WISC-III
K-TEA
Vineland Adaptive Behavior Scales, Survey Form
Parent and Child Interview

OBSERVATIONS

Teresa was pleasant, although initially somewhat shy, throughout the assessment. She communicated with me predominantly via ASL signs in English word order. As the evaluation began, Teresa became more animated. Test items were presented in sign language and, when using signs would change the difficulty level of the task (i.e., Verbal Scale and Achievement tests), in written or finger-spelled form Teresa responded to task demands as if she understood directions without difficulty. She was persistent in her approach to most tasks, although when presented with more difficult verbal items she tended to respond quickly that she didn’t know the answer. Given Teresa’s motivation and her understanding of the tasks presented, I believe that the results obtained are reliable and valid indicators of her current level of functioning in the areas assessed.

ASSESSMENT RESULTS

WISC-III Scale and Index scores

<table>
<thead>
<tr>
<th>Scale</th>
<th>Score</th>
<th>Confidence Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Scale</td>
<td>88</td>
<td>83–94 at 90%</td>
</tr>
<tr>
<td>Performance Scale</td>
<td>102</td>
<td>95–109</td>
</tr>
</tbody>
</table>

Verbal Comprehension = 85 (80–92)
Perceptual Organization = 104 (97–111)

WISC-III Subtest Scores

<table>
<thead>
<tr>
<th>Verbal Scale</th>
<th>Performance Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>Picture Completion</td>
</tr>
<tr>
<td>Similarities</td>
<td>Coding</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>Picture Arrangement</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Block Design</td>
</tr>
<tr>
<td>Comprehension</td>
<td>Object Assembly</td>
</tr>
</tbody>
</table>
K-TEA Brief Form Subtest and Composite Scores

- Math = 87 (19th percentile)
- Reading = 76 (5th percentile)
- Spelling = 82 (12th percentile)
- Battery Composite = 78 (7th percentile)

Vineland Adaptive Behavior Scales, Survey Form
(Teresa’s mother provided responses)

- Communication = 85 (74–96 at 90% confidence level); Adequate–Moderately Low range
- Daily Living Skills = 106 (98–114); Adequate range
- Socialization = 85 (74–96); Adequate–Moderately Low range
- Adaptive Behavior Composite = 96 (89–103); Adequate range (the Maladaptive Behavior Domain was not assessed; norms are in reference to normal-hearing age mates, not special deaf norms).

IMPRESSIONS

Teresa’s score on the Verbal Scale of the WISC-III should be interpreted as an indicator of her incidental knowledge and language acquisition—not her underlying cognitive ability. Her score on this scale is within the Low Average to Average range, reflecting performance between the 13th and 34th percentiles when compared to normal-hearing children her own age. Subtest scores are generally within the average range, with the exception of the Vocabulary subtest, which is significantly below average and is an area of relative weakness compared to Teresa’s level of performance on this scale.

On the Performance Scale, Teresa’s scores are in the Average range, reflecting performance between the 37th and 73rd percentiles when compared to normal-hearing age mates. Her score on the Picture Completion subtest is above average, and may reflect an area of relative strength for Teresa; all other subtest scores are within the average range.

The 14-point difference between Teresa’s VIQs and PIQ is sufficiently large to suggest her nonverbal abilities are stronger than those requiring English. However, differences this size are not rare in normal-hearing children (28% of the WISC-III standardization sample showed a difference of the same size or larger), and higher PIQs over VIQs are common in deaf children. A similar pattern is shown by the 19-point discrepancy between her VC and PO indexes, which is met or surpassed by only 15% of the standardization sample. Again, such differences are common in deaf children, because deaf children often do not acquire English language fluency (and consequently get lower verbal scores).

Teresa’s scores on the K-TEA reflect performance in the Well Below Average to Below Average range in Reading and Spelling, and the Below Average to Average range in Mathematics. These scores are commensurate with her actual classroom performance relative to normal-hearing children. However, scores in this range are common for deaf children.
Teresa’s Adaptive Behavior Composite is within the Adequate range, between the 23rd and 58th percentiles compared to children her own age. Her Communication and Socialization domain scores are in the Adequate to Moderately Low range. The Written Communication subdomain is significantly lower than the other areas within the Communication domain, as is the Interpersonal Relationships subdomain within the Socialization area. One possible account of the depressed Interpersonal Relationships subdomain score is Teresa’s limited access to friends and peers who can communicate easily with her via signs. An informal interview with Teresa indicates that her dissatisfaction with her current educational setting is mainly due to social factors. She reports that she has few friends in school and that it is difficult for her to interact with her peers, the vast majority of which do not know sign language. She also reports that she is reluctant to participate in classroom discussions, because she must depend on an interpreter to express her thoughts. She feels she is becoming more and more isolated socially as the school year goes on.

SUMMARY AND RECOMMENDATIONS

Teresa is a 13-year 10-month-old girl who was referred for this evaluation as part of the intake process for admission into this residential school. Although she is generally performing at acceptable academic levels in her current school setting, she reports feeling socially isolated. Teresa’s cognitive abilities are within the average range as assessed with the WISC-III Performance Scale. Her WISC-III Verbal Scale score reflects an acquired language and knowledge level within the low average to average range. Teresa has difficulty in reading and spelling, and in her ability to define various vocabulary words. Adaptive behavior scores in communication, daily living skills, and socialization suggest adequate adaptation to her community environment. Items relating to reading and written communication, and to interpersonal relations, suggest limited interpersonal opportunities.

1. A multidisciplinary team, including Teresa, her parents, and representatives from Teresa’s current school placement and district and the residential school for the deaf, should convene to discuss most appropriate placement for Teresa. A critical focus of the team should be to evaluate the degree to which the two settings can meet Teresa’s academic and social development needs.

2. Regardless of setting, Teresa’s educational placement should provide instruction (and, to the degree possible, social opportunities) in Teresa’s primary mode of communication (i.e., PSE, with ASL and English sign features).

3. These assessment results suggest no other educational disabilities at this time. That is, there is no evidence of cognitive disabilities, nor is there evidence that the gap between Teresa’s achievement and her (nonverbal) intellectual ability is due to any condition other than deafness. However, Teresa’s lack of peers with whom to communicate could contribute to motivational and emotional difficulties in the future.
REFERENCES


9. ASSESSMENT OF DEAF CHILDREN


INTRODUCTION

The WISC-R (Wechsler, 1974) has been the most widely used individual intelligence test for children for a number of years. However, it was increasingly criticized as being out of date (Witt & Gresham, 1985). As a result, The Psychological Corporation revised and updated the test in 1991 (Wechsler, 1991).

Although little doubt exists that the Wechsler scales will continue to be the most frequently employed measures of intelligence, there remain misconceptions as to what constitutes the neurobiological bases of intelligence and how these measures of general ability should be interpreted with respect to revealing clinically meaningful variability in brain–behavior relationships in children and adolescents. It is our contention that psychometric ‘g’ cannot be localized in the brain but that associated subprocess may be differentially localized within the brain. Wechsler (1958) has well expressed this view himself. As he has suggested,
One cannot expect anything like fixed centers of intelligence for purely logical reasons. Intelligence deals not with mental representations but with relations that may exist between them, and the relations cannot be localized. . . . For effective functioning intelligence may depend more upon the intactness of some rather than other portions of the brain, but in no sense can it be said to be mediated by any single part of it. Intelligence has no locus. (p. 20)

Although this indeed seems like a reasonable perspective at face value, it does challenge us to develop a better understanding of the neurobiological bases of intelligence because intelligence is so very sensitive to the effects of brain impairment (Boll, 1974; Seinberg, Giordani, Berent, & Boll, 1983). Thus, one might ask, how can something not localized within the brain in turn be so sensitive to brain impairment?

The answer lies in the notion that intelligence, or psychometric g, is a reflection of our ability to perceive relations, or as Spearman and Jones (1950) suggest, the ability to engage in the "education of correlates." Thus, although the various subtests or scales on intelligence tests may be useful in revealing some aspects of impairment in correlated neurological systems that may be more localized in the brain, a general intelligence quotient such as that provided by a full scale measure will be less useful in specific clinical interpretation.

So, where does this leave researchers conducting neuropsychological evaluations with respect to employing tests of intelligence or their scales or subtests, such as the Wechsler scales? Why might they be useful and what interpretations about neuropsychological functioning can one make from data provided by tests of intelligence such as the WISC-III?

In order to address these questions and reach conclusions supported by the literature, we first need to address the validity of the factor structure of the Wechsler scales. Then we will address the validity of using verbal–performance discrepancies in arriving at neuropsychological interpretations. Although much of the research relevant to this discussion is derived from studies employing the WISC-R, some pertinent new studies also exist. As will be seen, a number of studies do support the notion than some neuropsychological interpretations can be made employing tests of intelligence, especially with more well-researched populations. Finally, we will attempt to draw some conclusions that are supported by the extant literature and provide some preliminary data that illustrate the notion that certain subprocesses assessed on the WISC-III may indeed have some association with theoretically consistent neurological systems. We remain convinced, however, that general intellectual ability, or psychometric g, has no specific locus in the brain, only perhaps its associated subcomponents.

**FACTOR STRUCTURE OF THE WISC-III**

Factor analytic research with the WISC-R yielded strong support for the Full Scale (FSIQ), Verbal (VIQ) and Performance IQ (PIQ) scores along with a three-factor solution comprised of a Verbal Comprehension (VC) factor, a Perceptual
Organization (PO) factor, and a Freedom from Distractibility (FD) factor (Kamphaus, 1993). Factor analysis of the WISC-III (Wechsler, 1991) as reported in the examiner's manual, continues to demonstrate support for the FSIQ, VIQ, and PIQ scores as well as the VC and PO factors. With the addition of the Symbol Search subtest, however, the factor analysis now yields two additional factors rather than just one. The third and fourth factors are identified as FD and Processing Speed (PS). The new FD factor includes only two subtests, both auditory; the fourth factor (PS) is also composed of two subtests, both visual and involving a motor speed component.

VERBAL-PERFORMANCE DISCREPANCY

Research has been done relating performance on the WISC (Wechsler, 1949) specific to localizing brain dysfunction. There is some correlative evidence to support that the VIQ reflects left-hemisphere functioning whereas the PIQ reflects right-hemisphere functioning (e.g., Fedio & Mirsky, 1969; Rourke, Young, & Flewelling, 1971). Similar types of conclusions, with implications for clinical diagnosis, have been offered for the VIQ PIQs on the WISC-R (Wechsler, 1974), with particular importance attributed to discrepancies between VIQ and PIQ scores (e.g., Kaufman, 1979). These hypotheses are based on the distinction between verbal and nonverbal abilities that is historically evidenced in factor analytic support (Kaufman, 1979) and continue to be applicable for the WISC-III (Kamphaus, 1993).

