

## CHAPTER 21

**PSYCHOLOGICAL ASSESSMENT  
OF ETHNIC MINORITIES**

Antonio E. Puente  
Miguel Perez Garcia

**INTRODUCTION**

Whereas the study of abnormal behavior through the use of scientifically based psychological instruments has a relatively lengthy and interesting history (see introductory chapter), the assessment of individuals who fall toward the edges of the "bell curve" poses unique social, political, and scientific challenges. Traditionally, approaches to the study of individuals who are considered outside the mainstream of whatever society they belong to have been politically and socially based. The understanding of the study of these individuals was grounded on the assumptions that it is morally correct to understand these individuals (Fowlers & Richardson, 1996). Their "abnormal" functioning may be more saliently expressed by understanding their affiliation to a culture that is not appreciative or reflective of the majority group which rules or guides the social context in which they live. However, we propose that while such a motivation would appear reasonable and politically correct, it is still insufficient scientifically.

First, such an approach presupposes that the role of psychology is partially if not largely oriented toward righting the wrongs of a society's ancestors and, hence, primarily a social enterprise. While a reasonable goal, that would appear to us as insufficient. Second, one might assume that understanding others who are, by design, difficult to

understand, is again a reasonable goal. While we believe that for individual cases and in clinical situations this is not only desirable but also ethically appropriate, again this paradigm is insufficient. A third goal, rarely addressed by workers in this field, is that we believe that the study of culture and psychopathology combined (especially from a cognitive or neuro-cognitive perspective) provides a much larger pool of data about the human condition than previously used paradigms.

An example of this approach is found in a study from the World Health Organization (1979). They reported that with regards to schizophrenia, in Nigeria 58 percent and in India 51 percent of hospitalized individuals experienced complete remissions after two years after treatment. In contrast, in Denmark only 6 percent remission had been reported. The question then becomes what aspects of Nigerian and Indian culture is present (that are not in Denmark) which allow for such a high rate of recovery. The Basic Behavioral Science Task Force of the National Advisory Mental Health Council (1996) reported that in Los Angeles, Mexican Americans indicated that schizophrenia was a transitory condition associated with nervousness whereas Anglo-Saxon counterparts believed that schizophrenia was a permanent and total deterioration of mental functions. The question becomes how culture is defined and how it helps focus the more global issue of human function and dysfunction.

tion. Thus, we believe that basic principles about the human condition are the best understood by studying the interrelations of culture and psychopathology through the use of scientifically and cognitively (possibly neuro-cognitively) based psychological instruments. Such an approach could potentially yield unique insights into individual differences and general theories of psychological function and dysfunction.

However, individual differences and diversity are often viewed as impediments to the development of general principles of behavior. Of significance is the lack of understanding and sensitivity for larger group differences. Few would question the basic ability of specific psychological tests (e.g., Wechsler Memory Scale) to assist in the discrimination or classification of specific diagnostic groups (i.e., those with, from those without memory dysfunction). In contrast, few would disagree that affiliation with specific demographic groups (non-diagnostic) would be of great value in diagnostic classification. Presumably, this assumption is based on the concept that psychopathology (or for that matter, skills, abilities, or any other behavioral variable) is relatively free from the contamination of these potential confounds (Westermeyer, 1987a). Thus, this perspective suggests that other (non-diagnostic) group membership, while possibly important in some capacity, would have little or no effect on nosological issues. Such a belief is deeply rooted in nonempirical foundations, and its beginnings lie in a number of historical trends, none well documented or acknowledged. Thus, speculation rather than definitive analysis is the source for the following observations.

Traditionally, few attempts have been made to understand the behavior of individuals in minority groups. Brislin (1988) and others have cogently argued that psychologists for too long have categorically shown poor understanding of behavior traits and patterns of individuals who do not belong to groups associated with mainstream society. This limited perspective of the nature of behavior was first addressed by Frank Beach (1950) in his now classic article, "The Snark was a Boojum." In more contemporary terms, Robert Guthrie's (1976) book, *Even the Rat was White*, cites clear evidence not only of restricted sampling but of limited understanding of many other species (in the case of Beach) or other racial and ethnic groups (in the case of Guthrie).

One direct outcome of this situation, shown in recent statistics, suggests that few individuals

appear interested in studying how understanding racial and ethnic group membership may contribute to understanding behavior. The article by the American Psychological Association's (APA) Committee for Human Resources, "The Changing Face of American Psychology" (Howard, Pion, Gottfredson, Flattau, Oskane, Pfafflin, Bray, & Burstein, 1986) underscores the paucity of minorities pursuing study and being associated with all areas of psychology. Of special concern is the limited number of minorities in graduate schools and in faculty positions. Hall (1997) has suggested that "cultural malpractice" exists across all aspects of psychological pedagogy, research, and clinical activities. Bernal and Castro (1994) indicated that only 12 percent of all clinical programs require courses involving cultural issues but 89 percent of the programs indicated that they "integrated" such issues into their program. Interestingly, they reported that approximately half of all clinical programs did not have an ethnic minority on the staff. To add insult to injury, less than 10 percent of clinical students are of color. These disturbing trends persist a decade after the historical report in 1978 by the President's Commission on Mental Health.

The lack of understanding combined with the lack of resources to solve the problem will clearly lead to further complications of an already complex issue. Nevertheless, the common denominator is limited understanding. This limited understanding of minority populations has resulted in overrepresentation of minority groups in several distinct psychopathology groups. Maheady, Towne, Algozzine, Mercer, and Ysseldyke (1983) and others have observed that members of minority or underrepresented groups tend to be overrepresented in special education programs, especially programs for the mildly handicapped. However, it is unclear that less "biased" tests will produce less overrepresentation. Thus, we believe that basic principles about the human condition is best understood by studying

The overrepresentation of minority groups in handicapped conditions has, in turn, resulted in negative stereotypes. In 1991 the Department of Education reported that African Americans comprise 21 percent of total enrollment but an astonishing 42 percent of individuals labeled educable mentally retardate, 38 percent of those in educable mentally retardate, and 22 percent of those considered learning-disabled. Hispanics comprise 13 per-

cent of the total enrollment, but 10 percent of educable mentally retarded, 22 percent trainable mentally retarded, and 12 percent learning disabled. In contrast, for Asian students total enrollment is reflective of enrollment of special programs. Such stereotypes in the short term encourage the assignment of individuals to incorrect diagnostic groups (e.g., learning disabled). In the long term this stereotypical and grossly incorrect database may eventually serve as a foundation for potentially incorrect theories and research programs on racial and ethnic differences (e.g., Jensen, 1980). While all valid programs of inquiry should exist (Kuhn, 1970), constraints on the scientific process fueled by emotional and unempirical variables have little value for the discipline, for the science, for society, and most of all, for members of ethnic-minority groups. But as late as the end of the 20th century, we still are surprised and disappointed to read that "distinguished" historians of American culture (e.g., Graham, 1996) continue to misunderstand the very essence of the issues at heart.

The purpose of this chapter will be to avoid such an orientation by focusing as much as possible on the data that are available. Initially, this contribution will focus on providing both historical and clinical background of testing of ethnic-minority group members. Standard clinical and psychometric practices involving individuals of minority groups will be presented and critiqued. Suggestions for theoretical shifts as well as practical clinical and psychometric approaches will be outlined, with cognizance of the potential pitfalls, perceived or real, that presently exist.

This chapter is primarily intended for North American audiences. Numerous limitations in the available data set, whether clinical or otherwise, would make a more geographically ambitious approach impossible. In fact, it could be argued that the most fertile research database is found in the states. Nevertheless, the approach (though not necessarily the data) should be considered a model for workers in other cultures, groups, or locations (e.g., Native Indians in mainstream Brazilian culture) in order to address the issues of psychological assessment of ethnic-minority group members.

An initial step in understanding members of minority groups is to define such groups. According to accepted practice, individuals are different from larger groups if they are not members of that group. Group composition can be determined by social, legal, biological, statistical, and behavioral

variables. Possibly the easiest and most socially acceptable variable is biological, such as color of skin. However, other variables may also play a role. Statistical methods define group memberships by numerical scores obtained, while social and legal approaches may use societal tradition to define membership. Behavioral and psychological variables represent the most robust method as they should be free of bias due to the use of empirical methods and the criterion in question, the function of the person. After all, the color of an individual's skin is much less critical than their thinking patterns when it comes to understanding such issues as capacity to learn.

Standard practices have used overt and obvious variables to classify members into minority groups. For example, if an individual is not white (Caucasian) in North America he or she must belong to a minority group. One need look no further than the disciplines of animal behavior and neuropsychology to realize that gross morphological signs are often not well correlated with clear behavioral patterns. For this chapter, Brislin's (1988) classification system for human diversity is adopted. Contrary to popular belief, only three races exist. These include Caucasian (e.g., white), black, and Indian. The Indian race can be subdivided into Native American (e.g., Cherokee, Incas, etc.) and Asian (e.g., Japanese, Chinese, etc.). Ethnicity is another variable that can be used to differentiate mainstream from minority groups. Here, ethnicity is defined as a collective identity (e.g., Jew, Italian, etc.). Next, group composition can be determined by culture (e.g., southern, urban, etc.). This variable implies that groups can be defined according to social and personal identification. While less understood and accepted, other variables could also assist in determining group membership. These include, but should not be limited to, gender, sex, physical status (e.g., disability), social class, and religion. In 1990 the United States Bureau of the Census has more or less compressed these distinctions avoiding the differences between race, culture, and ethnicity. In a bold step, they proposed five different groups; Spanish/Hispanic/Latin Background or Origin, African American/Black/Negro, Asian (American) or Alaska Native, and White. The Spanish group contains five different groups whereas the Asian has seven separate subcategories and the Indians have three.

