
WISC-III CLINICAL USE AND INTERPRETATION

SCIENTIST-PRACTITIONER PERSPECTIVES

EDITED BY

AURELIO PRIFITERA
The Psychological Corporation
San Antonio, Texas

DONALD H. SAKLOFSKE
Department of Educational Psychology
University of Saskatchewan
Saskatoon, Saskatchewan, Canada



ACADEMIC PRESS

San Diego London Boston New York Sydney Tokyo Toronto

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 socioeconomic 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

TABLE 11.1 Ethnicity and Race According to the 1990 U.S. Census Data: Origin, Total Numbers, and Subgroups

Spanish/Hispanic/Latin Background or Origin	
Origin = Latin America or Spain; Total = 22,354,059	
Cuban	
Mexican/Mexican-American/Chicano	
Puerto Rican	
Hispanic Latin American (e.g., Panamanian, Peruvian, Venezuelan, Ecuadorian, Guatemalan, etc.)	
Spaniard	
African American/Black/Negro	
Origin = African or Caribbean; Total = 29,986,060	
Asian or Pacific Islander	
Origin = Far East, Southeast Asia, Indian Subcontinent, or Pacific Islands; Total = 7,273,662	
Asian Indian	
Chinese	
Japanese	
Korean	
Vietnamese	
Filipino	
Hawaiian	
Indian (American) or Alaska Native	
Origin = North America; Total = 1,959,234	
Aleut	
American Indian	
Eskimo	
White	
Origin = Europe, North America, Middle East; Total = 199,686,070	

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 $1/2$ to 1 standard deviation, and Blacks about 1 standard deviation below Whites. Intelligence test findings for North American Aboriginial 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*, Rushon 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 and 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 is 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 Priftera 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, & Constantino, 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-

ple, 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 (Kaufman & 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, Abbatiello, & 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 (Wechsler, 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 Figerola (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, Rodriguez, 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, Krichew, 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 Langan (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, & Walters, 1986). Saklofske and Schwan-Kowalchuk (1992) have discussed a number of factors that may impact on the test performance of children, ranging from race of 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 IQs 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