DISCREPANCY IN HETEROGENEOUS LD SAMPLES

It has been argued that a significant VIQ–PIQ discrepancy (PIQ > VIQ) is suggestive of an LD (e.g., Kavale, & Forness, 1984) and/or the presence of a speech or language impairment (Kamphaus, 1993). There are numerous studies that support the use of VIQ–PIQ discrepancy as a characteristic of LDs. For example, Newman, Wright, and Fields (1989) found that students with a reading disability demonstrated a significant VIQ–PIQ difference. Further, Clampit and Silver (1990) concluded that a large VIQ–PIQ difference had considerable face validity in identifying children with LDs based on the WISC-R standardization sample. Cheng, Dai, and Liu (1994) investigated the pattern of VIQ–PIQ discrepancy for Chinese children using the Chinese version of the WISC-R. Results suggested that children with LD showed a larger mean VIQ–PIQ discrepancy than that of controls. Cheng et al. concluded that the factor pattern for the children with learning disabilities was consistent with the pattern described by Bannatyne (1968).

In contrast, other researchers have not found the use of WISC-R VIQ–PIQ discrepancy patterns to be useful in the differentiation of children with LDs from other groups of children (e.g., Clampit & Silver, 1990; Humphries & Bone, 1993; Newman et al., 1989). In many studies, the VIQ–PIQ discrepancy found on the
WISC-R was small (Kamphaus, 1993), and similar results have been found with the WISC-III. In a study of 202 children with specific LDs (115 children with mental retardation and 159 children who were referred but not classified), Slate (1995) found that all three groups of children demonstrated a higher PIQ than VIQ on the WISC-III. The VIQ–PIQ discrepancy for the children in this study was smaller than that found in the normative sample. In a study of 40 children identified as language impaired, VIQ–PIQ discrepancies on the WISC-III were also found to be small (Phelps, Leguori, Nisewaner, & Parker, 1993).

**DISCREPANCY AND THE GIFTED**

Furthermore, a number of researchers have examined the profiles of children with FSIQs greater than 120 with results suggesting that large subtests scatter, VIQ–PIQ discrepancy (VIQ > PIQ), and high variability are relatively frequent for bright children (Saccuzzo, 1992; Taylor, Ziegler, & Partenio, 1984; Wilkinson, 1993). There is also some indication that the higher VIQ generally found with these students may be associated with higher achievement motivation (Kamphaus, 1993).

It has been noted that among bright students, gender differences with regard to profile variability were evident (Wilkinson, 1993). Cultural differences have also been found in the studies of gifted children. For example, Saccuzzo (1992) found that a significant VIQ–PIQ discrepancy (VIQ > PIQ) was common for gifted White children, but the opposite pattern (PIQ > VIQ) was evident for gifted children who were Filipino. Furthermore, Taylor et al. (1984) found that gifted children of Hispanic origin were likely to have a greater VIQ–PIQ (PIQ > VIQ) than White children. Other studies have also found that the frequency or extent of a VIQ–PIQ discrepancy was no higher in gifted children than in children with FSIQs in the average range (Patchett & Stansfield, 1992).

**DISCREPANCY AND EMOTIONAL OR BEHAVIORAL PROBLEMS**

Evidence to support a VIQ–PIQ discrepancy (PIQ > VIQ) on the WISC-R as predictive of delinquent behavior is also equivocal (Cornell & Wilson, 1992; Walsh, 1992; Walsh & Beyer, 1986). For example, Walsh (1992) found no correlation between the magnitude of the VIQ–PIQ discrepancy and measures of delinquency. Results suggested, however, that the presence of such a discrepancy may predict the severity of delinquent behavior once the youth has become a delinquent. Discrepancies between VIQ and PIQ (VIQ > PIQ) have also been attributed to depression; however, as with LDs and delinquency, the results of clinical studies have been mixed (Mokros, Poznanski, & Merrick, 1989).

**DISCREPANCY AND CULTURAL DIFFERENCES**

Various studies have demonstrated a higher frequency of VIQ–PIQ discrepancies in children for whom English is their second language (Gerken, 1978; Sac-
cuzzo, 1992; Taylor et al., 1984). Variations of English usage may also impact not only on the child's ability to understand the task and respond, but also the clinician's ability to interpret the response (Kamphaus, 1993). Other factors (e.g., lack of exposure) may also result in a VIQ–PIQ discrepancy. For example, it has been suggested that African-American children may have difficulty with some of the Performance Scale tasks due to the novelty of these tasks (Reynolds, 1981).

In summary, research has not supported the validity of reliance on VIQ–PIQ discrepancies for clinical purposes with the WISC-R. With regard to diagnosis or classification, it has been argued that reliance solely on the presence of a specific VIQ–PIQ discrepancy on the WISC-III is of no value (Kamphaus, 1993) and may result in misclassification (Prifitera & Dersh, 1993).

In interpreting a VIQ–PIQ difference, a variety of confounding factors should be addressed. These include the presence of speech or language problems, hearing problems, motor problems, achievement motivation, and linguistic and cultural differences as accounting for the VIQ–PIQ discrepancy (Kamphaus, 1993). Thus, although a VIQ–PIQ discrepancy may be used for hypothesis generation, the presence or absence of such a discrepancy should not be viewed as conclusive evidence for the presence or absence of a disability.

**THE THIRD AND FOURTH FACTORS OF THE WISC-III**

In the early studies of the WISC-R, the FD factor has been used as a clinical indicator of attention deficit/hyperactivity disorder (ADHD) (Barkley, 1990); however, a variety of studies have questioned its validity for this purpose (e.g., Cohen, Becker, & Campbell, 1990; Kamphaus, 1993; Kostura, 1993; Stone, 1992). Various studies have, for example, failed to support a correlation between the FD factor and attention of hyperactivity subscales of frequently used rating scales for ADHD (e.g., Cohen, Becker, et al., 1990). Cohen and colleagues found that the FD factor correlated in the expected direction only with the Anxiety subscale on the Conners Teacher Rating Scale (Conners, 1969, 1973; Goyette, Conners, & Ulrich, 1978). Cohen et al. concluded that the FD factor was not a reliable indicator of ADHD. They further surmised that the high correlation between the FD factor and the VC factor ($r = .70, p < .001$), and FD factor and the PO factor ($r = .60, p < .001$) suggested that the FD factor was actually measuring some other cognitive function such as numerical facility, working memory, or sequencing ability.

Similarly, Kostura (1993) concluded that the WISC-R FD factor was not a pure indicator of attention and should not be used when making ADHD diagnoses. He cited several studies (e.g., Barkely, DuPaul, & McMurray, 1990; Brown & Wynne, 1982; Milich & Loney, 1979) that concluded that factor scores on the WISC-R do not adequately differentiate children with ADHD from other groups. Stone (1992) also questioned the utility of the FD factor, suggesting that it was never a specific entity, but the result of psychometric underfactoring. Furthermore, Woodcock (1990) found a lack of any combination of loadings to support
the WISC-R FD factor. He suggested that the three subtests comprising the FD factor are single markers of separate and specific abilities (short-term memory, processing speed, and quantitative ability) as opposed to a common factor underlying the subtests.

The validity of the FD factor continues to be an issue with the WISC-III. The WISC-III manual includes data suggesting that both learning-disabled (LD) and ADHD children earn lower scores on both the FD and PS factors. However, similar to previous studies of the WISC-R FD factor with regard to ADHD diagnoses, Anastopoulos, Spisto, and Maher (1994) questioned this factor's diagnostic utility. Although they found that the FD factor on the WISC-III was a more valid indicator of the construct of inattention than the FD factor on the WISC-R, its use as a diagnostic measure for an ADHD diagnosis was also rejected. They found that as many as 78% of children who met stringent criteria for an ADHD diagnosis did not demonstrate any significant discrepancy between the FD factor score and other factor scores.

Kamphaus, Benson, Hutchinson, and Platt (1994) concluded that the four-factor structure was "somewhat supported" but that the "theoretical or clinical importance of the additional factors and the Index scores is unclear" (p. 185). Furthermore, they concluded that the distractibility label for the third factor was not supportable and that the PS factor lacked empirical support as well. More recently, the factor structure of the WISC-III was examined for a sample of 140 children with LDs, (e.g., reading disorders), ADHD, or both (Smith & Gfeller, 1995). Both confirmatory and exploratory factor analyses were employed and both a three-factor model (VC, PO, and PS) as well as the four-factor model provided in the WISC-III manual emerged. Furthermore, Woodcock (1990) noted that for effective measurement of a specific ability (i.e., processing speed) a battery requires at least three subtests.

In a recent study of 126 children referred due to academic and/or behavioral difficulty (mean age of 9.24 years; \( SD = 2.33 \)), a battery of neuropsychological tests was administered in order to further explore the meaningfulness and utility of the FD and PS factors (Riccio, Cohen, Hall, & Ross, in press). The FD and PS factors were found to correlate significantly with the VC and PO factors, as well as with each other. Neither the FD nor PS factors correlated significantly with visual and auditory continuous performance measures, and correlations of the FD and PS factors with the Wisconsin Card Sorting Test (Heaton, 1981) were low at best. In addition, correlations with behavioral rating scales (parent and teacher) did not consistently support an association of the FD or PS factors with inattention or hyperactivity. Rather, results suggest that the FD or PS factor scores may be influenced by a myriad of abilities, including but not limited to attention and concentration.

The study by Riccio and colleagues did, however, yield a significant association between the FD factor and measures of immediate memory. This was evident not only for performance on the Sentence Imitation subtest of the Detroit Tests of Learning Aptitude-2 (DTLA-2; Hammill, 1985), but also for the Spatial Memory and Hand Movements subtests of the Kaufman Assessment Battery for Children.
(K-ABC, Kaufman & Kaufman, 1983). Specific to the PS factor, significant correlations \((p < .001)\) were found with the Spatial Memory and Hand Movements subtests only, and were not as robust as the correlations between the immediate memory measures and the FD factor.

Finally, to assess the utility of the third and fourth factors in determining the presence or absence of ADHD in a clinical population, analysis of variance (ANOVA) was computed for the three clinical groups that did not have a co-occurring diagnosis: LD/no ADHD \((n = 30)\), ADHD Predominantly Inattentive \((n = 31)\), and ADHD Combined Type \((n = 47)\). The finding of no significant between-group differences suggests that the FD and PS factors do not provide information that facilitates differential diagnosis within a heterogeneous clinical sample of children with developmental disabilities (Riccio et al., in press).

In another study, Slate (1995) examined the performance of children diagnosed with specific learning disability, mental retardation, and children who were referred but not diagnosed. Results indicated that children with specific learning disability scored higher on the FD factor than the PO factor, whereas children with mental retardation and undiagnosed children demonstrated a higher PO factor score relative to the FD factor score. Between-group differences were not evident on the FD factor or VC factor.

Kamphaus (1993) asserted that further research is required to provide clearer guidelines for the interpretation and use of the third and fourth factors. Kaufman (1993) also noted the lack of empirical support for the third and fourth factors and recommended that the clinical value of the four-factor model be studied further. Based on research completed thus far, the clinical validity of the FD and PS factors remains questionable. Although these two factors may emerge in factor analysis, the results of studies to date do not support any clinical or theoretical importance of the FD and PS factors for differential diagnosis of ADHD. Neither of these factors appears to measure attention or concentration in a manner similar to other measures used to assess attention (behavior rating scales; continuous performance tests).