In the area of psychological assessment, race has been the most widely studied of the previous variables. Sex and, to a lesser degree, ethnicity have

**Table 21.1.** Ethnicity and Race According to the 1990 U.S. Census Data: Origin, Total Number, and Subgroups

---

Spanish/Hispanic/Latin Background or Origin
Origin = Latin America or Spain; Total = 22,354,059
Cuba
Mexican/Mexican-American/Chicano
Puerto Rican
Hispanic Latin America (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

---

been considered as potential though not highly salient variables. However, culture, physical status, social class, and religion have rarely been considered important in understanding human behavior. Whether this neglect is due to collective wisdom or ignorance is not known (nor is it the focus of this chapter).

Regardless of the variable used, ethnic-minority group membership will be defined as indicated previously by groups who are both politically powerless and sparsely represented in scientific inquiry. However, what may be a minority group in terms of ethnicity in 1990 may not be by the year 2000. Census figures suggest, for example, that people of color (including African Americans, Asian Americans, Hispanics, and Native Americans) who now constitute less than 20 percent of the U.S. populations will soon constitute approximately 50 percent of the American population (Basic Behavioral Science Task Force of the National Advisory Mental Health Council, 1996).

A necessary outcome of appropriately defining group membership is the implication that a minority member will engage in behavior that is different

from the mainstream norm but not necessarily abnormal. Thus, clearer understanding of human behavior is the goal. Such an understanding is not only academically useful but also contains treatment implications. The importance of ethnic-minority group membership for psychological treatment has been outlined by Sue and Zane (1987), while Lawson (1987) has reported its implications for psychopharmacological intervention. Caution should be inserted here. Careful between-group comparison often implies limited concern for within-group analysis. Using the Hispanic population in the United States as an example, the behavioral patterns of Cubans, Mexicans, and Puerto Ricans may actually differ more from each other than the entire group of Hispanics differs from Caucasians. In an ongoing translation and standardization of the Wechsler Intelligence Scale for children, Hispanics have been further subdivided into Central Americans, Cubans, Mexicans, Puerto Ricans, and South Americans. Thus, within-minority group analysis will eventually become as important as minority versus majority group comparisons.

### HISTORICAL PRECEDENTS

Galton's "Inquiries into Human Faculty and Its Development" written in 1883 is most often considered the beginnings of psychological assessment (Boring, 1950). In order to evaluate human disabilities (and not sins, as had commonly been the case prior to Galton), this British pioneer developed the "mental test." While the test intended to measure such variables as color discrimination and auditory reaction time, the purpose of establishing the Anthropometric Laboratory at the International Health Exhibition in London was to determine the range of human abilities. Together with the founding of the journal *Biometrika* and the Eugenics Laboratory, Galton attempted to develop the concept of racial improvement (Schultz & Schultz, 1996).

The discrimination of acceptable and unacceptable human characteristics has, unfortunately, found its way into present-day mental testing, possibly by way of James McKeen Cattell. After obtaining his Ph.D. from Wundt in Leipzig, Germany, Cattell came into contact with Galton (Boring, 1950), who in turn had enormous influence both directly (e.g., with numerous students) and indirectly (e.g., as editor of *Science*) on the study of mental ability in the United States. However, it was not until the appearance of Henry H. Goddard at Vineland Training School in New Jersey, and later Lewis Terman at Stanford University that a research program of psychological abilities became part of mainstream psychology.

Using "the evidence of mental tests," Terman (1916) indicated that "the average intelligence of women and girls is as high as that of men and boys" (p. 68). Nevertheless, he concluded in his book, *The Measurement of Intelligence*, that the "dullness" seen in "Indians, Mexicans, and Negroes raises the question of racial differences in mental tasks." Terman suggested, "Children of the group should be segregated in special classes and given instruction which is concrete and practical. They cannot master abstraction, but they can often be made efficient workers, able to look out for themselves" (p. 92). He continued, "There is no possibility at present of convincing society that they should not be allowed to reproduce, although from a eugenics point of view they constitute a grave problem because of their unusually prolific breeding" (p. 92).

Such an orientation is observed in Goddard's work and later in Robert Yerkes's groundbreaking

work with the Army Alpha and Beta tests during World War I. These tests were meant to classify A (intelligent) and D and E (feeble-minded) individuals with a mean mental age of 13.08. (This score may have prompted Goddard to term any adult with less than 13 years of mental age as "moron.") However, both immigrants and nonwhites tended to score lower, prompting Yerkes (1923) to write in *Atlantic Monthly* about noninherited racial differences. This conclusion readily supported the racist opinion of Madison Grant who considered Nordics superior to other races. Based on these observations, Yerkes and others encouraged strict immigration laws especially for "the negro." To curtail the reproduction of those already in the United States, several American followers of Galton (namely John H. Noyes and Victoria Woodhull) established a center for American eugenics in Cold Spring Harbor with financial support from the Carnegie Institution (Leahy, 1997). One of the greatest proponents of eugenics, Henry Goddard, published his famous book *The Kallikak Family: A Study in the Heredity of Feeble-mindedness* (1912). This book, probably more than any published work of the time, was used for the control of reproduction by ethnic minorities.

Reflecting the influence of this and similar works, sterilization and vasectomy became common phenomena. According to Leahy, one of the greatest landmark decisions on the issue was that of a mental patient, Carrie Buck. After giving birth to a retarded child out of wedlock, the "feeble-minded" Buck was involuntarily sterilized. She, in turn, sued the state of Virginia but lost in a split decision at the Supreme Court level. It seems as though unempirical (and presently considered unethical) approaches to the measurement of abilities are never easily resolved scientifically. Earlier in this century this issue was far from being resolved academically. Approximately 50 years later Stephen Jay Gould (1981) continues to argue that such strong conclusions have indeed been based on weak data.

Unfortunately, this approach to the understanding of minority behavior, at best weak and spurious, was the foundation for the Jewish genocide by the Nazis. However, not until 1954 did the judicial branch in this country make strides to erase this previously accepted and now embarrassing "scientific" orientation. The Brown case in 1954 allowed for desegregating of races in the school system. However, cases specifically referring to minorities and testing did not surface until the 1970s

(Reschly, 1984). Generally, the plaintiff in these cases represented the three major minority groups of the time—African American, Hispanic, and Native American—who had been poorly and unethically classified as retarded.

While most of the cases were won or favorably settled out of court, it was not until the legislative aspect of the litigation-legislation cycle occurred (Bersoff, 1981) that reform began to be developed and later implemented. According to Reschly (1984), the federal Education for All Handicapped Children Act of 1975 "was the most important and most widely applicable legislative act." This act opened the road for later litigation meant to define more succinctly the spirit of this law.

Perhaps linked to these legal efforts, psychologists have become increasingly aware of the need to document human abilities more carefully. In his introduction to the special issue, "Cultural Factors in Understanding and Assessing Psychopathology" (*Journal of Consulting and Clinical Psychology*), James Butcher (1987) stated that "the application of psychological procedure and methods with patients from different cultural backgrounds raises numerous methodological issues. 'Issues such as psychological equivalence, test reliability and validity, and test utility were some of the factors that Butcher considered critical. Five years later, Bethacourt and Lopez (1993) still believe that the study of ethnic minorities still hold 'at best second place' in mainstream psychology. This has occurred despite critical reports suggesting that such an approach would be detrimental not only to ethnic-minorities but to psychology at large (McGovern, Furomoto, Halpern, Kimble, & McKeachie, 1991).

In this decade undoubtedly the most significant and most detrimental work on this issue comes from Herrnstein and Murray (1994) in their highly controversial book, *The Bell Curve*. Herrnstein and Murray aggressively pursue the traditional concepts that ethnic minorities do not score well on standardized tests, including tests of achievement and intelligence, because of genetic and biological limitations. In many respects these authors provide a modern-day version of the ideas of Terman, Goddard, and others (e.g., Graham) linked to immigration laws, reproductive limitations, and intellectual and social segregation.

Sternberg (1997) has cogently addressed the importance of ultimate criterion, possibly life-long learning capacity, rather than test intelligence. Test intelligence may be a significant though, by design,

incomplete condition for understanding learning. Such intelligence appears, according to Sternberg, to be predictive of scholastic achievement. However, this type of achievement is only partially correlated with life-long success. Two other factors are suggested by him that are obviously not understood by Herrnstein and Murray; behavioral intelligence and intelligence. It is the concept of behavioral intelligence that we find particularly interesting. One could clearly argue that a migrant worker whose native language is Spanish would do particularly poorly on the SAT or, for that matter, the GRE. Thus, there would be little question that such a person would not gain admission to most selective colleges in the United States. Further, one could argue that such an individual would undeniably do poorly in a traditional university curriculum. But to argue that such a person is biological or genetically inferior seems downright stupid. For example, such a migrant worker has found a method to travel from rural Central America to, say the eastern United States, with little money, inadequate transportation, and limited understanding of the culture. They are able to find work, complete the task, live frugally, send money back to their homeland, and locate alternative employment within days of completion of the job at hand, often in another state. One could argue that a suburban-raised Anglo-Saxon who has played on the high school sports team and has dedicated his or her life to spectator sports and socially driven concerns could not under any circumstances go to Central America and replicate what their counterparts have accomplished in the United States. The question becomes, are we measuring "true" intelligence or some understanding of culture.

Of course, to the typical reader of this chapter, such an example seems rather extreme. Hence, we have chosen to provide another example, which should be closer to the experiences of most psychologists. In another and more recent article Sternberg and Williams (1997) suggest that the GREs, still the most widely used standardized measure of achievement for acceptance into graduate school, predicts little in terms of graduate school performance and maybe less than that in terms of career success. Additional and also sophisticated arguments against the limited arguments of *The Bell Curve* are also found in Gould (1996). Thus, the criterion for intelligence and achievement and, for that matter pathology, according to Sternberg (and suggested by Gould) as well as accepted here, is not test scoring (espe-

cially alone and out of context) but life-long ability to adapt to the demands of life. This chapter attempts to build on this newfound scientific interest in an effort to determine the needs, limitations, and directions associated with the psychological assessment of ethnic-minority populations in North America.