- American Psychological Association. (1993). Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations. *American Psychologist*, 48, 45-48.
- Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Psychology*, 17, 143-150.
- Ardila, A., Rosselli, M., & Puente, A. E. (1994). *Neuropsychological Evaluation of the Spanish Speaker*. New York: Plenum.
- Barona, A. (1992). Effects of test administration procedures and acculturation level on achievement scores. *Journal of Psychoeducational Assessment*, 10, 124-132.
- Barona, A. (1989). Differential effects of WISC-R factors on special education eligibility for three ethnic groups. *Journal of Psychoeducational Assessment*, 7, 31-38.
- Barona, A., Santos de Barona, M., & Faykus, S. P. (1993). The simultaneous effects of sociocultural variables and WISC-R factors on MR, LD, and nonplacement of ethnic minorities in special education. *Education and Training in Mental Retardation*, 28, 66-74.
- Beaucourt, H., & Lopez, S. R. (1993). The study of culture ethnicity and race in American psychology. *American Psychologist*, 48, 629-637.
- Berry, J., Poortinga, Y., Segall, M., & Dasen, P. (1992). *Cross-cultural Psychology: Research and Applications*. New York: Cambridge University Press.
- Berry, J. (1984). Towards a universal psychology of cognitive competence. *International Journal of Psychology*, 19, 335-361.
- Bontempo, R. (1993). Translation fidelity of psychological scales: An item response theory analysis of an individualism-collectivism scale. *Journal of Cross-Cultural Psychology*, 24, 149-166.
- Brace, C. L. (1997). Intelligence. *Neural History*, 6, 12.
- Brandt, E. (1984). The cognitive functioning of American Indian children: A critique of McShane and Pas. *School Psychology Bulletin*, 13, 74-81.
- Bristol, R. W. (1989). Increasing awareness of class, ethnicity, culture and race by expanding on student's own experiences. In I. S. Cohen (Ed.) *The G. Stanley Hall Lecture Series* (Vol. 8). Washington, D.C.: American Psychological Association.
- Christjohn, R., & Lanigan, C. (1986). Research on Indian intelligence: Review and prospects. In selected papers from the First Mokakiti Conference (1984) (pp. 50-57). Mokakiti Indian Educational Research Association, University of British Columbia.
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 1-7.
- Common, R., & Frost, L. (1988). The implication of the mismeasurement of Indian students through the standardized intelligence tests. *Canadian Journal of Native Education*, 15, 18-30.
- de Jesus, N., Perrin, G., & Blackwood, H. D. (1996). Conducting assessments of non-English speaking persons. *El Boletín*, 2, 3-6.
- Elliott, S. N., & Boeve, K. (1987). Stability of WISC-R IQs: An investigation of ethnic differences over time. *Educational and Psychological Measurement*, 47, 461-465.
- Ellis, B. (1990). Assessing intelligence cross-nationally: A case for differential functioning detection. *Intelligence*, 14, 61-78.
- Figuerola, R. (1983). Test bias and Hispanic children. *The Journal of Special Education*, 17, 431-440.
- Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment*, 7, 248-260.
- Friel-Patti, S. (1990). Otitis media with effusion and the development of language: A review of the evidence. *Topics in Language Disorders*, 11(1), 11-23.
- Garcia, J. G., & Marotta, S. (1997). In Jorge G. Garcia & Maria Cecilia Zea (Eds.), *Psychological interventions and research with Latino populations* (pp. 1-14). Needham Heights, MA: Allyn and Bacon.
- Geisinger, K. F. (1992). Fairness and selected psychometric issues in the psychological testing of Hispanics. In K. F. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 17-42). Washington DC: American Psychological Association.
- Greenough-Olson, S. C. (1993). *The influence of language on non-verbal intelligence test performance of Northern Saskatchewan Aboriginal children*. Unpublished Master's thesis, University of Saskatchewan, Canada.
- Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for white and black children from the standardization sample. *Journal of Educational Psychology*, 73, 227-231.
- Hanley, J. H., and Barclay, A. G. (1979). Sensitivity of the WISC and WISC-R to subject and examiner variables. *The Journal of Black Psychology*, 5, 79-84.
- Helm, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- Hernstein, R. J., & Murray, C. (1994). *The bell curve*. New York: The Free Press.
- Hickey, T. (1972). Bilingualism and the measurement of intelligence and verbal learning ability. *Exceptional Children*, 39, 24-28.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jones, J. M. (1994). *Our similarities are different: Toward a psychology of affirmative diversity*. San Francisco, CA: Jossey Bass.
- Jones, J. M. (1991). *Racism: A cultural analysis of the problem*. Berkeley, CA: Cobb & Henry.
- Kaufman, A. S. (1993). King WISC the Third assumes the throne. *Journal of School Psychology*, 31, 345-54.
- Kaufman, A. S., & Doppelt, J. E. (1976). Analysis of standardization data in terms of stratification variables. *Child Development*, 47, 165-171.

- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Koh, T., Abbateio, A., & McLoughlin, C. S. (1984). Cultural bias in WISC subtest items: A response to Judge Brady's suggestion in relation to the PASE case. *School Psychology Review*, 13, 89-94.
- La Fromboise, T., Coleman, H. L., & Gertton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin*, 114, 395-412.
- Lane, C. (1994). The tainted sources of *The Bell Curve*. *The New York Review*, 26, 14-19.
- Laosa, L. (1991). The cultural context of construct validity and the ethics of generalizability. *Early Childhood Research Quarterly*, 6, 313-321.
- Laosa, L. (1984). Ethnic, socioeconomic, and home language influences upon early performance on measures of ability. *Journal of Educational Psychology*.
- Laosa, L. (1975). Bilingualism in three United States Hispanic groups: Contextual use of language and children and adults in their families. *Journal of Educational Psychology*, 67, 617-627.
- Lewontin, R., Rose, S., & Kamin, L. (1984). *Not in our genes*. New York: Penguin Books.
- Malgady, R. D., Rogler, L. H., & Constantino, G. (1987). Ethnoculture and linguistic bias in mental health evaluation of Hispanics. *American Psychologist*, 41, 228-234.
- Manuel-Dupont, S., Ardila, A. A., Rosselli, M., & Puente, A. E. (1992). Bilingualism. In Puente, A. E., & McCaffrey, A. J. (Eds.) *Handbook of neuropsychological assessment*. New York: Plenum.
- Melnick, M. (1997). Methodological errors in the prediction of ability. *American Psychologist*, 52, 74-75.
- McClelland, D. (1973). Testing for competence rather than for intelligence. *American Psychologist*, 28, 1-14.
- McCullough, C., Walker, J., & Dlessner, R. (1985). The use of the Wechsler scales in the assessment of Native Americans in the Columbia River Basin. *Psychology in Schools*, 22, 23-28.
- McShane, D. (1983). Explaining achievement patterns of American Indian children: A transcultural and developmental model. *Peabody Journal of Education*, 61, 1, 34-48.
- McShane, D., & Plas, J. (1984). The cognitive functioning of American Indian children: Moving from the WISC to the WISC-R. *School Psychology Review*, 13, 11, 61-73.
- Moore, E. G. J. (1987). Ethnic, social milieu and black children's intelligence test measurement. *Journal of Negro Education*, 56, 44-52.
- More, J. (1987). *Native American learning styles: A review for research*. Paper presented at "Meeting Their Needs," Winipeg, Manitoba.
- Mulcahy, R., & Marfo, K. (1987). Assessment of cognitive ability and instructional programming with Native Canadian children. In L. Stews & S. McCam (Eds.), *Contemporary education issues: The Canadian mosaic* (pp. 157-178). Toronto: Copp Clark Pinnam Ltd.
- Musler, H., Mulcahy, R., Wilgosh, L., & Waters, B., & Mancini, G. (1986). An analysis of WISC-R item responses with Canadian children. *Alberta Journal of Education Research*, 32, 12-36.
- Munford, P. R., & Munoz, A. (1980). A comparison of the WISC and WISC-R on Hispanic children. *Journal of Clinical Psychology*, 36, 452-458.
- Murdoch, J. (1988). Cree cognition in natural and educational context. In J. Berry, S. Irvine, & E. Hunt (Eds.), *Indigenous cognition: Functioning in cultural context* (pp. 231-255). Boston: Martinus Nijhoff Publishers.
- Naglieri, J. A. (1986). WISC-R and K-ABC comparison for matched samples of black and white children. *Journal of School Psychology*, 24, 81-88.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1995). *Intelligence: Knowns and unknowns*. Washington, DC: American Psychological Association.
- Nobles, W. W. (1995). *Psychological research and black self-concept: A Cultural Review*. New York: University Press.
- Olmedo, E. (1981). Testing linguistic minorities. *American Psychologist*, 36, 1078-1085.
- Padilla, A. M. (1995). *Hispanic Psychology: Critical issues in theory and research*. Thousand Oaks, CA: Sage Publications.
- Padilla, A. M. (1988). Early psychological assessments of Mexican-American children. *Journal of the History of the Behavioral Sciences*, 24, 111-117.
- Pike, R. (1966). *Language in relation to a unified theory of the structure of human behavior*. The Hague: Mouton.
- Prewitt-Diaz, J., Rodriguez, M., & Rivera, D. (1986). The predictive study of the Spanish translation of the WISC-R (Elwyn-R) with Puerto Rican students in Puerto Rico and the United States. *Educational and Psychological Measurement*, 46, 401-407.