Of most interest from the early research with the WISC-III is the association between the FD factor and measures of both auditory–verbal and visual–spatial immediate–working memory. Additional research specific to the relationship between the FD factor and immediate–working memory appears warranted. The PS factor also correlated significantly with measures of visual–spatial immediate–working memory, however, to a lesser degree. The PS factor may be better explained by Kaufman's (1993) assertion that there is greater emphasis on speed with the WISC-III and/or the incorporation of a motor component. Thus, this factor may be assessing immediate–working memory and fine motor skills in combination with PS. Given the apparent relationship between FD–PO factors and immediate–working memory, it would appear reasonable for clinicians to evaluate as child's ability to learn and remember new material (e.g., California Verbal Learning Test for Children; Delis, Kramer, Kaplan, & Ober, 1994; The Children's Memory Scale; Cohen, 1997) when the child exhibits poor performance on either of these factors.
It is beyond the scope of this chapter to discuss the various batteries and eclectic approaches that are currently being used in the neuropsychological assessment of children and adolescents. The reader is referred to chapters by Hynd and Willis (1988), Teeter (1986), and Tramontana and Hooper (1987) for such discussion. Regardless of which approach to assessment is applied, the goal of the evaluation is to analyze the integrity of the child's higher cortical functioning in order to (a) determine the child's present pattern of strengths and weaknesses, (b) characterize future patterns of functioning that might be expected as the child matures, and (c) develop appropriate remedial and compensatory programs in order to assist the child so that he or she can perform at their maximum capability (Cohen, Branch, Willis, Weyandt, & Hynd, 1992).

According to Luria's (1980) functional system approach to assessment, a given behavior is the product of several interrelated cortical and subcortical components functioning in concert. Therefore, if any one or more of the components become dysfunctional, the functional system is effected in such a way that one sees not only a quantitative but also a qualitative deviation from normal performance. The dysfunction also affects other functional systems that require the particular component in question. Those functional systems that do not require the particular component remain unaffected. Thus, a child neuropsychologist must not only be proficient in test administration and interpretation from a quantitative (norm-referenced) perspective, but also skilled at qualitative analysis (observation) of how a child or adolescent passes or fails certain test items.

Perhaps the best place to begin the neuropsychological assessment is with intelligence testing. The intelligence test affords a reference point for all higher cortical functions in that it provides an estimate of a child's intellectual potential or psychometric g. Furthermore, intelligence testing provides the child neuropsychologist with a cross-sectional panorama of the child's higher cortical functioning. Thus, by combining the "seven-step" quantitative approach to WISC-III interpretation as described by Kaufman (1994), and the qualitative approach offered by Kaplan (Appendix, Chapter 1, this volume), the examiner can use the WISC-III to begin developing hypotheses concerning the child's higher cortical strengths and weaknesses, which will require further verification. The remaining portion of the chapter will focus on the clinical research findings related to the WISC-R and WISC-III in various clinical groups commonly referred for child neuropsychological assessment.

**EPILEPSY**

It is estimated that 100,000 new cases of epilepsy are diagnosed in the United States each year (Kurtzke & Kurland, 1984), and of these cases, approximately 70% experienced their first seizure during childhood (Dreisbach, Ballard, Russo, &
Schain, 1982). Indeed, within the pediatric population, epilepsy is considered by many to be the most prevalent chronic neurologic disorder (Bolter, 1986) and is estimated to occur in 4 to 8 children per 1,000 (Blom, Heijbel, & Bergfors, 1978).

Specifically, epilepsy is a disorder that is characterized by two or more unprovoked seizures. Often, the words, "seizures" and "epilepsy" are used interchangeably, however they are two distinct entities. Epilepsy is a disorder that is the result of two or more unprovoked epileptic seizures, whereas a seizure is a sudden electrical discharge of neurons from the brain resulting in impaired functioning ( Henrikson, 1990). Impaired functioning within the areas of behavior, emotion, cognition, and academic performance have long been recognized as a result of epilepsy, and some authors have postulated that subclinical seizure activity may result in specific types of LDs (Hynd & Willis, 1988). In addition, a relationship between cognitive impairment and the effects of antiepileptic medications, and the epileptic focal point has also been recognized (Chaudry, Najam, DeMahieu, Raza, & Ahmed, 1992). Therefore, given the immense annual growth rate of this population, in conjunction with the resulting impairments that occur, the need to address the intellectual functioning of school children that fall within this population is vital.

Several investigators have examined the relationship between intelligence and epilepsy in children. In general, four major findings have repeatedly surfaced. First, the intelligence of children with epilepsy is skewed toward the lower end of normal functioning (Bolter, 1986; Farwell, Dodrill, & Batzell, 1985; Giordiani, Berent, Sackellares, et al., 1985; O'Leary et al., 1983; O'Leary, Seidenberg, Berent, & Boll, 1981). Specifically, 40-48% of the children evaluated in these studies had FSIQ scores below 90. In a study of epileptic children using the WISC-R, Bolter (1986) found that as a group the children fell within the average range. However, only 13% had above-average FSIQ scores, whereas 48% were below average.

Cohen, Branch, Riccio, and Hall (1991) studied the intellectual performance of children followed in a tertiary care medical center for epilepsy. Analysis of their performance on the WISC-R and the WISC-III indicated that this population generally functioned in the borderline to low average range across the VIQ, PIQ, and the FSIQ, as well as on the Factor/Index scores. In addition, like the standardization sample, children with epilepsy scored slightly lower on the WISC-III than the WISC-R. However, this decline was found to be significantly ($p < .05$) greater for VIQ and FSIQ in the sample of children with epilepsy as compared with that reported for the standardization sample. Furthermore, it is felt that the slightly lower mean scores obtained for this particular sample of children with epilepsy may be explained by the fact that this sample was taken from a population of children being treated at a tertiary care epilepsy center. In general, children treated at such a center tend to have more severe and more difficult-to-control seizure disorders that may result in lower intellectual functioning (Holmes, 1987). This interpretation is further supported by studies evaluating the intellectual performance of institutionalized children or children with severe epilepsy that have reported IQs in
the borderline to mentally handicapped range (Kugelmass, Poull, & Rudnich, 1938; Tenny, 1955).

The second major finding reported by several authors (e.g., Bourgeois et al., 1983; Farwell et al., 1985; O’Leary et al., 1983) but not all (Ellenberg, Kirtz, & Nelson, 1986) indicates that the onset of seizures in early childhood is associated with a higher risk for intellectual impairment than when seizures began in late childhood or adolescence. Third, children with symptomatic epilepsy have a higher risk for intellectual impairment than children with idiopathic epilepsy (Bourgeois et al., 1983; Ellenberg et al., 1986). Fourth, children with typical absence, generalized tonic clonic, and partial seizures tend to score higher on IQ tests than children with atypical absence, akinetic, and myoclonic seizures (Farwell et al., 1985; Giordiani et al., 1985; Sofijanov, 1982).

Finally, there appears to be few antiepileptic medications that have positive effects on cognitive functioning. In a review by Holmes (1987), the author concludes that most of the antiepileptic medications used to treat seizure disorders impede intellectual functioning, especially when there drugs are used in combination. Perhaps the worst offenders are phenobarbital (barbiturates) and dilantin (phenytoin), which have been shown to consistently produce negative effects on attention and concentration, motor speed, and processing speed. Thus, children with epilepsy should be carefully monitored for drug toxicity and associated functional deficits. Furthermore, the use of antiepileptic medication in combination should be avoided whenever possible.

**HYDROCEPHALUS**

Within the normal brain, there is a stable balance between the production and absorption of cerebral spinal fluid (CSF) maintaining suitable size of the ventricles. When a disruption of this balance occurs, hydrocephalus is often times the result. Hydrocephalus is a condition by which an increase of cerebral spinal fluid occurs in the ventricles of the brain resulting in an increase of intracranial pressure (Hynd & Willis, 1988). This increase of CSF is the result of obstruction of the CSF flow, that is, the overproduction or malabsorption of CSF (Bigler, 1988). The increased intraventricular pressure produces a compression, especially concerning the white matter, which in turn results in a thinning of the cortical mantle (Rowland, 1981).

Hydrocephalus often develops in conjunction with other neurological disorders or postnatal insult, thereby making it difficult to provide an accurate estimate of the incidence of true hydrocephalus. However, as a congenital disorder, it is estimated that the incidence rate ranges from .9 to 1.5 per 1,000 births (Milhorat, 1982). The incidence rate for hydrocephalus in association with spina bifida and myelomeningocele rises to 1.3 to 2.9 per 1,000 births (Myrianthopoulos & Kurland, 1961).

Hydrocephalus appears not only to have physical ramifications but cognitive ramifications as well (Bigler, 1988). Cognitive characteristics of hydrocephalic patients are low intelligence and poor academic performance (spelling, reading,
and arithmetic). Even in patients with mild hydrocephalus, IQ and academic achievement is reported to be approximately one standard deviation below the mean for the normal population. In addition, poor abstract reasoning, poor visual-perceptual skills, and a higher verbal than nonverbal IQ were also so noted (Hurley, Dorman, Laatsch, Bell, & D'Avignon, 1990).

Several studies (e.g., Anderson, 1975; Lorber, 1971; Tew & Laurence, 1975) have noted that on intellectual testing, verbal performance is typically higher than nonverbal performance. The verbal performance of hydrocephalic patients is one that is characterized by good form and structure producing adequate verbal fluency, but poor substance and monitoring of output, that is, a poor understanding of language. The speech of hydrocephalic patients has been described as lacking substance (Spain, 1974), with their conversation containing a high proportion of improper words (Swisher & Pinsker, 1971). Indeed, this verbal characteristic is one of several characteristics that in combination have been described as the "cocktail party syndrome." This syndrome is characterized by good fluency and articulation, verbal perseveration and trivial loquaciousness, overuse of social phrases, and an apparent overfamiliarity with the listener. Thus, the elevated VIQ typically seen in hydrocephalic patients may reflect these verbal abilities, which are emphasized on the verbal portion of IQ testing (Holler, Fennell, Crosson, Boggs, & Mickle, 1995).

Based on the type of nonverbal deficits exhibited by this population, some researchers have suggested right-hemisphere white matter lesions (Harnadek & Rourke, 1994; Hurley et al., 1990; Rourke, 1989). Another area that is hypothesized to be effected based on the observed cognitive patterns is the frontal lobes (Hagberg & Sjorgen, 1966).

**TRAUMATIC BRAIN INJURY**

In 1985, the National Head Injury Foundation estimated that 1 in 80 people born in that year would receive a brain injury from an auto accident. Indeed, traumatic brain injury (TBI) is so prevalent that it has become the second leading cause of death within the pediatric population (Annegers, 1983). Automobile accidents account for \( \frac{1}{2} \) to \( \frac{3}{2} \) of all TBIs within the adolescent population (Filley, Cranberg, Alexander, & Hart, 1987), whereas pedestrian auto accident accounts for the majority of TBIs in younger children (Chadwick, Rutter, Thompson, & Shaffer, 1981). Falls are the second leading cause of TBI (Field, 1979). While TBIs are not uncommon in infants from birth to 2 years of age (Goldstein & Levin, 1987), child abuse is the major cause of TBI in children under 1 year of age (Lange-Cosack & Tepfer, 1973). Finally, although there is little disagreement that males experience a higher frequency of TBI as opposed to females, ratio estimations vary from 2:1 to 3:1 (Goldstein & Levin, 1987; Rimel & Jane, 1983).