## THEORETICAL ISSUES

In a chapter of this type it would be essentially impossible to address all pertinent theoretical issues that apply to the psychological assessment of ethnic minorities. We have chosen to focus on three main issues, bias, acculturation, and culture believing that they are the three most critical issues involved in this area of study.

### Bias

Kenneth Eells pioneered the concept of bias in mental measurement, specifically the mental test. While his work focused only on whites, it did address the importance of difference—in this case, social class—in the assessment of mental function (Eells, 1951). Although the reasons for doing so are not entirely clear, some workers in psychometrics generalized his findings to other populations, namely African Americans. This incorrect generalization launched a wave of poorly developed and executed studies on bias in testing.

One of the most controversial figures in mental bias research is Jensen, of the University of California at Berkeley; his most controversial book is *Bias in Mental Testing* (1980). According to Jensen, mental testing has been criticized because of one or more of the following reasons:

1. cultural bias
2. specific test items
3. inability to define or measure intelligence
4. tests that measure too narrow a range of abilities
5. failure to measure innate capacity
6. IQ tests that measure only learned skills
7. IQs that are inconsistent
8. test scores that are contaminated by extraneous factors
9. misuses, abuses, and undesirable consequences of testing

According to Jensen (1980), these criticisms are largely unfounded and confused with other factors. As he wrote, "Anxiety about one's own status, or the importance of the traits measured by tests, or sympathy for the less fortunate, may prompt the acceptance of criticisms of tests without evidence" (p. 23). Unfortunately, such critiques tend to focus on IQ tests and are emotionally interpreted. They complicate the question and prevent adequate understanding of the valid issues.

Reynolds and Brown (1984) presented a set of reasons, which are applicable to bias for a wider range of tests. These include:

1. inappropriate content
2. inappropriate standardization samples
3. examiner and language bias
4. inequitable social consequences
5. measurement of different constructs
6. differential predictive validity

Regardless of the source of bias, the definition of bias must also be considered. Unfortunately, numerous definitions are available in the literature—some more heuristic and plausible than others. The following are two samples of the many available.

[Bias results from] differences in the extent to which the child being tested has had the opportunity to know and become familiar with the specific subject matter or specific process required by the test item. (Eells, 1951, p. 54)

Psychometric bias is a set of statistical attributes conjointly of a given test and two or more specified sub-populations (Jensen, 1980, p. 375)

Flaugher (1978) has suggested that test bias can mean more than simple knowledge or psychometric deficiencies. Indeed, bias could be represented in a wide variety of concerns including, but not limited to, both psychometric issues (mean differences, differential validity, item content, internal validity) and test usage (over-interpretation, selection model, and atmosphere). He concluded that in 1978 the research was promising, but the results were still disappointing. Twenty years later, the research and the results are both disappointing.

Among the more current research findings, an excellent example is Drasgow's (1972) article, "Biased Test Items and Differential Validity." In this review the author addresses differences between majority and minority groups in validity coefficients. The results of his study provide sup-

port for earlier findings suggesting that validity coefficients may not prove useful in examining test bias. He concludes: "Test scores *can* be used to predict criterion performance for minority group members. Nevertheless, it *may be inappropriate to compare test scores for minority group members with test scores for majority group members*" (p. 529, italics added). In a similar vein, Cole (1981) concluded in her article, "Bias in Testing," that "there is no large-scale, consistent bias against minority groups." Nevertheless, both "subtle aspects of the testing situation" and presumably more refined understanding still evade workers in the field. In contrast, Humphries (1986) has argued that even if items differ between groups, these items should not be labeled as biased if adequate measurement properties are taken into account.

Recently, the American Psychological Association released the results of a task-force study on these issues (Neisser, Boodoo, Bouchard, Boykin, Bordy, Ceci, Halpern, Loehlin, Perloff, Sternberg, & Urbina, 1996). Presumably the focus of this task force was to address authoritatively the issues brought out by *The Bell Curve*. Whereas the article addressed numerous critical issues, it fell far short when addressing ethnic-minority issues. The task force assumes, without question, that if tests are to be used as predictors of future performance, these tests do not seem to be biased against African Americans" (p. 93). However, less clear evidence of bias is presented against other ethnic-minority groups. Unfortunately, one is left with the sense that outside Asian Americans, other ethnic minorities score below their Anglo-Saxon counterparts. Further, they suggest that numerous factors, but not necessarily bias, ranging from economics to genetics, many be playing a role in these differences. Despite these opinions, may questions have yet to be formulated and, of course, answered. Until then, as Reynolds and Brown have concluded, the verdict on test bias is still not in (scientifically).

However, Susuki and Valencia (1997) have suggested that a significant drop in bias research exists. They believe in this precipitous drop is due to the fact that belief that bias does not exist. This erroneous belief is due to the following reasons:

1. Some minority groups have not been studied extensively and, in some cases, not at all.
2. Test bias in school-based tests has been done only with some but not a large variety of tests.

3. In actuality, there are mixed results. They report several studies by Valencia and colleagues which suggest that the K-ABC contain bias in predictive validity and content but not in construct validity or reliability.
4. Bias research has traditionally been done with nonpatient populations.

Malgady (1996) proposes an interesting twist to the theoretical foundations of bias research. He proposes that what is necessary is to reverse the null hypothesis. That is, we must assume that bias exists. If one commits an error in measurement, then it is better to assume that bias exists so that precautions are taken to protect the individual.

### Acculturation

If a minority group does poorly on a test, relative to a majority group, two interpretations may be used to account for the discrepancy. A rather emotional one is provided by Jensen (1980)—that the difference is accounted for by biological factors such as genetics. A less popular interpretation used by researchers studying integration of an immigrant group into a majority or mainstream culture is that of acculturation.

Assimilation into a larger, more mainstream culture allows an individual to understand and adjust to the cultural, social, and psychological requirements of that culture. Conversely, those who do not adapt are considered to exhibit greater degrees of psychopathology. An illustration of the lack of adaptation was reported by Hoffman, Dana, and Bolton (1985) who found that Sioux Native Americans with strong ties to tribal values and language were more likely to exhibit psychopathology as measured by the MMPI. These findings have also been replicated with other minority groups, including Hispanics (e.g., Montgomery & Oroz, 1984). Focusing on cognitive style and intelligence, Gonzales and Roll (1985) reported differences between Mexican Americans and whites on several test measures. However, no group difference was observed between Anglo-Americans and a subgroup of the original sample of Mexican Americans who had been shown to be acculturated to Anglo-American culture.

One method to determine whether acculturation has been achieved, and thus controlled, is to administer an acculturation scale. Marin, Sabogal, Marin, and Otero-Sabogal (1987) have developed



a 12-item scale, which measures acculturation in Hispanic populations. The validation criteria included generation, length of residence in the United States, age at arrival, ethnic self-identification, and an acculturation index. These findings have been extended to children (e.g., Franco, 1983) as well as to other cultural groups such as Asian Americans (Suinn, Rickard-Figueroa, Lew, & Vigil, 1987). Preliminary findings suggest that age (younger), sex (male), and length of exposure to the predominant culture (Bumam, Telles, Kamo, & Hough, 1987) as well as cultural awareness and ethnic loyalty (Padilla, 1985) are critical factors in the acculturation process. Another scale used for acculturation is the Acculturation Rating Scale for Mexican Americans (ARSMA and ARSMA-II) (Cuellar, Arnold & Maldonado, 1995; Cuellar, Harris & Jasso, 1980). While this is a promising scale, more research is necessary to generalize use to other Hispanic and ethnic-minority populations. Fradd and Hallman (1983) concluded that until an individual has been taught strategies to build bridges from a previous to a current domain of knowledge, the validity of test measures is questionable.

The process of acculturation must be understood, however, as a dynamic rather than static process. Acculturation does not imply reaching an imaginary threshold at which time one becomes clearly acculturated. Knight and Kagan (1977) reported that it took about three generations for Mexican-American children to develop modal responses on Anglo-Saxon children with regard to social motives.

Four stages have been postulated (Basic Behavioral Science Task Force of the National Advisory Mental Health Council, 1996). They are; assimilation (becoming part of the majority culture), acculturation (adapting to the majority culture), alternation or biculturalism (adequately engaging two cultures), and multiculturalism (holding on to a personal and non-majority identity while participating in a goal-directed activity of the majority culture).

Further, acculturation is not dichotomous, instead it is multifaceted (Phinney, 1996; Magana, de La Rocha, Amsel, Magana, Fernandez, & Rulnick, 1996). Triandis (1980) has suggested that culture could be physical (e.g., buildings, tools, etc.) or subjective (e.g., social norms, roles, beliefs, and values). The subjective could include family dynamics, religious beliefs, language limi-

tations, individualism, and so forth. Thus, one could conceivably be adapted to a culture physically, live and appear to be American (i.e., live in North Carolina, dress in Brooks Brothers clothing, etc.) but have specific behavioral patterns that would clearly identify the person as non-North Carolinian (i.e., native language would be Spanish, have extended family, practice Catholicism, and so forth).

Berry (1990) has proposed an interesting theory of acculturation. The process involves three levels:

1. Antecedents—internal, external, and traditional.
2. Processes—cultural change, acculturation, psychological acculturation,
3. Consequents—changed cultural and social system, changed psychological status of persons.

Such a comprehensive approach appreciates the multidimensional and phasic nature of acculturation. Bethancourt and Lopez (1993) have also suggested that lack of acculturation implies perceived stress, presumably because of the individual's inability to "fit" with the majority culture. The problem is in determining why that "stress" exists by defining the specific cultural values (e.g., lack of understanding of social norms) that produce the lack of acculturation and eventually the perceived stress. And even though not suggested by the authors, but certainly implied, is that perceived stress must be controlled in order to make certain that what is measured is acculturation and not stress secondary to limited acculturation.