- Rechsy, D. (1978). WISC-R factor structures among anglos, blacks, Chicanos, and Native American Papagos. *Journal of Consulting and Clinical Psychology*, 46, 417-422.
- Rechsy, D. J., & Rechsy, J. E. (1979). Brief reports on the WISC-R. *Journal of School Psychology*, 17, 355-361.
- Reynolds, C. R. (1995). Test bias and the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 545-576). New York: Plenum.
- Reynolds, C. R., & Wilson, V. L. (1983). *Methodological and statistical advances in the study of individual differences*. New York: Plenum.
- Rodriguez, J. M., de Torres, I. R., Herrans, L., & Aponte, M. R. (1994). Comprehension of instructions and other critical issues in the test standardization process. *Revista Interamericana de Psicología*, 28, 179-190.
- Rowe, D. C. (1994). No more than skin deep. *American Psychologist*, 49, 215-216.
- Rushon, J. P. (1997). Race, IQ, and APA report on *The Bell Curve*. *American Psychologist*, 52, 69-70.
- Saklofske, D. H., & Janzen, H. Z. (1990). School-based assessment research in Canada. *McGill Journal of Education*, 25, 1, 5-23.
- Saklofske, D. H., & Schwane-Kowalechuk, V. (1992). Influences on testing and test results. In M. Zeidner & R. Most (Eds.), *Psychological testing: An inside view* (pp. 89-119). Palo Alto, CA: Consulting Psychologist Press.
- Sanchez, G. I. (1932). Group differences in Spanish-speaking children: A critical review. *Journal of Applied Psychology*, 16, 549-558.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology*, 47, 919-927.
- Sandoval, J., & Whelan, M. P. (1980). Accuracy of judgements of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology*, 48, 249-253.
- Sandoval, J., Zimmerman, I. L., & Woo-Sam, J. (1983). Cultural differences on WISC-R verbal items. *Journal of School Psychology*, 21, 49-55.
- Saskatchewan Indian Institute of Technology. (1990). *Aboriginal literary plan*. Saskatoon, Saskatchewan.
- Sattler, J. (1988). *Assessment of children* (3rd ed.). San Diego: Jerome Sattler Publications.
- Scaidwell, W., Frane, J., & Cookson, D. (1985). Individual assessment of Chippewa, Muncey, and Oneida children using the WISC-R. *Canadian Journal of School Psychology*, 1 (1), 15-21.
- Scarr, S., & Weinberg, R. A. (1976). I. Q. test performance of black children adopted by white families. *American Psychologist*, 31, 726-739.
- Seyfort, B., Spreen, D., & Lahmer, V. (1980). A critical look at the WISC-R with Native Indian children. *Alberta Journal of Educational Research*, 26, 1, 14-21.
- Slate, J. R., & Jones, C. H. (1995). Preliminary evidence of the validity of the WISC-III for African American students undergoing special education evaluation. *Educational and Psychological Measurement*, 55, 1039-1046.
- St. John, J., Kirchev, A., & Bauman, E. (1976). Northwestern Ontario Indian children and the WISC. *Psychology in the Schools*, 13, 407-411.
- Tanner-Halverson, P., Burden, T., & Sabers, D. (1993). WISC-III normative data for Tohono O'Odham Native American children. *Journal of Psychoeducational Assessment*. WISC-III Monograph, 125-133.
- Taylor, R., & Richards, S. (1991). Patterns of intellectual differences of black, Hispanic, and white children. *Psychology in the Schools*, 28, 5-9.

- Terman, L. M. (1916). *The measurement of intelligence*. Cambridge, MA: Riverside.
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Cambridge, MA: Houghton Mifflin.
- Terrell, F., & Terrell, S. L. (1983). The relationship between the race of the examiner, cultural mistrust, and the intelligence test performance of black children. *Psychology in the Schools*, 20, 367-369.
- U.S. Bureau of the Census. (1992). *Survey of Business Owners and Self-Employed Persons: Form MB-1*. Washington, D.C.: Department of Commerce.
- Wechsler, D., (1974). *Wechsler Intelligence Scale for Children—Revised Edition*. New York: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition: Manual*. San Antonio: The Psychological Corporation.
- Westermeyer, J. (1987). Cultural factors in clinical assessment. *Journal of Consulting and Clinical Psychology*, 55, 471-478.
- Wilgosh, L., Mulcahy, R., & Walters, B. (1986). Assessing intellectual performance of culturally different Inuit children with the WISC-R. *Canadian Journal of Behavioral Science*, 18, 3, 270-277.
- Williams, R. L. (1975). The BITCH-100: A culture-specific test. *Journal of Afro-American Issues*, 3, 103-116.
- Yee, A. H. (1997). Ending the controversy. *American Psychologist*, 52, 70-71.