In nearly all children with TBI, residual motor deficits have been observed (Dalby & Obrzutz, 1991). These include paresis (Filley et al., 1987), unsteadiness of gate, mild tremors, fine motor incoordination, difficulties with copying tasks, difficulty with rapidly alternating movements, and clonus (Dimitrijevic, Dimitrijevic,
Persistent speech and language impairments do occur, but are most influenced by the etiology and extent of the lesion (Chadwick, 1985). Acquired aphasia in children is most often the result of brain trauma (Dalby & Obrzut, 1991), involving the left hemisphere. However, right-hemisphere lesions have been reported to produce linguistic deficits in 3½- to 6-year-olds (Hecaen, 1976). Expressive language difficulties include anomia for objects, decreased performance in verbal fluency, describing functions of objects, sentence repetition, and verbal associative fluency (Dalby & Obrzut, 1991). Receptive language deficits include poor linguistic comprehension (Levin & Eisenberg, 1979) and difficulties comprehending lengthy or multipart oral directions (Fuld & Fisher, 1977). Written expression deficits appear more evident in children than adolescents (Hecaen, Perenin, & Jennerod, 1984). Such writing difficulties occurred more frequently in right-handed children, 10 years of age and younger with left-hemisphere lesions.

Case studies by Fuld and Fisher (1977) have revealed that in general, intelligence scores are lower when compared to the preinjury estimates. However, some controversy does exist on WISC VIQ–PIQ discrepancies. Some studies have indicated a lower PIQ in relation to the VIQ (Chadwick, Rutter, Thompson, & Shaffer, 1981; Levin, Benton, & Grossman, 1982). Nevertheless, no significant VIQ–PIQ differences have also been reported (Fuld & Fisher, 1977). It has been suggested that lower PIQs may actually be the result of deficits in visual-spatial perception or reduced processing speed (Dalby & Obrzut, 1991). Furthermore, because attention span, memory, and motor skill or speed are frequently effected, any one or combination of these deficits could result in decreased performance on the PIQ subtests.

A greater number of children in comparison to adolescents are more effected in their cognitive functioning. In one study, (Levin, Eisenberg, Wigg, & Kobayashi, 1982) after a 6-month follow-up, it was found that 33% of the children (12 years and younger) in the sample had WISC-R VIQs lower than 80 compared to their matched adolescents who were all above 80. Similarly, 40% of children demonstrated WISC-R PIQ deficits, whereas no adolescent scored below 80.

There is still debate as to whether the preinjury condition of children with TBI is ever fully restored (Dalby & Obrzut, 1991). The foundation of this debate is based on the lack of accuracy in establishing preinjury estimation of intellectual functioning. A review by Ewing-Cobbs, Fletcher, and Levin (1986) suggests that children with severe TBI do not achieve preinjury intellectual functioning. Thus, data from previous studies indicate that premorbid intellectual functioning does not occur in children with TBI resulting in bilateral generalized damage.

**UNILATERAL LESION**

Neuropsychologists should be careful when inferring lateralized brain dysfunction based upon VIQ–PIQ discrepancies on the Wechsler scales. Based on the
oversimplistic notion that the left hemisphere primarily mediates linguistic functions whereas the right hemisphere is typically involved in the mediation of visual-spatial or nonverbal functions, it is tempting to conclude that individuals with unilateral lesions involving the left cerebral hemisphere should demonstrate VIQ–PIQ discrepancies in favor of PIQ. Similarly, individuals with unilateral lesions involving the right cerebral hemisphere should exhibit VIQ–PIQ discrepancies in favor of VIQ. However, this notion finds little support in the adult literature. For example, Kaufman (1990) reviewed 33 studies (30 WAIS and 3 WAIS-R) reporting VIQ–PIQ discrepancies in samples of adults with unilateral brain insult due to various etiologies. If one insisted upon at least a 6-point difference between mean VIQ and PIQ in the predicted direction, only 9 of the 33 samples (27%) exhibited the predicted VIQ–PIQ discrepancy pattern.

Research within the pediatric population is even more equivocal than that which is reported in the adult literature. In a study by Lewandowski and DeRienzo (1985), the authors investigated the WISC-R performances of children with hemiplegia who were 6 to 12 years of age. They reported a significant association between right-hemisphere insult and depressed PIQ, contrasted by a nonsignificant relationship between left-hemisphere insult and depressed VIQ. Shapiro and Doltan (1986) reported a nonsignificant relationship between location of lesion and VIQ–PIQ discrepancy on the WISC-R in their sample of children with unilateral brain damage. Morris and Bigler (1987) investigated the WISC-R performance of children (6 to 12 years of age) with unilateral brain damage and found a more substantial relationship between left-hemisphere insult and depressed VIQ than between right-hemisphere insult and depressed PIQ. Finally, Cohen, Branch, McKie, and Adams (1994) reported the neuropsychological test performance of a small group of children (6 to 16 years of age) with sickle cell anemia who experienced a single stroke in the left or right cerebral hemisphere. Results indicated that children with left-hemisphere insult demonstrated a global decline on intelligence testing with a very small VIQ–PIQ discrepancy in favor of PIQ (4 points). In contrast, children with right-hemisphere insult demonstrated a marked decline in PIQ only, which resulted in a large (13 points) VIQ–PIQ discrepancy in favor of VIQ.

In order to comprehend the disparity of the results concerning VIQ–PIQ discrepancy in children, one must consider the notion of plasticity in the developing brain. Although the exact physiological mechanisms underlying the recovery of functioning remain unclear, numerous physiological mechanisms, including axonal regeneration, collateral sprouting, and denervation supersensitivity have been postulated to account for the recovery processes. Furthermore, factors such as premorbid level of functioning, age at lesion onset, the type, size, and location of lesion, and subsequent habilitation efforts must also be considered (Cohen, Branch, Willis, Weyandt, & Hynd, 1992).

Specifically, there are two general theoretical perspectives regarding recovery of function following early brain insult. Isaacson (1976) proposes that early brain insult must be considered more detrimental than insults occurring later in life. This speculation is based on the fact that the infant brain often shows a reduction in size following insult, whereas the adolescent brain typically does not show a
similar reduction. Consequently, loss of function in the adolescent tends to be
more highly localized. This contention is supported by a study conducted by Rei-
tan (1981), which examined the learning potential of children experiencing brain
insult at different ages. Results indicated that younger children demonstrated a re-
duction in overall learning potential as compared with children who experienced
brain insult at an older age.

Based upon the work of Kennard (1936, 1942), who compared the effects of uni-
lateral motor cortex lesions in infant and adult monkeys, the concepts of "sparing of
function" and "brain plasticity" were developed. Specifically, Kennard found that
lesions in the infant monkey brain resulted in milder deficits than those occurring
within the adult monkey brain. These concepts encompass two major postulates
(Fletcher & Satz, 1983): (a) younger organisms exhibit better outcomes following
brain injury, and (b) there is greater plasticity in the immature brain, which explains
the better outcomes. For example, studies by Rasmussen and Milner (1977), and
Woods and Teuber (1973) indicate that early damage to the language centers of the
left hemisphere lead to a functional shift in language abilities to the right hemisphere.
However, this transfer is typically not complete in that subtle deficits in linguistic
ability are often evident later in life (Dennis & Whitaker, 1976, 1977). Furthermore,
children experiencing early left-hemisphere insult typically exhibit evidence of
"crowding" in which deficits in visual-spatial functioning are also evident (Lansdell,
1962, 1969). Thus, it is common to see higher VIQs, as opposed to PIQs, in children
who have experienced early (prior to 5 years of age) left-hemisphere insults (Cohen,
Hynd, & Hartlage, 1983). In contrast, children who experience early right-hemisphere
insults tend to demonstrate impairment of both VIQ and PIQ (Nass, Koch, Janowsky,
& Stiles-Davis, 1985; Woods, 1980). Woods interpreted this finding as being consis-
tent with an early right-hemisphere contribution to linguistic functions.

The concepts of "sparing of function" and "brain plasticity" can be used to
help explain the large disparity of results regarding VIQ-PIQ discrepancies in
children with unilateral brain damage. In addition, they also serve to emphasize
the fact that statements regarding brain dysfunction based solely upon WISC-III
VIQ-PIQ discrepancies are not warranted (Hynd & Willis, 1988).

DEVELOPMENTAL DYSLEXIA

Historically, developmental dyslexia (severe LD in reading) has been concep-
tualized as a homogeneous clinical entity in which researchers directed their ef-
forts in search of a single cause for the disorder. However, during the last 30 years,
considerable research has been accumulated that indicates that developmental
dyslexia is best conceptualized as a heterogeneous disorder with multiple etiolo-
gies. This conceptualization makes sense, if one assumes that the development of
reading skills requires the complex integration of various higher order processes,
which work together to form a higher cortical functional system for reading. Dys-
function in any one of these component processes will disrupt the total mecha-
nism and result in a qualitatively different subtype of developmental dyslexia
(Hynd & Cohen, 1983). Thus, it is possible for children with developmental dyslexia to experience reading failure for a variety of reasons including deficits in the areas of phonological processing, expressive/receptive language, auditory/verbal memory, visual/spatial perception, and visual/nonverbal memory.

Review of the dyslexic subtyping literature indicates that clinically reproducible subtypes do exist. Although there are several subtyping approaches in this area of literature, one approach is to use WISC VIQ–PIQ discrepancies as a clinical criteria for subtyping inclusion. For example, Kinsbourne and Warrington (1963) developed a subtyping model based upon a child’s performance on the WISC or the WAIS. Employing this rationale, they found two distinct subgroups of reading-disabled children. Group 1 was composed of children with at least a 20-point discrepancy between verbal and nonverbal performance (VIQ < PIQ). In addition, this group also exhibited delays in speech acquisition, verbal comprehension, and verbal expression. Group 2 also demonstrated a 20-point discrepancy but in the opposite direction (VIQ > PIQ). Furthermore, this group exhibited significant deficits in right–left confusion, arithmetic, and constructional abilities. Group 2 was labeled the “developmental” Gerstmann Group as reflected by similar symptoms seen in adults with Gerstmann Syndrome.

Using a neuropsychological model that involved the administration of a comprehensive neuropsychological battery that included the WISC-R, Pirozzolo (1979, 1981) was also able to identify two clinically distinct dyslexic subtypes. Subtype 1 (auditory-linguistic) exhibited the following pattern of test performance; average to above average PIQ, VIQ < PIQ, developmentally delayed onset of language, expressive speech deficit, anomia, agrammatism, phonological errors in reading, phoneme-to-grapheme association errors in spelling, a letter-by-letter decoding strategy, normal eye movements, and intact visual–spatial abilities. In contrast, Subtype 2 (visual–spatial) exhibited the following test performance pattern; average to above average VIQ, PIQ < VIQ, right–left confusion, early evidence of mirror or inverted writing, spelling errors reflecting letter/word reversals and omissions, finger agnosia, spatial dysgraphia (e.g., poor handwriting, poor use of space), visual type errors in reading, a phonetic decoding strategy, faulty eye movements during reading, and intact oral language abilities.