While it is clear that acculturation is seen as the ability of immigrants (e.g., Hispanics and Asians) to adapt to a majority group culture (e.g., the United States), the possibility is considered that an analogous concept could be applied to nonimmigrant ethnic minorities already living in North America. If, as it is argued in the next section, intelligence is largely, if not completely, a cultural phenomenon, then acculturating to the majority culture is a prerequisite for the development of successful learning strategies and eventual intelligence. Thus, it is argued that African Americans living in the United States but not participating fully in the American culture may not be acculturated. Hence, the same issues would apply to this ethnic-minority group as it would to Hispanics and Asians.

## Culture and Ethnicity

According to recent position papers, neither bias nor acculturation may be the most salient variables that need to be addressed in understanding ethnic minorities. Bethancourt and Lopez (1993) suggested that most studies to date on the issues at hand have been at best descriptive of the differences between cultures. Whereas such approaches appear on the surface useful, they suggest that they are at best an initial step in the more important question of culture. To date, they argue, the field does still not understand the role of culture in behavior and cognition. For example, they believe that using race as a variable, either dependent or independent, is inappropriate. Zuckerman (1990) has reported that within-race differences appear larger than between-race differences with regard to biological variables. The hypothesis is made that similar assumptions can be generated with regard to behavioral and cognitive variables as well. Indeed, social class might be a more salient variable in grouping individuals than race. In other words, individuals from a high social class would more likely be different than persons from a low social class than a white and black from similar social classes. First, they suggest the bottom up approach which starts with the data generated from cross-cultural studies in order to then alter existing theories about human behavior. They suggest the work of Triandis and colleagues as a benchmark for this approach. Second, and a more novel approach, is what they term the top down approach. The essence of this approach is to determine how culture helps define the larger concept of human behavior and cognition. In other words, culture is seen, much like psychopathology in the earlier ideas of Neal Miller, as a unique way of being able to understand "normal" human function. Instead of examining how cultures are different, the focus would shift to determining what are the most salient variables in helping distinguish individuals. Using Hispanics as an example, one would ask not how Hispanics are different than say Anglo-Saxons and what variables contribute to those differences. In this ethnic-minority group the following variables might be pertinent; language, social orientation, family dynamics, and religion. Then, the question would become how, for example, does language affect a person's behavior to the point of excluding them from a majority group essentially making them pathological. A third-tier question then would be what should be pathologi-

cal. Essentially, the final question would be what is the criterion for labeling pathology.

Phinney (1996) has furthered this concept by proposing that ethnicity, a subset of culture, could be studied at three levels. First, she suggests cultural values, attitudes, and behaviors that help define a group. Second, group identification is partially based on a subjective sense of what it means to belong to the group. Finally, she suggests that group identification is based partially on the specific experiences associated with that group identification.

Finally, an alternative to doing cross-cultural investigations is to begin by understanding culture. According to Greenfield (1997), "ability assessments don't cross cultures". Specifically, she suggests that values and meanings, knowledge, models of knowing, and conventions of communication are not easily translatable across cultures and could be culture-specific. That is, the criterion of a particular meaning must be understood before it is "translated". She concludes that "tests are not universal instruments".

## Social Policy

Whether tests are biased or culturally free, whether an individual belongs to a minority or a majority group, whether different groups are biologically equal or unequal, group differences still exist. To deny the obvious would be inappropriate. Certain minority groups perform differently, more often than not worse, than majority groups on specific items, tasks, or tests. These differences drive social policy. Academic psychology would undoubtedly prefer to research these problems and discrepancies more thoroughly before allowing findings to affect the judicial and legislative process, because the data for any of these questions are at best inconclusive as well as emotional and at worst confusing.

However, policy must be and will be made in the absence of adequate data and in the presence of emotion (see Bersoff, 1981). This reality could explain why Cole (1981) concluded that test bias research is likely to have only a small impact on complex social policy issues. Regardless, there are issues that relate to the available data. In the first edition of this handbook, Reschly (1984) addresses the concept of fairness. According to him, two approaches have been adopted. Equal treatment implies no bias or documentation in selection pro-

cedures and that all candidates, regardless of demographic affiliation, are treated equal. An alternative to this approach is equal outcomes, which implies that selection should match population demographics. Regardless of the approach and the data, North American society has adopted in principle the concept of fairness. The question remaining is which method described by Reschly will be chosen and what, if any, implications will the current paucity of data and lack of scientific agreement have on social policy formation and implementation. Another and politically limited approach would be to assume that representation of the American population (or for that matter whatever criterion population was chosen) is a criterion of choice. Next, one could use the currently used measures described in this chapter within the context of subsamples. Specifically, if selection to a college is the goal, then a college would first choose to accept representation from all groups as desirable. Next, they would apply the traditional standards (e.g., standardized test scores) within the accepted or chosen subsamples (e.g., Caucasian, Hispanics, African Americans, etc.). Considering the predictive validity of these tests, it is hypothesized that within-group (or subsample) variance would be greater than between-group (i.e., high-scoring blacks and high-scoring whites) variance.

To assist policymakers, researchers need to place greater importance on studying issues of race, culture, ethnicity, and related variables. The findings must then be applied to broaden our limited understanding of differences in psychological test performance of minority group members. Hall (1997) suggests the following steps in attempting to reach these objectives; (1) Ensure that the psychology curriculum is culturally diverse, (2) recruit and retain diverse faculty, (3) recruit and retain diverse students, (4) monitor, for the sake of accountability, efficacy of initiatives, (4) encourage culturally diverse research and publications, (5) increase the number of editors and reviewers of diverse background, (6) ensure minimum cultural competency for psychology students, (7) understand state-of-the-art research on topics of diversity, and (8) increase diversity within membership and leadership of the American Psychological Association.

Of course, there is the issue of who is to pursue these questions, both in academic and research settings. In the seminal article, "The Changing Face of American Psychology" (Howard et al., 1986), the future for ethnic-minority group representation

is presented as quite dismal. While women have made significant strides, African Americans, Hispanics, Asian-Americans, and Native Americans continue to lose ground, in terms of representation in graduate school ranks (Hall, 1997). Similar trends exist in academic ranks, and presumably in clinical settings as well (Bernal & Castro, 1994). Programs within the American Psychological Association, including the Minority Fellowship Program and the Minority Neuroscience Fellowship Program, may aid talented minorities to pursue graduate training. Unfortunately, undergraduate majors in psychology mirror the same trend (Puente, 1993). Indeed, by the time minorities have chosen a college, they most likely have committed to a course of study. Simply put, despite the urgency of the questions raised, the future for a better understanding of psychological assessment of ethnic-minority group members looks bleaker than its past especially when issues such as "pipeline" of prospective students is considered. Whereas one would hope that the natural forces or evolution of psychology would "take care of the problem", social policy initiatives may have to jump-start what psychology has verbalized yet never realized.

## ASSESSMENT METHODS

This section of the chapter will focus on specific assessment methods, including interviews, standard measures, culturally sensitive methods, and behavioral assessment methods. As feasible, each section will cover a variety but, not an exhaustive set, of tests or assessment strategies including application (and/or translation), norms, limitations and cautions, and suggestions for use.

### Interview

The interview, whether structured or unstructured, remains the initial step of any psychological assessment and also the most commonly used method for obtaining information. The interview is a frequently used method for obtaining data in cross-cultural contexts. As Zubin (1965) and others have pointed out, however, the unstructured interview poses problems since it may yield unreliable data resulting from a host of uncontrolled factors.

Structured interviews may help in avoiding these pitfalls. Numerous interview methods,

**Table 21.2.** Several Structured Interviews Applicable to Ethnic-Minority Populations

INTERVIEW	REFERENCES
Brief Psychiatric Rating Scale	Overall & Gorham (1962)
Inpatient Multidimensional Rating	Lorr & Lett (1969)
Mental Status Schedule	Spitzer, Endicott, & Flenn (1967)
Present State Examination	Wing (1970)
Structural Clinical Interview (DSM III)	Spitzer & William (1983)
International Classification of Disease Organization (1992)	World Health
Interview (for ICD-10)	
Mini-Method Station Examination	Folstein, Folstein & McHugh (1975)

including several presented in this volume, seem generally well suited for use with minority populations, especially since they are often based on objective diagnostic criteria (e.g., Research Diagnostic Criteria). Several of these methods are found in Table 21.2.

Although most of these structured interviews have been well studied and validated, validity studies often use the judgment of the clinician as the criterion variable. Further, it is well accepted that cultural and ethnic variables—such as behavior patterns, nonverbal cues, translation equivalence, concept equivalence, gender differences, and general cultural beliefs—are often misunderstood by even the most sensitive clinician (Hall, 1997; Westermeyer, 1987a). Recent research has also revealed that expression of psychological symptoms is differentially affected by culture. Interviews and diagnostic conclusions are based on signs and symptoms which could be considered normal (versus abnormal) in specific cultures (e.g., belief in the devil, describing somatic abnormalities using metaphors, etc.). The end result is confounding symptoms with culture (Basic Behavioral Science Task Force of National Advisory Mental Health Council, 1996).

One way to avoid this complication is to use interview methods that either have been formally validated or are in current use with these populations. For example, the Present State Examination was an interview used for the international pilot study of psychopathology (World Health Organization, 1973). Another method is that of using a translator or someone knowledgeable about ethnic-minority groups. However, even this approach has limitations. It is not unusual for the translator to be a lay person with limited understanding of psychological principles as well as an individual with personal interest in the patient. Further, translators may not approximate a balanced bilingual, or

worse yet, not understand the culture in question. Distortion or misconception further impairs data gathering, especially with severely disorganized patients or individuals whose culture is very different from that of the diagnostician. Velazquez, Gonzales, Butcher, Castillo-Canez, Apocada, and Chavira (1997) suggested that an important and often over-looked first step in an evaluation is to allow the patient to choose the language to be used in the evaluation.

Several steps might be taken to attempt to control interview distortion. First, in order to bridge the language and cultural gap between patient and psychologist, rapport should be established prior to the interview. Greenfield (1997) has reported that ease in speaking to strangers, even though they are professionals, varies across cultures. For collective cultures (e.g., Asian and Hispanic), it is typical to limit discussions to only known individuals (Kim & Choi, 1994). In other words, you only relate one's problems with intimate or close friends or family friends. In contrast, in North America providing personal information to a stranger, but presumably a professional, increases the perception of objectivity and effectiveness.