Finally, support for the use of VIQ–PIQ discrepancies in the clinical subtyping of children with dyslexia is derived from the work of Cohen and colleagues (Cohen, Hynd, & Hugdahl, 1992; Cohen, Krawiecki, & DuRant, 1987). In the first study, the dichotic listening performance of subtypes of children with developmental dyslexia was examined. In the second study, the effectiveness of the neuropsychological approach to remediation was examined. In both studies, children with dyslexia were diagnostically classified into three subtypes; language disorder, visual–spatial, and mixed. These children were then placed into two groups. Examination of the VIQ–PIQ discrepancy patterns of the three groups indicated that in both studies the language-disorder group demonstrated a significant VIQ–PIQ discrepancy in favor of PIQ, whereas the visual–spatial group demonstrated a significant VIQ–PIQ discrepancy in favor of VIQ.
Although there is still much work to be done in the area of diagnosis and treatment of children with developmental dyslexia, that is, subtype validation to ensure that the subtypes are clinically distinct and reproducible, this area of research has resulted in the educational and medical communities viewing dyslexia as a heterogeneous clinical entity with multiple etiologies.

**CONCLUSIONS**

What can be concluded about the usefulness of the WISC-III in terms of complementing information gathered from the comprehensive neuropsychological evaluation of children and adolescents? Based on our review, a number of conclusions seems warranted.

First, considerable scientific evidence points to the validity of the WISC-III Verbal and Performance scales. Furthermore, these scales may indeed differentially tap left- and right-hemispheric process, respectively. Consequently, in a child with a left-hemispheric lesion one might expect depressed VIQ relative to relatively intact PIQ. However, as we have noted, there are considerable factors that mediate the confidence one can employ in using these scales in isolation in determining the locus of a lesion. As a result, one should only employ these scales in differentiating hemispheric impairment if other collaborative evidence attests to the locus of a lesion (e.g., neuroimaging [CT/MRI] evidence, etc.).

Second, numerous studies attest to the validity of the VC and PO factors both with the WISC-R and WISC-III. In fact, since these factor scales are relatively more "pure" in a psychometric sense, one might expect that they could be more useful in confirming the presence of intact or impaired VC or PO abilities.

Some experimental evidence, in fact, supports this notion. For example, Morgan, Hynd, Hall, Novey, and Eliopoulos (1997) reported a correlation of .63 (p < .05) between the length of the left temporal bank of the planum temporale and the VC factor score from the WISC-III among normal children. Figure 10.1 illustrates the location of the left planum temporale, which has historically been associated with Wernicke's area, known to be vitally linked to language comprehension. In a larger sample of normal, ADHD, and dyslexic children, Morgan et al. (1997) found that irrespective of group, as plana asymmetry increased in favor of the left plana (L > R), so did ability on VC factor on the WISC-III. Figure 10.2 illustrates how the more normal pattern of left greater than right plana asymmetry increased when the VC factor score is >100. It can be seen that among children scoring <100 standard score on the VC factor, only 25% had L > R plana asymmetry. However, 61% of children scoring >100 standard score on the VC factor had left plana greater than the right. Thus, it might be concluded tentatively that the VC factor score may have a link with the length of the left-hemispheric language cortex such that better ability is associated with a larger left planum temporale.
FIGURE 10.1 A magnetic resonance imaging scan showing a lateral sagittal view of the left cerebral hemisphere. The region of the planum temporale lies in the posterior end of the Sylvian fissure between the two white arrowheads.

FIGURE 10.2 Planum asymmetry and the WISC-III Verbal-Comprehension (VC) factor. The incidence of greater left over right (L > R) asymmetry increases when the VC score is >100 standard score (based on N = 47 children). This may support the notion that left planum length is positively associated with greater verbal comprehension abilities in children.
Third, it is clear that insufficient research exists regarding the usefulness of the third and fourth factors of the WISC-III. Thus, in a neuropsychological context, clinicians should not make statements regarding the process these factors and their associated subtests assess in relation to neuroanatomy. Rather, they should be interpreted as revealing process abilities which to date have not been related to specific brain regions. This seems an appropriate caution considering the lack of validation regarding these third and fourth factors on the WISC-III.

Overall then, we must reiterate that the VIQ and PIQ scores as well as the VC and PO factors may have relevance in revealing the intactness of neurological systems associated with the right and left hemispheres but that caution should be applied in making statements regarding the integrity of neurological regions or structures.

General ability, or psychometric g, has no locus as it reflects the ability to perceive associations across neuropsychological abilities and processes as well as more localized subprocesses. However, because general ability may be impacted by dysfunctional processes, general ability may be sensitive to the intactness of the functioning of the brain and its associated processes. Clearly, as more advanced forms of neuroimaging, such as functional magnetic resonance imaging, become more widely used by the community of scholars, future research may reveal more about the locus of all of the mental subprocess assessed on tests of intelligence, such as the WISC-III.

ACKNOWLEDGMENTS

Supported in part by a grant (RO1-HD26890-03) awarded to the first author (GWH) from the National Institute of Child Health and Human Development (NICHHD), National Institutes of Health (NIH).

REFERENCES


11

Assessment of Minority and Culturally Diverse Children

Antonio E. Puente and Gabriel D. Salazar
Department of Psychology
University of North Carolina at Wilmington
Wilmington, North Carolina

INTRODUCTION

"Are the inferior races really inferior, or are they merely unfortunate in their lack of opportunity to learn? Only intelligence tests can answer these questions" (Terman, 1916, p. 20). By 1932, alternative suggestions for addressing the issue began to appear in the psychological literature. Sanchez (1932), often considered the founder of Chicano psychology, published in the Journal of Applied Psychology that mental testing biases existed against Mexican children. Well over half a century later these questions remain unanswered, but the debates continue in the public and professional sectors. The role of culture, ethnicity, and race are central and controversial issues in both the definition and measurement of intellectual functions. One need not look further than the recent publication of Herrnstein and Murray's (1994) The Bell Curve for illustrations of the importance and timeliness of this topic. The relevance of these variables in the unbiased assessment of intellectual and general cognitive abilities is critical (Betancourt & Lopez, 1993) and are highlighted in the most recent American Psychological Association (APA) guidelines on this topic. In 1993, APA published the "Guidelines for Providers of

Parts of this chapter, including the section on Aboriginal and Native American Children, were written by D. H. Saklofske.
Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations” (APA, 1993). Interestingly, the focus of those guidelines appear to be centered more around intervention rather than assessment practices.

The purpose of this chapter, despite the importance of addressing more central issues of the role of ethnicity, race, and related variables on testing (see Puente, in press, for a broader review), will be to examine the relationship between ethnicity and race and the three Wechsler scales for assessing children’s intelligence, especially the earlier versions of this test where much of the research on this issue is found. Furthermore, as a means to better understand the role of these complex variables on cognitive functions as measured by the Wechsler tests, special emphasis will be placed not on race but on ethnicity. If Lewontin, Rose, and Kamin (1984) are correct, race accounts for approximately 6% of genetic variability. Hence, it is difficult to account for large-scale differences between groups using race as a factor. Indeed, within-race variability, if different ethnic-racial group configurations are considered, is greater than between-race variability, especially if no large-scale cultural variation (e.g., adding Whites from two different countries such as United States and Russia) is introduced.

Here it will be assumed that whatever differences exist in intelligence and cognition are more culturally than racially determined. This assumption does not negate or even presuppose the existence of racial differences in intellectual functions. What it does do is help to reconfigure reported racial differences based on cultural rather than biological variables. Thus we believe that culture may have a greater impact on cognitive functions than race. Further, it could be that what is measured on standardized tests of cognitive abilities when race is an independent variable is simply the culture associated with the race rather than race itself. In other words, even though data are sparse on this topic, we believe that Blacks from two quite different cultures (e.g., USA and Zaire) may be more dissimilar than Caucasians and Blacks both living in similar geographic regions of the United States. This issue has been raised in descriptions of Aboriginal children where the belief may be held that there is, for example, only one type of hunting-gathering society. In contrast, the socioecological context of Aboriginal children is most important in describing their intellectual, behavioral, and social characteristics. Murdoch (1988) contends that whether Aboriginal children are from “hunter-gathering,” “pastoral,” or “agrarian” backgrounds has critical implications for the types of sociocultural adaptations they manifest. Furthermore, “the cultural and linguistic experiences of most Native American children, however, differ considerably from those of the middle-class, monolingual, English-speaking students upon whom most standardized intelligence tests were normed” (Tanner-Halverson, Burden, & Sabers, 1993, p. 125).

The pioneering studies of Scarr (e.g., Scarr & Weinberg, 1976) also suggest that culture may be more critical than race in determining differences on tests. This view has been further elaborated by Moore (1987), who found not only that the ethnicity of the family but also the ethnic milieu help explain IQ score differences. It is not only the family culture but the culture in which the child interacts.
daily that help to explain cognitive test differences. Further support for this departure from conventional thinking is warranted based on the overstating of race as a salient factor in intellectual functioning (Rowe, 1994) and the understating of the impact of culture on basic cognitive functions (Ardila, 1995). The work of Ardila (Ardila, Rosselli, & Puente, 1994) has also suggested that culture plays a critical role in the measurement of basic neuropsychological functions.

This readjustment from a racial to a cultural perspective may help further clarify some of the existing literature on this topic. However, an initial step in this approach is to define the differences between race, ethnicity, and culture. The next section of the chapter outlines the definition of culture, race, and ethnicity, especially as it pertains to the issue of how children from a nonmajority group (e.g., African Americans) fit into a majority group (e.g., White, Anglo-Saxon).

DEFINING RACE, ETHNICITY, AND CULTURE

Besides the fact that these variables have traditionally not been well researched in psychology, there is a general confusion about their differences. Although Jones (1991) has argued that race is difficult to operationalize, there is some agreement as to how to generally describe races. Obvious biological characteristics include skin color, facial features, and hair type (Betancourt & Lopez, 1993). However, other variables might include size and other related physical characteristics. According to Brislin (1989), there are three races; Caucasian, Negroid, and Indian (with two subcomponents, Asian and American).

Ethnicity is more diffuse and is behaviorally rather than biologically based. Specifically, ethnicity refers to a set pattern of behaviors that might include rituals, beliefs, customs, common ancestry, as well as family, social, and even marital restrictions. In contrast, culture is a wider defined pattern of behavior, which is generally more widely accepted across a number of ethnic groups. These patterns are more ingrained, socially less questioned, and often considered to be universal in scope. For example, Western culture is very focused on time, and more ethnic groups in the Western culture consider time to be of critical importance in everyday life. However, different ethnic groups view time in different ways. Ethnic groups in northern climates, especially with European ancestry, consider time a critical component in a variety of factors, ranging from intelligence to common courtesy. In contrast, ethnic groups closer to the Equator view time as something that occurs, and secondary to the enjoyment of the activity in question. Thus, it is not surprising that when time, namely speed, plays an important role in arriving at a conclusion regarding cognitive ability, people with ethnic origins close to the Equator may appear less intelligent than those with ancestry in northern climates.

It is important to emphasize that these three variables are not necessarily unique or independent. Indeed, the opposite may be more true than the principles of experimental methodology would allow to be easily detected. For example, there are Blacks that have their ancestry in the warm climates just as there are several
generations of Hispanics that have been raised strictly in northern climates (e.g., New York, Chicago). Furthermore, there is the complication of socioeconomic status. Laosa (1984), among others, reported that economic status is often not controlled for or measured, but when it is this variable has a critical impact in explaining the effects of race, ethnicity, and culture. This important issue has also been raised in chapter 1 of this book.

Another issue raised above that is rarely addressed is that of subgroup heterogeneity. Not all Whites have similar ethnic backgrounds. This is best illustrated with Hispanics. According to the Census Bureau, there are at least five different types of Hispanics. Cuban, Mexican/Mexican-American/Chicano, Puerto Rican, Hispanic Latin American (e.g., Panamanian, Peruvian, Venezuelan, Ecuadorian, Guatemalan), and Spaniard (originally from Spain). Table 11.1 provides a more specific breakdown according to the different ethnic groups as well as the 1990 Census data. As this table indicates, blacks outnumber other groups but only by a relatively small amount when compared with Hispanics. When considering the different subgroups, the question arises as to whether within-group heterogeneity could tend to obscure between-group differences. For example, when Hispanics...
are subdivided into smaller groups (e.g., Puerto Ricans, Cubans, etc.) it appears that Cubans parallel the general majority culture, whereas Puerto Ricans show the most differences (Garcia & Marotta, 1977).

**CONTROVERSIAL ISSUES**

Underlying the potential differences in performance on intellectual tests, in general, and the Wechsler tests in particular, are a host of variables that differentially affect cognitive functions and functioning. These variables have been the focus of two major if not controversial reports in recent years. In the highly controversial book, *The Bell Curve*, Herrnstein and Murray (1994) present some provocative ideas and interpretations that warrant attention when attempting to identify ethnic group differences in intellectual performance. A major premise is the rise of the "cognitive elite." This social class, highly predicted by intelligence test scores, are physically segregated through educational and vocational channels and institutions. Their review of the literature suggests that Asians have higher IQs, whereas Hispanics score ½ to 1 standard deviation, and Blacks about 1 standard deviation below Whites. Intelligence test findings for North American Aboriginal samples often show further that score differences exist not only in culturally loaded items but also in more neutral items. For example, differences are evident in both digit span forward and backward, and more so on the backward items. Such differences, according to Herrnstein and Murray, are due to motivation, knowledge of standard English, success and failure expectations on these tests, and so forth. The authors argue that up to 60% of the variance in intelligence test scores appears to be inherited and, thus, intelligence may not be that malleable through environmental interventions. To complicate matters, the authors propose that individuals with social problems are heavily represented in the lower portion of the cognitive ability distribution. Without actually resolving matters, this book has spurred active discussions on the historically controversial topics discussed by Galton during the 19th century and by Terman at the beginning of this century.

Probably the major and most official response from the psychological community was the report "Intelligence: Knowns and Unknowns" published by a Task Force appointed by the Board of Scientific Affairs of the American Psychological Association (Neisser et al., 1995). In this report, a number of major topic areas were reviewed, including the concept of intelligence, intelligence tests and their correlates, genetics and intelligence, environmental effects on intelligence, and group differences. In a particularly important section of this paper, ethnic group differences were considered. One possible way to explain the apparent differences between ethnic groups was to consider socioeconomic and related cultural factors. The authors concluded, "Thus the issue ultimately comes down to a personal judgement," (p. 35) and they add that at present scientific answers for these differences are still elusive.

However, the Neisser et al. report and the Herrnstein and Murray book have had critics and criticism. According to Lane (1994), the sources cited in *The Bell*
Curve, appear to be "tainted," thus resulting in biased and incorrect conclusions. Lane stated that a large number of sources are from individuals associated with the periodical The Mankind Quarterly and the Pioneer Fund, both associated with "race betterment," presumably of the white race. A number of in-depth reviews of this book were published in the School Psychology Review (volume 24, number 1) in 1995. The Neisser et al. report has similarly been critiqued, though not as aggressively. For example, as part of series of commentaries in the American Psychologist, Rushton and Yee (1997) suggested that IQ differences need to be considered as a function of not just environmental but hereditary issues. Yee also (1997) described the suggested links between IQ and socioeconomic status and race.

TECHNICAL ISSUES

The proposed differences between different ethnic groups may be partially or even largely attributable to methodological issues in research. It could be that such differences exist not because of internal variables that are neurologically and genetically based determinants of cognitive capacity but because of external variables, such as test content or administration. For example, Terrell and Terrell (1983) reported that race of examiner appears to have an effect on Black children's performance on standardized cognitive tasks. Saklofske and Janzen (1990) have described the potential problems that might arise when tests developed and standardized in one country are then used in another country. Cicchetti (1994) provided an interesting perspective on how to evaluate normed and standardized instruments in psychology. His approach included careful attention to standardization procedures, norming procedures, test reliability, and test validity. However, no reference in the text nor in the citations is made to potential issues of bias so comprehensively discussed in the psychological literature (e.g., Reynolds, 1995). However, the usual criteria for evaluating the efficacy of instruments appears a good one from which to begin the difficult task of examining the efficacy of intellectual tests for ethnic-minority children. This section attempts to address some of the more salient methodological issues that could help account for the differences described in The Bell Curve and the journal article, "Intelligence: Knowns and Unknowns" but which are often ignored in the applied world of assessment (a particularly thorough assessment of test validity as applied to Hispanics is found in Geisinger, 1992).

Padilla (1988, 1995) Olmedo (1981), and Westmeyer (1987) are modern pioneers following in the footsteps of Garcia, who in the 1930s suggested that people from minority groups (e.g., linguistic, racial, etc.) have to be understood from a different vantage point. To do otherwise would confound "abnormality" (statistical or clinical) with culture-based behavior patterns or communication difficulties. According to Olmedo (1981), when linguistic minorities are tested several factors are often ignored: type of test used, socioeconomic class, degree of bilingualism, language-based factors, ability to communicate in the nonmajority language, ac-
culturation, and cultural equivalence. Pragmatic factors in these circumstances include language proficiency, language of test examiner, language used in the evaluation, translations, translators, etc. For example, Hanley and Barclay (1979) reported that race of the child and the tester interacted negatively on WISC and WISC-R scores. Cross-cultural researchers have sensitized psychologists to these observations. “Emic” and “etic” were coined by Pike (1966) to represent two viewpoints in the study of human behavior. The “etic” viewpoint seeks to discover universals in a system, but when universals are assumed, this is termed “imposed etic.” It may be argued that “great risk attends the use of an imposed etic, since there would be no way of knowing whether it makes any sense to use it in any culture than that of its origin” (Berry, Poortinga, Segall, & Dasen, 1992, p. 54).

The “emic” viewpoint studies behavior from within the cultural system. Here it is recognized that understanding can only occur in reference to the context in which behavior takes place. Thus an intelligence test that is rooted in one culture but used as though it was valid for some other culture may result in very inaccurate and even tragic conclusions. Examples of potential item bias on the WISC-III for Native American children is described by Tanner-Halverson et al. (1993) and for Canadian Aboriginal Children by Greenough-Olson (1993). The assumption that mainstream culture, socialization, and language are the same for Native children as for the White children is an imposed etic. In fact, the research literature shows some quite consistent findings for the performance of Native children on the WISC-R and WISC-III, with the general trend being lower Verbal IQ (VIQ) than Performance IQ (PIQ) scores (e.g., McShane & Plas, 1984; Scaldwell, Frame, & Cookson, 1985; Wilgosh, Mulcahy, & Walters, 1986), as well as lower Full-Scale IQs (FSIQs) in relation to the standardization data (ie., $\bar{x} = 100, SD = 15$); (e.g., Tanner-Halverson et al., 1993).

From these complex set of variables, possibly a common factor could be extracted. Beyond the obvious problems of communication (especially when English is not the original language) is the underlying issue of cultural equivalence. Although much attention has been given to addressing the issue of test bias, whether due to socioeconomic status, language, or whatever, the question of intent still remains. Specifically, what is the goal or purpose of intelligence testing. Helms (1992) has argued that ethnic differences in cognitive ability are actually differences in culture, and no necessarily due to either biological or environmental determinants. In other words, if the goal is to measure the intellectual ability of the child, then testing a Spanish-speaking child in English, or an Aboriginal child from a remote northern settlement by a White examiner from the ‘deep’ south, may result in a less than contextually sensitive assessment. In contrast, the alternative might be to test a Spanish-speaking child’s ability to understand the intellectual demands of the Anglo-Saxon culture. The second intention would not reflect a bias but rather a different and more difficult goal.

Defining the goal of assessment establishes a direction and a set of guidelines from which one can carry out the necessary testing. If the question is one of intellectual acculturation, then testing variations such as the use of English IQ tests
for Spanish children might be appropriate. If the question relates more to an assessment of intellectual ability, then matching the test to a larger set of intellectual concepts that are global rather than cultural would appear to make more sense. La Framboise, Coleman, and Gerton (1993) provide support for a psychological viewpoint on biculturalism and a perspective from which to address this problem. Thus an initial and critical issue is to determine what the goal of the intellectual testing might be. In most cases, the location or position of the child within an intellectual spectrum as defined by the majority culture would be one obvious choice. However, underlying this criterion would be an important secondary question; what capabilities does the child have that will enable him or her to assimilate and accommodate critical 'data' from the majority culture?

Several important technical variables may influence our description of the child within a cognitive framework. Regardless of ethnic background, these variables have received some attention in the literature. The variables addressed here include samples tested, language and communication, and acculturation.

SAMPLES

It is often assumed that a child does not belong to a majority group (i.e., the reference group), they belong to a homogeneous minority group. Homogeneity is often associated with limited variation in a host of variables, including race, language, socioeconomic background, and even religion in some instances. Such an approach is fraught with problems. Chrisjohn and Lanigan (1986) have commented on the frequent presence of "Pan-Indianism" in the research literature. Pan-Indianism refers to the treatment of members of different native nations as a homogeneous group. However, they argue against this viewpoint by stating that not a single study has demonstrated a universal Aboriginal cognition. It is this tremendous variability in intelligence test scores that has led Brandt (1984) to state that "the high degree of variability of scores in Native American populations is the major reason researchers have little confidence in the WISC-R as an assessment instrument" (p. 75).

To further illustrate this point, Hispanics are a nonmajority group composed of a variety of subgroups, including Mexicans, Puerto Ricans, Central Americans, Cubans, and so on. Furthermore, these individuals may either be black or white, yellow, or some combination thereof (most often seen in the Caribbean). Finally, these individuals may be monolingual, bilingual, or mixed. Thus, when a test is normed on Hispanics, African Americans, or Aboriginal peoples in either Canada or the United States, for example, a number of problems are bound to arise. An African American who lives in affluent America with an ancestry going back several generations in the United States would be more likely to perform like a White person with an affluent background than a newly emigrated Haitian Black with limited resources and a different cultural history. Thus, it is extremely important to sort through these kinds of variables when considering the appropriateness of both the test and the test norms.
Many of the issues raised throughout this chapter mainly focus on black and Hispanic children, in part due to the large numbers they represent in the WISC-R and WISC-III standardization samples and in the general population of the United States. However, it is important to note that these issues in intelligence testing probably apply to other ethnic minority groups as well. Further, studies are now available (e.g., Yang, Su, Qhang, & Ta, 1995, see also chapter 1, by Prifitera et al., this volume) that have examined the Wechsler test performance of children in different countries and even translated the WISC-R and WISC-III into other languages.