Westermeyer (1987b) suggested that diagnostic interviews may take up to twice the usual time of a standard interview. Also, the clinician should make sure that ambiguous (whether real or imagined) questions or answers are clarified. Confrontation, the hallmark of some structured interview methods, should be avoided if possible since it may adversely affect client-clinician rapport.

By far the most important aspect of any diagnostic interview is to place the client in his or her own bio-psychosocial context and not the psychologist's context. Otherwise, a patient's behavior could be incorrectly interpreted as maladaptive (Adebimpe, 1981). To avoid erroneous conclusions, the psychologist must put special emphasis

on understanding the patient's culture, race, ethnicity, class, or social context that grants him or her membership in a minority group. Not only must that context be understood but it should be understood as it relates to the patient's relationship to majority culture (e.g., Mexican migrant worker employed as a field hand in Colorado). Finally, and possibly most important, the clinician must understand his or her own limitations in other sociocultural situations. To enhance his or her understanding of others, the psychologist must become aware of, and possibly experience, other cultures and ethnic behavior patterns and cognitions. Hall (1997) has suggested that all clinicians must be well versed in these issues, initially in completing a multicultural graduate course and subsequently in clinical training.

### Intellectual

Tests which attempt to measure the construct of intelligence are not only the most commonly used psychological tests (see related chapters in this volume) but also the most criticized (Neisser et al. 1996, Sternberg, 1997). The literature is replete with controversies about the efficacy of the construct of intelligence and its measurability (Gould, 1996; Helms, 1992), and strong and often emotional arguments have been levied against tests of intelligence by members of ethnic-minority groups (Herrnstein & Murray, 1994). Before these arguments are considered, the most commonly used tests of intelligence will be reviewed relative to their applicability to minority populations.

The application of intelligence tests to children of minority populations has yielded the most empirical data as well as the most controversy. Of the tests applicable to children, the Wechsler Intellectual Scale for Children (WISC) is one of the most popular psychometric tests of intelligence (Puente & Salazar, 1998). Despite that the WISC-III was published in 1991 (Wechsler, 1991), most data on this topic exists with the first two versions of this test. Excluding Asian children, the results, in general suggest up to one standard deviation difference between ethnic-minority groups and the criterion sample, Anglo-Saxon children. Using Hispanic children as an example, it appears that these differences are erased if the child is a third-generation American. Thus, the issue might be more that the WISC

might be measuring some type of acculturation process.

Nevertheless, conflicting and nonconclusive evidence is often found. For example, in one thorough review of the literature, the race of the examiner did not seem to affect the validity of intelligence scores in African-American children (Graziano, Varca, & Levy, 1982). Using the criteria outlined by Jensen (1980) for determining bias in testing, Sandoval (1979) concluded that the "WISC-R appears to be non-biased for minority group children." Other factors are presented by Sandoval to explain observed minority versus majority group scores. These findings are supported by Ross-Reynolds and Reschly (1983) in a study involving Anglo, African American, Hispanic, and Native-American Papago. While no bias in the WISC-R was found against African Americans and Hispanics, ceiling effects influenced the response pattern of the Papagos.

Language, however, may be confounded in bilingual children and thus needs to be clarified prior to the administration of the WISC-R. Sandoval (1979) examined the evidence of cultural bias for Anglo, Hispanic, and African-American children. Further, the Spanish version of the WISC-R does not have acceptable norms for each cultural or ethnic group and should be used with extreme caution. Concern is also cited by Dana (1984) who indicated that the WISC-R is biased for traditional Native-American children. He indicated that a pattern of spatial > sequential > conceptual > acquired knowledge exists across both ages and tribes. The difficulties associated with using the WISC in other ethnic-minority cultures is discussed in Puente and Salazar (1998).

Lampley and Rust (1986) examined the validity of the Kaufman Assessment Battery for Children and found that African Americans scored significantly lower on this test. These findings are supported by others (e.g., Sandoval & Mielle, 1979). Nevertheless, these conclusions are in direct contrast to those of Hickman and Reynolds (1986-1987) who reported that "blacks did not perform significantly better in the test form developed solely on their own item statistic."

It seems that regardless of the data, contrasting interpretations abound. An interesting and eloquent attack on these issues was leveled by George Jackson, chair, Association of Black Psychologists in 1975. A more balanced perspective on this issue is presented by Cole (1981) as well as Reynolds and Brown (1984) and Helms (1992). Additional

commentaries and rebuttals are found in the 1985 article by Jensen in *Behavioral and Brain Sciences*.

Little information is found for adult intelligence testing with the Wechsler Adult Intelligence Scale Revised (WAIS-R) and, due to its recent publication date, the WAIS-III (though the items appear to be much less culturally biased and the norms are more reflective of the American population). For example, in the first edition of this handbook, Lindenmann and Matarazzo (1984) indicated that the Army Alpha was developed for literates and the Army Beta for the non-English speaking. The implicit assumption is that non-English-speaking individuals were illiterate. Of course, if the dominant language becomes that of the client, then it is the psychologist who is illiterate.

Using both the WAIS and the WAIS-R, Whitworth and Gibbons (1986) reported that differences were found using both tests and that the most significant differences appeared to be the conversion of race to scale scores. Reynolds, Chastain, Kaufman, and McLean (1987) re-analyzed the data for the 1981 standardized sample of the WAIS-R and reported a 14 1/2 point difference between whites and African Americans on the Full Scale IQ. In attempting to resolve these discrepancies, Grubb (1987) examined the IQ differences in profoundly and severely mentally retarded individuals using Weschler's test. He reported no differences between whites and African Americans in this sample of subjects and concluded that lower IQ scores of African Americans were not biologically determined and, instead, were attributable to other factors.

Unfortunately, few data other than the results of the Weschler tests exist on measures of intellectual abilities. While one might expect that such tests as the Raven Progressive Matrices and the Beta would be less ethnically biased, the data provide little support for this (or contradictory) views. For example, using minority group offenders, Hilton-Smith and colleagues (1984) reported that these subjects actually scored lower on the Beta than on the WAIS-R.

Obviously, one of two things must be occurring. There is either incorrect measurement of intellectual function or some difference (not deficiency) is present. Before accepting the possibility of difference, measurement error must be eliminated or reduced to the lowest possible level. One possible way to address this is to use greater care and ingenuity in the construction of intellectual tests. Using

the WISC as an example. Care must be taken in the development of items that are not culturally biased, both across cultures and within subcultures (e.g., Cubans versus Mexicans). Second, greater care must be taken in the standardization process. A typical protocol might include two phases, a try-out phase that helps develop further an item pool and a standardization phase that would closely mimic the U.S. population.

Another possibility is to consider intelligence from a totally different perspective. For example, Sternberg (1996) has suggested that intelligence is really nothing more than success in life. Hence, tests such as the Learning Potential Assessment Device (Feuerstein et al., 1979), Cognitive Assessment System (Naglieri & Das, 1996), as well as standard tests from the neuropsychological literature provide a more unique way to address the possible underlying variable, problem solving, in intelligence. Indeed, it is expected that future tests of intelligence will have strong foundations in neuropsychological performance.

### Neuropsychological

It is often assumed that brain functions are not affected by non-neurological variables. A review of the table of contents of major neuropsychological texts of the 1980s and 1990s suggests that issues of culture, ethnicity, and race have not been addressed to date. Even more revealing are the reference sections of the books, which indicate that very few articles on these issues exist. A review of the existing journal literature also exposes the paucity of references surrounding neuropsychological assessment and the effects of culture, ethnicity, and race. In *Reliability and Validity in Neuropsychological Assessment*, Franzen (1996) presents an excellent overview of issues concerning most measures of neuropsychological ability. While different forms of validity are considered, no mention is made of the application of the tests to minority group members.

Most of the sparse data that do exist on this topic are found in the non-neuropsychological literature. For example, Lopez and Romero (1988) assessed intellectual functions in Spanish-speaking adults using both the WAIS and the Puerto Rican version of the WAIS. While the authors report that differences did exist, test equivalence is generally elusive and its application for these tests to a neuropsychological sample would be at best hap-

hazard. On a more theoretical orientation, Drasgow (1972) addressed test-item bias and differential validity by using a "profoundly" biased test. However, in this case (as with all others), no direct or indirect mention is made of neuropsychological tests.

Anecdotal and clinical evidence indicate that these variables may have little, if any, effect on specific sensory and possibly motor measures. Some support for this contention exists. For example, Roberts and Hamsher (1984) administered both the Facial Recognition and Visual Naming Tests of the Multilingual Aphasia Examination to African Americans in a consultation setting. They reported negligible racial bias. In contrast, Adams, Boake, and Crain (1982) found that bias did exist with regard to several variables, including ethnicity, in neuropsychological performance. In both brain-damaged and normal samples, African Americans and Mexican Americans exhibited more errors than did Caucasian participants. One may extrapolate from early (though questionable) motor-learning studies on race that motor measures may be affected by race. However, as implied, the data are questionable because of numerous methodological and theoretical issues. Other individual variables are definitely affected. Language, for example, is a difficult variable to measure across groups because it contains syntactical, grammatical, and cultural content that precludes a direct translation/interpretation of a specific concept. For example, the location in a sentence of nouns and verbs differs across certain languages. Another example involves the Spanish alphabet, which contains two additional letters, ñ and ll. Cognitive styles may similarly be affected because of variables, which directly affect cognitive manipulations, such as specific style or analysis of information. Additionally, indirect variables may play a role. Asians or Hispanics not acculturated to North American norms may find it difficult to permit a professional to examine "their minds." In certain subcultures this probing is allowed only by medicine men, witch doctors, or "curanderos." Thus, it may be impossible to obtain valid data because of the client's fear of testing.