**COMMUNICATION AND LANGUAGE**

The assumption is made in any testing situation that meaningful communication has been effectively established between the examiner and the test taker. Furthermore, when communication has not been established an assumption may be made that language differences exist (e.g., the English-speaking psychologist is testing a Spanish-speaking child), which may then hinder the meaningful communication of instructions and concepts (Malgady, Rogler, & Constantinio, 1987). This communication, which is most frequently verbally established, is not the sole means of transmitting information. A host of other communication variables exist; social boundaries, use of slang, eye and physical contact, and the relationship between the tester and the examinee when cultural and linguistic differences are present. Thus, communication is obviously multifaceted, with far-reaching effects on the measurement of intellectual functions. One such effort to address this issue was the somewhat controversial publication of *The Black Intelligence Test of Cultural Homogeneity*, which contained 100 vocabulary words used mainly in African-American settings (Williams, 1975).

Bilingualism may also present a problem when assessing some children. It is often commonly assumed that if a child can speak enough English to take an IQ test, then they should be proficient in that language. Laosa (1975) suggested that this is an oversimplification of a more complicated situation. In interviewing nearly 300 Hispanic children and their families, he found the following patterns of communication; English is the single most frequently used language, Spanish is the single most frequently used language, both English and Spanish were used without mixing, and English and Spanish were equally used with mixing.

Thus assumptions that are made about a child’s proficiency in one or more languages could actually be wrong. Hickey (1972) tested two groups of 100 students, monolingual (Spanish and English) and bilingual. He reported that Mexican-American children had difficulty associating English verbal nouns with pictures because of the differences between the two languages. Manuel-Dupont, Ardila, Rosselli, and Puente (1992) have argued that bilingualism is a complex concept rooted both in sociocultural traditions and in neurological substrates. Such variables as the method of language acquisition, age and sequence of acquisition, and the structure of the languages in question all affect the ‘imprinting’ of language in the brain and subsequent use (e.g., speaking, comprehension).
Finally, it is important to appreciate that bilingualism is not restricted only to obvious differences in the language. In other words, subtle differences in verbal communication (e.g., standard and black English) are probably sufficient to result in some of the problems and issues outlined in this section. As early as the mid-part of this century language differences between Welsh and English-speaking children in the United Kingdom had been reported (Jones, 1952). A recently published statement on conducting assessments of non-English individuals (de Jesus, Perrin, and Blackwood, 1996) provides an overview of general principles of communication with individuals of a nonmajority group culture.

**ACCULTURATION**

Perhaps the most salient factor that is related to all the previous variables discussed above is that of acculturation. Acculturation in this instance is not considered in the classical sense but instead is viewed in a broader context, as described by Laosa (1991). He identified the problem of construct validity in the testing of minority population as a critical ethical problem. Specifically, the assumption is made that the intellectual domains of the test in question reflect a cultural 'g' of some sort. The test could therefore be biased in a very specific sense. For example, one of the WAIS-R Picture Completion items involves a map of the United States without the state of Florida. Even very impaired or culturally limited individuals living in Florida were able to obtain a correct response, whereas the same could not be said of similar individuals living in North Carolina. Similarly, Aboriginal children from remote settlements in northern Canada often give very different answers to test items, such as what they should do if they see a house on fire. Here there are no fire departments, police, or phones to call 911.

The question of what is intellectually salient may be more critical. This might include very basic questions involving some of the following variables: response to authority figure (tester), acknowledgment and manipulation of time as a critical element, expression of confidence as either an unwarranted coping mechanism or a lack of courtesy to others, and the understanding of complex cultural concepts. Also, the role of sociocultural variables needs to be considered. Barona, Santos de Barona, and Faykus (1993) reported that these variables accounted for a significant degree of the variance in Mexican-American students diagnosed with mental retardation.

Ellis (1990) examined cross-national comparisons of intelligence using translated versions of the Wilde Intelligence Test and the Career Ability Placement Survey. She concluded that cross-national comparisons of intelligence and abilities might result in incorrect conclusions when translation equivalence is not established. The question of what is important and relevant for one majority group versus another must also be considered together with the question of cultural equivalence. Understanding the basic concepts of the culture in question is essential to appreciating its similarities and differences with other cultures. For exam-
II. WISC AND ETHNICITY

pie, Nobles (1995) has provided an interesting background of what is called "African philosophy." This philosophy provides a critical backdrop from which to understand culturally based intellectual concepts.

The concept of equivalence is not restricted just to the words of a language but to a host of other factors all contained under the general rubric of culture. Casagrande (1954; in Bontempo 1993) has identified four types of translations. These include pragmatic, aesthetic-poetic, ethnographic, and linguistic. It is assumed that when a test is "translated" by a test administrator for a child, what is typically occurring is pragmatic translation. The question involves something like, "We need to obtain a general idea of the intellectual abilities of this child... do what you can." This approach reduces translation fidelity (Bontempo, 1993) and increases measurement error. It is for all of these reasons and more that different cultural and language groups contend that the commonly employed tests and testing practices in schools may provide an inaccurate and unfair assessment of particular examinees. For example, in Canada, the Saskatchewan Indian Institute of Technology (SIIT, 1990) recognized this problem in their Aboriginal Literacy Action Plan as follows:

Most of the testing instruments are inappropriate. They are culturally irrelevant and geared toward white middle class society. Indian educators should be able to develop their own testing instruments using means and methods that will be relevant and familiar to Indian Students. The testing that our Indian students are presently being subjected to is another reason for their frustration and discouragement, because they do not do well on these tests. (p. 44).

ASSESSING CHILDREN'S INTELLIGENCE
WITH THE WESCHLER SCALES

Before describing the findings from studies of the WISC performance of children from different groups, several relevant issues regarding such test findings will be raised here. First, the study of intelligence across cultural, ethnic groups may be viewed from within three broad paradigms described by Berry (1984). The general intelligence paradigm assumes that intelligence is a single construct that is common across all people so that comparing culturally different groups will indicate who has more or less of this general intelligence, if there are differences at all. The "specific abilities" paradigm reflects our emic perspective, in which the emphasis is on the culturally relative nature of cognition. Here there is no assumption relating to universal patterns of intelligence so that cross-cultural comparisons are not especially relevant or meaningful. The cognitive styles paradigm is also based on a position of cultural relativism but in addition searches for systematic connections among abilities. For example, More (1987) has outlined the cognitive style characteristics of Aboriginal children while also recognizing that there are many within-group differences. These differing paradigms remind us
that studies of the WISC across cultural, ethnic, and even racial groupings reflects but one perspective in the study of cognition and intelligence.

Another major issue related to the cross-cultural study of intelligence is tied to the contentious problem of bias. While Jensen (1980) is often credited with “leading the charge” on the problem of bias and ethnic differences in intellectual tests and assessment, much had been written before him (Kauffman & Doppelt, 1976; Reschely, 1978; Sandoval, 1979). The early writings of Terman (Terman, 1916; Terman and Merrill, 1937) provide an interesting insight into the difficulties surrounding this issue. More recently Sattler (1988) has summarized some of the key themes thought to underlie test bias. The finding of differences in mean scores between two groups has been suggested as evidence of test bias. However, Sattler contends that such differences are not indicative of test bias since differences between the test scores of a minority group and the majority group may reflect differences in socioeconomic status. Another suggestion that a test may be biased is related to how good a predictor it is of some criteria across two or more groups. A third criterion for evaluating test bias relates to whether the instrument measures the same abilities across different groups. The items or test content have been suggested as a cause of that bias. Reynolds and Wilson (1983) has also included examiner and language bias and inappropriate standardization samples among the reasons for suspected bias and criticism of intelligence tests.

Two of these points will be briefly commented on to illustrate how they relate to the Wechsler Scales for assessing intelligence in children. Although the WISC-III (Wechsler, 1991) was published in 1991, most studies involving ethnic-minorities are based on data from the WISC and the WISC-R. Hence, the majority of discussions center around these studies, although chapter 1 of this book reexamines the WISC-III performance of Black and Hispanic children. It is important to note that despite the relatively minimal bias that had been reported for earlier versions of the WISC, the WISC-III used item-bias statistics to eliminate potentially biased items (primarily restricted to Information, Vocabulary, and Comprehension subtests) and also obtained expert’s reviews of items for potential bias.

The sample obtained for the WISC-III reflects the 1988 United States Census survey. A particularly interesting approach to ethnic identification involved the use of the parents of Hispanic children to identify their children as Hispanic or otherwise. Furthermore, Hispanic ethnicity was not confounded with race. Blacks were sampled in exact proportion to their representation by age group, whereas Hispanics were very closely sampled relative to their representation in the different age groups. Also, similar representation was noted for ethnicity by geographical region, which is a particularly difficult task in a large sampling study of this type. Furthermore, during standardization, an extra 400 minority children were used. This effort did not go unnoticed by Kaufman (1993), who reported that of the seven major changes in the WISC-III, one involved the new (and presumably comprehensive) standardization with a better definition of ethnicity.
AFRICAN AMERICANS

Based on allegations that inappropriate educational placements had occurred as a result of the test evaluation of two African-American girls in Chicago, the Chicago School Board and eventually Judge Grady ruled over a decade ago that tests of intelligence, including eight items from the WISC and WISC-R were so culturally biased that their use was considered inappropriate (Koh, Abbatello, & McLoughlin, 1984). This was a particularly problematic issue in light of the fact that the revision of the WISC had taken into account perceived racial differences (Weschler, 1974). Yet, according to several studies by Munford and colleagues (e.g., Munford, Meyerowitz, and Munford, 1980), the differences had actually increased. However, in more refined studies with larger samples, alternative conclusions have been reached. For example, Koh et al. (1984) administered the eight alleged biased items in the Grady decision to 360 educable mentally handicapped (EMH) White and Black children. The results indicated that "the children who constituted this sample could not be discriminated on the basis of ethnicity" (p. 93). Factor analysis of the WISC-R have also been conducted for white and black children (Gutkin & Reynolds, 1981). The results suggested similar factor structures for both sets of children, and according to the authors provide support for the "growing body of research supporting the construct validity of the WISC-R across race" (p. 230).

These studies could be considered in direct contrast to those of other researchers. Kaufman and Kaufman (1983) reported up to 16-point differences between Black and White children on the WISC-R. Others (Naglieri, 1986) have reported somewhat smaller differences (i.e., 9 points). Slate and Jones (1995) recently reported on the validity of the WISC-III for African-American students undergoing special education evaluation. In general, and as expected, the results suggested lower scores for the WISC-III when compared to the WISC-R. A salient explanation for such discrepancies probably lies both in the questions asked and the methodologies used as well as the newer norms for the WISC-III. To compare studies of racial differences on the WISC by contrasting, say, the Kaufman and Kaufman (1983) with the Koh et al. (1984) results seems inappropriate at best. Indeed, in this instance one study uses the entire test, the other eight items; one uses "normal" children, the other uses EMH. Furthermore, other salient variables are not controlled. Considering the earlier discussions in this chapter, particularly important variables are not considered or even controlled for, such as socioeconomic status, parental educational achievement, race matching between tester and test taker, and so forth.