While few individual neuropsychological tests have been adapted or translated, the two most widely used batteries, the Halstead-Reitan and Luria-Nebraska Neuropsychological Batteries, have been used with diverse populations. Both of these batteries have been translated into Spanish (HIRNB by Melendez; LuriaNebraska by Puente

and colleagues) and are presently being used in other cultures (e.g., Chinese). Of the two, the Halstead-Reitan may prove, at least initially, to be more adaptable since the focus is less on language function than is the Luria-Nebraska. In both cases, however, the lack of data from diverse populations is presently hindering their application.

The data that do exist, though extremely sketchy, may indicate the direction for future research. For example, complications are introduced in a report on sex, age, developmental variables, and cognitive functioning by Denno, Meijs, Nachshon, and Anrand (1982). Differences were noted on a variety of cognitive tests (e.g., Stanford-Binet) but only for four and eight year olds. Specifically, "white males scored the highest on all tests, followed by white females, black females and black males." Thus, variables such as sex and age may interact with race (and other variables). If these studies are found to be valid examples of neuropsychological measures, then a clear and easy identification of variables contributing to diversity of neuropsychological performance may not be feasible.

More recently, Perez-Arce and Puente (1996) reviewed the literature with a focus on understanding ecological validity of neuropsychological tests for Hispanics living in North America. If neuropsychological assessment focuses on problem solving, they suggest that different problem-solving strategies are employed by Hispanics. For example, many neuropsychological tests use time in assessing brain dysfunction. Whereas, in a competitive culture, like the American, time is a critical variable not to be wasted, and so on, the opposite is often true for other cultures. Hence, slowed performance, which could actually be interpreted by the Hispanic as prolonging a task of interest, would be incorrectly interpreted as "brain-damage". Unfortunately, in their review the authors conclude that outside the work of a few researchers such as Ardila (e.g., Ardila, Rosselli, & Puente, 1994), little is found in the literature.

### Personality and Pathology

Tests of personality could be generally categorized as one of two types—projective or objective. Projective or cognitive-perceptual tests (e.g., Rorschach) are quite commonly used with minority members because of their inherent ease of administration and superficial adaptability and interpre-

tation. According to preliminary analyses by Exner and Sciarra (personal communication, July 7, 1989), the Rorschach, an internationally accepted measure of cognitive-perceptual status, does not appear to be biased against Asian Americans, African Americans, or Mexican Americans. However, it is important to state that no data exist to support this contention.

In contrast, limitations of test adaptability are more readily investigated for issues such as bias with objective measures. Clearly, the most often used test of personality and psychopathology is the Minnesota Multiphasic Personality Inventory (MMPI). The homogeneity of the original MMPI sample limits its ready application to minority groups. According to Dahlstrom, Welsh, and Dahlstrom (1972), the norms used in the original MMPI sample were Caucasian, married, rural, blue-collar workers, with an eighth-grade education. However, Dahlstrom, Diehl, and Lachar (1986), and Lachar, Dahlstrom, and Moreland (1986) have suggested that even when important demographic variables are taken into account (e.g., race and socioeconomic status), approximately 12 percent to 13 percent of the total variance of the basic scales is accounted for. Still, the popularity of the test has resulted in translation into approximately 100 languages (Butcher, 1984; Williams, 1987) and a wealth of cross-cultural, ethnic, and racial studies based on research using this instrument have been published.

In an early review of ethnicity and the MMPI, Greene (1987) did an exhaustive examination of studies. Over 100 studies were analyzed according to type of scale and item level across groups including African American white, Hispanic white, Asian American, and Native American. Greene concluded that too many variables and too few adequately completed studies prevent conclusions of bias. The variables in question include subject parameters, ethnic group membership profile validity, moderator variable, and scores analyzed. Additional methodological considerations include appropriate statistical analyses, adequate sample size, and validity of statistical (versus clinical) significance. Based on his review, Greene provided the following four conclusions:

1. At this stage of our understanding, it is too premature to develop norms for specific ethnic and racial groups.
2. Subjects have to be identified with an ethnic group using subjective self (not clinician or experimenter) identification.
3. Empirical and not clinical differences should be emphasized.
4. Finally, more research needs to be focused on the special scales of the MMPI.

As exhaustive as the review is and as heuristic as Greene's conclusions may be, others advocate different orientations. For example, Gynther (1981), Gynther and Green (1980), and others argue that specific norms—and, in some cases, items—be developed, using an empirical methodology rather than a review of the literature. Better understanding of ethnic, cultural, and race differences and their application to interpretation of *T* scores, specific scale scores, or patterns preclude widespread use of the MMPI with minorities. For example, it seems foolish to group all Hispanics together as Greene and others have done. As Sue and Zane (1987) have indicated, being culturally sensitive is being aware of within-group heterogeneity. Further, little understanding appears evident in the MMPI research with regard to differences among culture, ethnicity, and race. Until such issues, as well as those outlined by Greene (1987), are resolved, not only will the MMPI data as it now stands be premature; it will be incorrect. According to a recent announcement from the Restandardization Committee of the University of Minnesota Press (1989) concerning the MMPI-2, published in 1989, the revised version will have "national norms that are much more representative of the present population of the U.S." (p. 4).

Dana (1996) criticizes the sampling methods of the MMPI. In the MMPI, for example, Hispanics were underrepresented by 2.8 percent. In the second MMPI-2 similar limitations are still noted. Further, differences are noted in scale performance between Hispanics and their Anglo-Saxon counterparts. Main differences are noted in the following scales; L, K, 3, and 4. He suggests that clinical interpretation must hinge on understanding the acculturation of the individual.

Using the recommendations of Velazquez, and colleagues (1997) for the use of the MMPI with Hispanics, the following recommendations are suggested for all ethnic minorities:

1. When options are available, use the most recent version of the test.



2. Administer the entire, and not a short, version of the test.
3. Appreciate prior test-taking experience of the test-taker.
4. Test in the language selected by the test-taker.
5. Evaluate results within a bio-psychosocial context.
6. Appreciate the effects of acculturation on test results.
7. Interpret results based on research literature and not on cultural stereotypes.
8. Always use a variety of test sources to arrive at conclusions.

### Achievement, Aptitude, and Interest

Achievement tests are still widely used in a variety of settings. A starting point involving achievement assessment is that of achievement motivation (Basic Behavioral Science Task Force of the National Advisory Mental Health Council, 1996). Indeed, it assumed that motivation plays a relatively small role in achievement testing. Nevertheless, access to models, ethnic-minority status, and related variables produce an initial "handicap" in such testing. Unfortunately, motivation appears highly correlated with scores on achievement tests as well as academic performance. Hence, what one might be measuring with ethnic minorities is not achievement as much as motivation.

What does exist, as with many other psychometric instruments, is a paucity of data. In chapter 7 of this handbook, there is a comprehensive review of achievement tests. Of the tests discussed in that chapter, the California Achievement Test (in education) and the Wide Range Achievement Test (in education and clinical application) are two of the most frequently used tests, which have been applied to nonmajority samples of the U.S. population. Initial findings regarding test bias in these measures reflect the conclusions outlined by Fox and Zirkin (1984) in the first edition of this handbook. Specifically, they suggest that while attention should be paid to the possibility of such bias, and while it may be intuitive that such bias would exist (at least on specific items), these tests should not be considered biased. This conclusion is in direct contrast to others, however. For example, Weiss (1987), considered the Scholastic Aptitude Test especially biased in the verbal section. While Golden Rule procedures have been applied to

reduce such biases, the reliability and validity of these tests may be in jeopardy (Linn & Drasgow, 1987). Thus, conflicts exist in terms of having a useful but unbiased test of achievement.

For tests of interest, even fewer data exist. While separate scales for sex are the rule and not the exception for measures of occupation it is generally assumed that other variables are of little importance. The same applies for interest surveys. For example, the Kuder Occupational Interest Survey (Form DD) (Kuder, 1966) as well as the Holland Interest Inventories (1978) consider academic major, occupational status, and even personality type, but not cultural, race, or ethnic factors. The Strong-Campbell is available in Spanish but the norms presumably are from non-Spanish-speaking samples. In a recent study, Drasgow and Hulin (1987) attempted to answer the question of whether scores on the Job Description Index (a vocational measure) varied across different Hispanic populations. Specifically, they compared bilingual Mexicans in Mexico City to other Hispanics residing elsewhere. While few differences were noted between the New York and Miami samples, large differences were noted between the U. S. and Mexican samples. Drasgow and Hulin concluded that both linguistic and cultural measurement equivalence must be addressed in measures of vocational interest.

What appears to exist is that differences between ethnic minorities and the criterion counterparts exist as early as the first grade (Task Force of the National Advisory Mental Health Council, 1996). Asian students appear to perform better than either Anglo-Saxons or other ethnic minorities. Such differences are even greater by the fifth grade. Asian families tend to explain this in motivational differences. In contrast, American mothers consider good performance of their children to be related to "natural" ability. Unfortunately, similar data are not available on Hispanic and African-American children. However, it is often thought that, even when intelligence is controlled, achievement differences may be due to biological variables in these ethnic-minority groups. According to the task force report, "Americans regard the need to try harder as evidence of low innate ability and are less likely to value or encourage such effort" (p. 725).

While it is assumed that ethnicity, race, and related variables have been explored by the Educational Testing Service and related psychological test corporations, again few scientific data exist in the public domain regarding tests of apti-

tude. Terman (1916) helped develop the now widely used Stanford Achievement Test for pre-college screening with no reference to minority groups. At the college level, the College Advanced Placement Examination is also widely used and accepted. However, data on minority populations is still lacking for both of these instruments since relatively few ethnic minorities enroll in such programs.

### Culturally Sensitive Measures

Traditionally, one method of avoiding test bias with regard to culturally different populations is to use instruments that are sensitive to and factor out cultural variables. Of the attempts to diminish test bias, the most significant effort has been by Cattell (1963). His Culture-Fair Intelligence Test measures intellectual abilities that allegedly factor out culture.

Cattell's basic aim was to factor out both cultural and educational variables from intellectual factors. Items were developed on common rather than culturally specific knowledge. Based on initial speculation, Cattell suggested that fluid intelligence was a function of biological factors including genetic and constitutional ones. In contrast, crystallized intelligence was a result of the development of fluid intelligence through environmental and cultural opportunities. While the Culture-Fair Test has been regularly used in the United States, its popularity has extended to non-North American populations. To date the instrument has been used with Nigerian (Nenty, 1986), Bulgarian (Paspalanova & Shtetinski, 1985), Italian (Stepanile, 1982), Spanish (Ortega-Esteban, Ledesma-Sanz, Lopez-Sanchez, & Prieto-Adanez, 1983), Israeli (Zeidner, 1987), and Indian (Ravishankar, 1982) groups. Unfortunately, the test has been shown to exhibit bias in some (e.g., Nigerian) though not all populations. In addition, these studies were completed with individuals residing in their own culture. It would be interesting to explore the efficacy of this test with minority cultures residing in the United States. While this test is promising in theory, additional research both in the United States and abroad will have to occur prior to its wider clinical acceptance. It is interesting to note that Cattell who died very recently has been accused as being biased. Indeed, this accusation was dramatically brought to the attention of psychologists by the 1997 American Psychologi-

cal Association Keynote Speaker, Ellie Weisel. Unfortunately, the blue-ribbon panel that was convened to explore these allegations was disbanded before any discussions ensued (partially due to the fact that Cattell removed his name from further consideration for the award for which he was being considered).

Of all standardized tests, the WAIS has received most attention with regard to cultural standardization. Two excellent examples are the Canadian and Puerto Rican versions of the test. Violato (1984) administered the standard or a revised version of the WAIS to 101 Canadians. The revised version contained eight items that were changed to increase face validity for Canadians. While bias effects were limited, the author did suggest that changes for Canadian administration of the WAIS were necessary. The WAIS has also been translated and standardized with Puerto Rican populations (1980). It was assumed that all translations would be appropriate; this assumption, however, is incorrect. Puerto Rican, Chicano, Mexican, Latin American, South American, and Castilian Spanish not only have their own dialects and idiosyncrasies but in many cases, their own language. Thus, the Puerto Rican translation of the WAIS has limited usefulness with non-Puerto Rican subjects. Further, though yet to be researched, the issue of norms needs to be addressed. For example, Puerto Rican norms may differ from Argentinean norms. Also, there is the question of when an individual, from one culture but residing in another, becomes acculturated enough to be administered the "new" culture's tests. These and related questions remain to be answered.

Other tests of intellectual ability which are purported to be culture-reduced or fair include Raven's Progressive Matrices—both Colored and Standard versions—as well as the Peabody Picture Vocabulary Test, the Quick Test, and the Army Beta. However, little evidence exists on the ability of these tests to be culture free. With the Picture Vocabulary Test serving as an example, several of the pictures on this test are useful for North American but not British populations. Another interesting example is that of the LuriaNebraska Neuropsychological Battery. Certain sections and stimuli are deemed culture free or culture reduced; but several of the visual stimuli come from Denmark and not Nebraska, making clear identification of specific items (e.g., nutcracker) an often difficult if not impossible task.

Less and less bias research is being done, as indicated earlier, since the belief that bias does not exist is so prevalent (Suzuki & Valencia, 1997). Hence, one might assume that tests such as the ones developed by Cattell will continue to lose favor as psychologists continue to, possibly incorrectly, assume that culture is not a critical factor in psychological assessment.

### Behavioral Assessment

In another section of this handbook, chapters on behavioral assessment are found. One major focus of this type of assessment is the assumption that behavioral, versus psychometric, approaches to assessment reduce the risk of focusing on psychic and nonobservable attributions. Psychometric focus may increase the potential for incorrect understanding of the behavior in question and, of course, is more likely to introduce bias in the assessment process.

Behavioral assessment focuses on empirically based methods of understanding behavior and, thus, the application to minority populations seems obvious. If psychometric tests are riddled with questions of culture, race, and ethnicity, then an assessment procedure, which focuses on behavior, and places the individuals in question in their environmental context, would seem an excellent alternative. Hence, it is surprising to note that this application has not been considered and researched adequately.

What scientific literature does exist is limited and, at best, preliminary. For example, Slate (1983) attempted to compare three nonbiased "behavioral" measures in retarded and non-retarded children across race and social class. Unfortunately, the results are so convoluted that they preclude an adequate understanding of any of these measures. Further, the possibility exists that behavioral measures may themselves be biased, partially due to the rater as well as the rating instruments. Letherman, Williamson, Moody, and Wozniak (1986) examined the effects of the race of the rater on the rating of social skills of African-American and white children. The results support earlier findings that the race of the child affects the ratings received. In addition, the researchers reported that racial bias effects were noted with both African-American and Caucasian raters.

While intuitive support exists for the use of behavioral assessment with non-mainstream popu-

lations, complications are evident in the literature. First, little data and even less clinical application of this approach are available. In addition, initial studies suggest that bias may still be present both in terms of the rated and the rater.

### SUMMARY

Understanding human behavior requires an understanding of human diversity. Unfortunately, historical foundations have dictated an incorrect understanding of how culture, race, ethnicity, and related demographic variables affect human behavior. This situation is evident in the traditional and current use of psychological tests to measure such variables as intelligence, achievement, abilities, aptitude, personality, and neuropsychological function.

Two factors appear to have guided this incorrect measurement of human diversity. First, pioneers such as Terman not only suggested that minorities were inferior but that their "proliferation should be controlled." Legislation and adjudication addressing minority bias continues to this day, especially in California and Texas, even at the level of the Supreme Court. Indeed, recent rulings on affirmative action call these issues into question. Second, few researchers, academicians, or clinicians have devoted time and effort to answering pertinent questions on human diversity, and even fewer have studied psychological assessment of diversity. Recently published statistics indicate that fewer ethnic minorities than in earlier years are pursuing graduate training in psychology or the study of human diversity. The lack of interested personnel is mirrored in faculty and clinical positions throughout North America.

Of course, the possibility exists that what is actually occurring is what has been previously described as "the false uniqueness effect" which is the tendency to overestimate one's personal positive attributes and underestimate other's abilities (Basic Behavioral Science task force of the National Advisory Mental Health Council, 1996). This Task Force reports that American children tend to think better of themselves when compared to others, presumably ethnic minorities. As adults, Americans tend to think of themselves as more attractive and intelligent than average. Further, 60 percent of students believed that they were in the top 10 percent of ability to get along with others—clearly an impossibility. One might assume

that what has transpired is that American psychology has failed to understand the possibility of "false uniqueness effect" and has confused such variables as test intelligence, which are highly influenced by economic and related factors, with biological and genetic superiority (see Jensen, 1980 for further information). It appears that psychologists have decided that when an ethnic-minority group does better than the majority group (as in the case for Asian Americans), the attributing variable is motivation and cultural variables (e.g., Susuki & Valencia, 1997). In contrast, when an ethnic minority does poorer than a majority group it is often ascribed to biological variables. Differential attribution of ethnic-minority differences—if better, it must be due to effort; if worse, it must be due to genetics—represents intellectual imperialism.

The obvious outcome is a field lacking in adequate data and much emotionality. The data that are available are clouded not only by a host of methodological problems but by researchers' gross misunderstanding of ethnic-minority group members and membership (including but not limited to within-group heterogeneity), especially in the context of majority-group behavior patterns. Regardless of the absence of data, social policy continues forward—often guided by political but not scientific correctness. Thus, much effort needs to be directed to the areas of research, teaching, and services to minority group members. Until additional adequate information is available, extreme caution should be used in the application of present knowledge of the psychological assessment of ethnic-minority group members and in the acceptance of previously considered "universal" theories of human function.

## REFERENCES

- Adams, R. L., Boake, C., & Crain, C. (1982). Bias in a neuropsychological test classification related to education, age, and ethnicity. *Journal of Consulting and Clinical Psychology, 50*, 143–145.
- Adebimpe, V. R. (1981). Overview: White norms and psychiatric diagnoses of black patients. *American Journal of Psychiatry, 138*, 279–285.
- Ardila, A., Rosselli, M., & Puente, A.E. (1994). *Neuropsychological evaluation of the Spanish speaker*. New York: Plenum.
- Basic Behavioral Science Task Force of the National Advisory Mental Health Council (1996). Basic behavioral science research for mental health. Sociocultural and environmental processes. *American Psychologist, 51*, 722–731.
- Beach, F. (1950). The snark was a boojum. *American Psychologist, 5*, 115–124.
- Bernal, M., & Castro, F. (1994). Are clinical psychologists prepared for service and research with ethnic minorities? *American Psychologist, 49*, 797–805.
- Berry, J.W. (1990). Psychology of acculturation. In J.J. Berman (Ed.), *Nebraska symposium of motivation, 1989: Cross-cultural perspectives, 37*, 201–234.
- Bersoff, D. N. (1981). Testing and the law. *American Psychologist, 36*, 1047–1056.
- Betancourt, H., & Lopez, S.R. (1993). The study of culture, ethnicity, and race in American psychology. *American Psychologist, 48*, 629–637.
- Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Brislin, R. W. (1988). Increasing awareness of class, ethnicity, culture, and race by expanding students' own experience. In E. S. Cohan (Ed.), *The G. Stanley Hall lecture series* (Vol. 8, pp. 137–180). Washington, DC: American Psychological Association.
- Bumam, M. A., Telles, C. A., Kamo, M., & Hough, R.L. (1987). Measurement of acculturation in a community population of Mexican Americans. *Hispanic Journal of Behavior Science, 9*, 105.
- Butcher, J. N. (1984). Current developments in MMPI use: An international perspective. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment*. Hillsdale, NJ: Erlbaum.
- Butcher, J. N. (1987). Introduction to the special series: Cultural factors in understanding and assessing psychopathology. *Journal of Consulting and Clinical Psychology, 55*, 459–460.
- Butchner, J. N., Braswell, L., & Raney, D. (1982). A cross-cultural comparison of American Indian, black & white inpatients on the MMPI and persisting symptoms. *Journal of Consulting and Clinical Psychology, 51*, 587–594.
- Cattell, R. R. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.
- Cole, N. S. (1981). Bias in testing. *American Psychologist, 36*, 1067–1077.
- Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation Rating Scale for Mexican American II: A revision of the original ASRMA scale. *Hispanic Journal of Behavioral Science, 17*, 275–304.
- Cuellar, I., Harris, L.C., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical population. *Hispanic Journal of Behavioral Science, 2*, 197–217.