HISPANICS

According to Figeroa (1983), the argument that bias is not present in the WISC is due primarily to the adopted model and definition of bias and not the lack of bias itself. The assumption that underscores the item bias theory tested by Sandoval (1979) was that there is a specific and relatively inflexible learning curve of
language and cultural knowledge across a society. Further, that learning curve is shared when others, including those with different language backgrounds, do not interact with the majority culture on an active and ongoing level. Jensen (1980) had already suggested that bias might occur when intelligence is measured in a language different than the original language of the test. Thus, one could argue that even if the tests themselves were equivalent, they would not be in other and presumably more important ways (e.g., knowledge of the culture in question). Hence, it is not surprising that despite the fact that the WISC-R was intended to be less biased than the WISC according to Wechsler (1974), that indeed the discrepancy between Hispanics and Anglos was actually greater with the WISC-R.

An alternative to this situation was to translate and standardize the WISC into Spanish (Rodriguez, de Torres, Herrans, & Aponte, 1994). This was done with varying results. In reviewing the literature, the work of Prewitt-Diaz, Rodrigues, and River (1986) provides a glimpse into the problems in question and reflects the difficulties previously outlined. The Escala de Inteligencia de Wechsler para Ninos was intended to be a test which could be used across all Spanish-speaking children. Unfortunately, the sample was based only on Puerto Rican children. A more recently published and updated version of this test reflects many of these same problems. Some of the questions did not appear to have wide generalizability to other Hispanic subcultures because the items were reflective of only indigenous Puerto Rican culture. However, the newer version of this test (Rodriguez et al., 1994) provides updated norms, alternative and better instructions for Similarities and Digit Backwards, as well as improvement on several specific items. For a historical analysis of mental testing with one group of Hispanic (Mexican) children, the reader is referred to Padilla (1988).

ABORIGINAL AND NATIVE AMERICAN CHILDREN

Although the research and clinical literature on the Wechsler test performance of Aboriginal children is relatively small, it does serve to raise issues related to differential test performance, test bias, and research methodology. Such findings are most relevant in the clinical interpretation and reporting of intelligence test data for children of Native ancestry.

A number of investigations have tended to report a pattern of lower VIQ in contrast to more average PIQ scores on the Wechsler scales (e.g., McShane & Plas, 1984). However this general finding must be further viewed within the context of other observations. St. John, Krichev, and Bauman (1976) tested 160 Cree and Ojibway children and youth on the WISC and WAIS and found that VIQ ranged from 69.7 at 6–7 years to 91.1 at 18–20 years. In contrast PIQ ranged from 99.8 at 9–10 years to 103.4 at 18–20 years. This large Verbal-Performance (V-P) discrepancy was found among the youngest children and decreased with age. Language spoken at home was significantly related to VIQ scores. Seyfort, Spreen, and Lahmer (1980) administered the WISC-R to 177 Aboriginal children from three different West Coast Canadian bands. Again average PIQs and lower VIQs
were reported but more important was the finding that a large number of test items did not contribute to the total test score or score variance. Similarly Mueller, Mulcahy, Wilgosh, Watters, and Mancini (1986) reported other item difficulty data suggesting "that these figures for the Inuit sample represent a significant increase in the overall WISC-R difficulty compared to that shown by majority children" (p. 35). Wilgosh, Mulcahy, and Walters (1986) also examined the WISC-R scores of 366 Inuit children and observed that 77% earned VIQs less than 70 in contrast to only 5.7% with PIQs less than 70. FSIQs of less than 70 were attained by 32% of the children. The Information and Vocabulary subtests accounted for the majority of items unanswered or answered incorrectly.

Turning to studies of the predictive validity of the WISC, St. John et al. (1976) found that the WISC-R VIQ and PIQ was significantly correlated with year-end school grades for only one of the four age groups in their study of Aboriginal children. A study comparing the predictive validity of the WISC-R for samples of Anglo, Black, Chicano, and Native American Papagos children found that FSIQ and the Verbal Comprehension factor were the best predictors of achievement defined by teacher ratings and the Metropolitan Achievement Test (Reschly & Reschley, 1979). However, the validity coefficients were lowest for the Native American children in comparison with the other three groups. These cumulative research findings led McCullough, Walker, and Diessner (1985) to conclude that caution is advised in the use of the WISC-R and WAIS with Native Americans. Significant Verbal—Performance deviations have been found across the tribes. The predictive validity of the Wechsler tests for academic achievement may vary across the Native American Cultures. (p. 29)

Several published reports have focused on the construct validity of the Wechsler Scales for children. Mulcahy and Marfo (1987) suggested that factor analytic studies of the WISC-R with Inuit children generally support the construct validity of the test for children aged 12 to 15 years but not for those 7 to 11 years of age. Chrisjohn and Lanigan (1986) argue that there is a lack of research substantiating the construct validity of the WISC-R when used with Aboriginal groups. They contend the following:

The WISC-R may indeed measure intelligence in non-Indian populations, but fail to measure it in Indian groups. Or the Performance Subtests of the WISC-R may measure intellect well enough and the verbal subtests not. Or the WISC-R may work for "acculturated" Indians and not for less acculturated groups. Mean comparisons under the condition of not knowing whether the test behaves equivalently in experimental groups are largely meaningless. (p. 7)

Finally, the argument that factors outside of the test may impact on the WISC performance of Aboriginal children, as with other culturally different children, has been raised by various authors. Sattler (1988) states that whether the use of a particular test in a particular situation results in discrimination . . . will depend on such factors as the purpose to which the results are put, how the results are interpreted, and how the test is administered. (p. 568)
Many of the same factors that pertain to studies of the WISC performance of Black or Hispanic children may be raised in relation to Aboriginal children, such as the uniqueness of cultural experiences, linguistic differences, health issues, and factors associated with the testing experience. McShane (1983) cites the high incidence of otitis media as the “single leading identifiable disease among Indian populations” (p. 37), which in turn can compromise efficient language learning (Friel-Patti, 1990). McShane also notes the higher incidence of vision problems of Native children and the problem of fetal alcohol syndrome or fetal alcohol effect associated with the problem of alcoholism in some communities. Tanner-Halverson et al. (1993) argue that the cultural and linguistic experiences of Native-American children are quite different from the experiences of English-speaking, middle-class children. They contend that the WISC-III has certainly shown improvements over the WISC-R, but there is still the potential for other bias. Furthermore, they raise questions regarding the sampling strategies and test norms and contend that although this is the proper representation of these minority groups, the scores derived from the national standardization norm tables by no means assume that this will be unbiased for the minority group. (p. 126)

In order to address this issue, Tanner-Halverson et al. have suggested that local norms be generated for the WISC-III and provide data for 110 randomly selected Tohono O'Odham Native American children.

Insensitivity to cultural differences may not only impact on the integrity of test administration but also interpretation (Wilgosh, Mulcahy, & Waiters, 1986). Saklofske and Schwean-Kowalchuk (1992) have discussed a number of factors that may impact on the test performance of children, ranging from race of examiner-examinee to test anxiety. Common and Frost (1988) conclude that misdiagnosis is an apparent danger when tests such as the WISC are used with Aboriginal students, because of the kind of factors described above. Although it may be argued that a “theory of Indian intelligence must eventually be constructed from within Indian ranks, with Indian perspectives and concerns reflected in its development” (Chrisjohn & Lanigan, 1986), it is imperative that research examining the test performance of Aboriginal children continues and that culturally sensitive test administration, scoring, and interpretation is “the order of the day.”

**CROSS-GROUP COMPARISONS**

Considering that everything in science is relative and the efficacy of a comparison hinges on the integrity and representativeness of the group in question, an alternative to simply examining how single ethnic groups compare to Anglos would be to compare how different ethnic groups compare to each other as well. A study by Sandoval, Zimmerman, and Woo-Sam (1983) is reflective of the complexities and subtleties in cross-group comparisons. The WISC-R was administered to
7½ and 10½ year old Anglos, African Americans, Chicanos, and Bermudians. The results indicated that item difficulty curves were "remarkably parallel." In addition, similar patterns were reported when the factor structures were compared across different ethnic groups. For example, Reschly (1978), Reschly and Reschly (1979), and Sandoval (1979) reported that the factors do not vary much and that the correlation's between IQ's and subtest scores were also generally similar. Another approach to addressing item bias was used by Sandoval and Whelan (1980), who tested 100 college students from different ethnic backgrounds to assess the face validity and item difficulty of the WISC items. No differences in cross-ethnic groups were found, suggesting that item difficulty was generally equally rated by the different groups of college students.

An alternative approach to cross-group comparison would be to control general IQ rather than ethnic group identity alone, and then to compare different ethnic groups. Taylor and Richards (1991) controlled overall IQ and then examined the intellectual patterns of African-American, Hispanic, and White children. In general, White children had the highest subtest test scores with African-American children scoring higher than Hispanics on the verbal subtests, and Hispanics scoring higher than African Americans on the performance subtests. When FSIQ was covaried, the Hispanic group was highest on Picture completion, Block Design, and Object Assembly, the White group on Information and Similarities, and the African-American group on Vocabulary. This study underscores some very important issues. Children of different ethnic groups appear to have different general intellectual patterns before and after FSIQ is controlled, and the subtest patterns differ considerably. Thus, one could conclude that different ethnic groups seem to vary on both the overall IQ scores as well as in the score patterns (e.g., White children attain higher FSIQ scores than African Americans and Hispanics, African Americans are better at Verbal tasks, and Hispanics are better at Performance tasks. However, when FSIQ is covaried, the patterns are much more subtle, suggesting large within-group differences, which may often be masked by the perceived large between-group differences. These findings, however, may be tempered by the overall intellectual status of the student. The Taylor and Richards study compared ethnic groups composed of normal children. Will the same patterns exist with special needs children? Barona (1989) reported that for children with mental retardation no significant differences were found across the major WISC-R factors between African-American, Mexican, and White children. However, for learning-disabled children, White children score higher on the verbal scale while African Americans score lower on the perceptual organization factor.

Finally, it is worth noting that such differences appear to be stable over time. Elliott and Boeve (1987) reported that handicapped Anglo, African American, and Mexican-American children did not have large "clinically significant" changes in their score patterns over a 3-year period. It would be easy yet erroneous to simply say that either no differences exist between ethnic groups or that easy-to-understand differences are evident. The complex truth appears to lie somewhere in
between. Furthermore, ethnic differences, though apparently reliable over time, appear to be modulated by intellectual status, and most likely by other variables not frequently measured nor considered (i.e., socioeconomic status, the educational attainment of parents, acculturation level, and so forth).

**SUMMARY**

Ethnic and race differences may best be explained not by speculation, popular opinion, or emotionally based arguments but through more of an anthropological or cultural understanding of the issues in question. The WISC-III is a measure of ‘g,’ which many psychologists accept as a universal description of intelligence. However, the content of intelligence tests and the normative data reflecting performance may not be universal but vary across as well as within groups. Foster and Cone (1995) discussed the importance of a cohesive hypothetical construct in assessment as well having a clear understanding of the purpose for which the test reflecting this construct is intended. Even though the WISC-R and the WISC-III may be used to assess the intelligence of American children, ethnic differences are still observed. The task will then be to determine exactly what those differences are, how are they manifested when important variables are controlled, and finally, what do these differences suggest. Even then, we must always appreciate the diversity of the children and the society we seek to understand and serve. In this context, the Wechsler scales can serve as useful measures of childrens’ intelligence.

**REFERENCES**


11. WISC AND ETHNICITY