- Dahlstrom, W. G., Diehl, L. A., & Lachar, D. (1986). MMPI correlates of the demographic characteristics of black and white normal adults. In W. G. Dahlstrom, D. Lachar, & L. E. Kahstrom (Eds.), *MMPI patterns of American minorities* (pp. 104-138). Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Lachar, D., & Dahlstrom, L. E. (Eds.) (1986). *MMPI patterns of American minorities*. Minneapolis: University of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook. Volume I: Clinical interpretation*. Minneapolis: University of Minnesota Press.
- Dana, R. H. (1984). Intelligence testing of American Indian children: Sidesteps in quests of ethical practice. *White Cloud Journal*, 3, 35-43.
- Dana, R.H. (1995). Culturally competent MMPI assessment of hispanic populations. *Hispanic Journal of Behavioral Sciences*, 17, 305-319.
- Denno, D., Meijs, B., Nachshon, I., & Aurand, S. (1982). Early cognitive functioning: Sex and race differences. *International Journal of Neuroscience*, 16, 159-172.
- Drasgow, F. (1972). Biased test items and differential validity. *Psychological Bulletin*, 92, 526-531.
- Drasgow, F., & Hulin, C. L. (1987). Cross-cultural measurement. *Revista Interamericana de Psicologia*, 21, 1-24.
- Eells, K. (1951). *Intelligence and cultural differences*. Chicago: University of Chicago Press.
- Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science*, 241, 31-35.
- Feuerstein, R., Rand, Y., & Hoffman, M. (1979). *The dynamic assessment of retarded performers: The Learning Potential Assessment Device, theory, instruments, and techniques*. Baltimore: University Park Press.
- Flaughner, R. L. (1978). The many definitions of test bias. *American Psychologist*, 33, 671-679.
- Fowers, B.J., & Richardson, F.C. (1996). Why is multiculturalism good? *American Psychologist*, 51, 609-621.
- Fox, C. H., & Zirkon, B. (1984). Achievement tests. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment*. New York: Pergamon Press.
- Fradd, S., & Hallman, C. L. (1983). Implications of psychological and educational research for assessment and instruction of culturally and linguistically different students. *Learning Disability Quarterly*, 6, 468-478.
- Franco, J. N. (1983). An acculturation scale for Mexican-American children. *Journal of General Psychology*, 108, 175-181.
- Franzen, M. D. (1989). *Reliability and validity in neuropsychological assessment*. New York: Plenum Press.
- Goddard, H. H. (1912). *The Kallikakfamily: A study in the heredity of feeble-mindedness*. New York: Macmillan.
- Gonzales, R. R., & Roll, S. (1985). Relationship between acculturation, cognitive style, and intelligence: A cross-sectional study. *Journal of Cross-Cultural Psychology*, 16, 190-205.
- Gould, S. J. (1981). *The mismeasurement of man*. New York: W. W. Norton & Company.
- Gould, S. J. (1996). *The mismeasurement of man*. New York: W. W. Norton & Company.
- Graham (1996).
- Grazianu, W., Varca, P., & Levy, J. (1982). Race of examiner effects and the validity of intelligence tests. *Review of Educational Research*, 52, 469-497.
- Greene, R. L. (1987). Ethnicity and MMPI performance: A review. *Journal of Consulting and Clinical Psychology*, 55, 497-512.
- Greenfield, P.M. (1997). You can't take it with you. Why ability assessment don't cross cultures. *American Psychologist*, 52, 1115-1124.
- Grubb, N. J. (1987). Intelligence at the low end of the curve: Where are the racial differences? *Journal of Black Psychology*, 14, 25-34.
- Guthrie, R. (1976). *Even the rat was white*. New York: Harper & Row.
- Gynther, M. D. (1981). In the MMPI an appropriate device for blacks. *Journal of Black Psychology*, 7, 67-75.
- Gynther, M. D., & Green, S. B. (1980). Accuracy may make a difference but does a difference make for accuracy? A response to Pritchard and Rosenblatt. *Journal of Consulting and Clinical Psychology*, 48, 268-272.
- Hall, Ch.C. (1997). Cultural malpractice. The growing obsolescence of Psychology. *American Psychologist*, 52, 642-651.
- Hayes, S. C., & Nelson, R. O. (Eds.) (1986). *The conceptual foundations of behavioral assessment*. New York: Guilford Press.
- Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- Herrnstein, R.J., & Murray, C. (1994). *The bell curve*. New York: The Free Press.
- Hickman, J. A., & Reynolds, C. R. (1986-1987). Are race differences in mental test scores an artifact of psychometric methods? A test of Harnington's experimental model. *Journal of Special Education*, 20, 409-430.
- Hiltonsmith, R. W. (1984). Predicting WAIS-R scores from the Revised Beta for low functioning minority group offenders. *Journal of Clinical Psychology*, 40, 1063-1066.
- Hoffmann, T., Dana, R., & Bolton, B. (1985). Measured acculturation and MMPI-168 performance

- of Native American adults. *Journal of Cross-Cultural Psychology*, 16, 243-256.
- Holland, J. L. (1979). *The self-directed search professional manual*. Palo Alto, CA: Consulting Psychologists Press.
- Howard, A., Pion, G. M., Gottfredson, G. D., Flattau, P. E., Oskame, S., Pfafflin, S. M., Bray, D. W., & Burstein, A. G. (1986). The changing face of American psychology: A report from the Committee on Employment and Human Resources. *American Psychologist*, 41, 1311-1327.
- Humphries, L. G. (1986). An analysis and evaluation of test and item bias in the prediction context. *Journal of Applied Psychology*, 71, 327-333.
- Jackson, G. D. (1975). On the report of the Ad Hoc Committee on educational tests with disadvantaged students: Another psychological view from the Association of Black Psychologists. *American Psychologist*, 30, 88-93.
- Jensen, A. (1980). *Bias in mental testing*. New York: Free Press.
- Jensen, A. R. (1986). Construct-validity and test bias. *Phi Delta Kappan*, 58, 340-346.
- Kuder, G. F. (1966). *General manual: Occupational interest survey form DD*. Chicago: Science Research Association.
- Kim, U., & Choi, S-H. (1994). Individualism, collectivism, and child development: A Korean perspective. In P.M. Greenfield & R.R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 227-257). Hillsdale, NJ: Erlbaum.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Lachar, D., Dahlstrom, W. G., & Moreland, K. L. (1986). Relationship of ethnic background and other demographic characteristics to MMPI patterns in psychiatric samples. In W. G. Kahlstrom, D. Lackar, & L. E. Dahlstrom (Eds.), *MMPI patterns of American minorities* (pp. 139-178). Minneapolis: University of Minnesota Press.
- Lampl, D. A., & Rust, J. V. (1986). Validation of the Kaufman Assessment Battery for Children with a sample of preschool children. *Psychology in the Schools*, 23, 131-137.
- Lawson, W. B. (1987). Racial and ethnic factors in psychiatric research. *Hospital and Community Psychiatry*, 37, 50-54.
- Leahey, T. H. (1997). *A history of psychology* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Lethermun, V. R., Williamson, D. A., Moody, S. C., & Wozniak, P. (1986). Racial bias in behavioral assessment of children's social skills. *Journal of Psychopathology and Behavioral Assessment*, 8, 329-332.
- Lindenmann, J. E., & Matarazzo, J. D. (1984). Intellectual assessment of adults. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment* (pp. 77-79). New York: Pergamon Press.
- Linn, R. L., & Drasgow, F. (1987). Implications of the Golden Rule settlement for test construction. *Educational Measurement Issues and Practice*, 6, 13-17.
- Lopez, S., & Romero, A. (1988). Assessing the intellectual functioning of Spanish-speaking adults: Comparison of the EIWA and the WAIS. *Professional Psychology: Research and Practice*, 19, 263-270.
- McGovern, T.V., Furumoto, L., Halpern, D.F., Kimble, G.A., & McKeachie, W.J. (1991). Liberal education, study in depth, and the arts and sciences major-Psychology. *American Psychologist*, 46, 598-605.
- Magana, J.R., Rocha, de la O., Amsel, J., Magana, H.A., Fernandez, M.I., & Rulnick, S. (1996). Revisiting the dimensions of acculturation: Cultural theory and psychometric practice. *Hispanic Journal of Behavioral Sciences*, 18, 444-468.
- Maheady, L., Towne, R., Algozzine, B., Mercer, J., & Ysseldyke, J. (1983). Minority overrepresentation: A case for alternative practices prior to referral. *Learning Disability Quarterly*, 6, 448-456.
- Malgady, R.G. (1996). The question of cultural bias in assessment and diagnosis of ethnic minority clients: Let's reject the null hypothesis. *Professional Psychology: Research and Practice*, 27, 73-77.
- Marin, G., Sabogal, F., Marin, B., & Otero-Sabogal, R. (1984). Development of a short acculturation scale for Hispanics. *Hispanic Journal of Behavioral Sciences*, 9, 183-205.
- Montgomery, G. T., & Orozco, S. (1984). Validation of a measure of acculturation for Mexican Americans. *Hispanic Journal of Behavioral Sciences*, 6, 53-63.
- Naglieri, J.A., & Das, J.P. (1996). *Das Naglieri Cognitive Assessment System*. Chicago: Riverside.
- 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. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Nenty, H. J. (1986). Cross-culture bias analysis of Cattell Culture-Fair Intelligence Test. *Perspectives in Psychological Researches*, 9, 1-16.
- Olmedo, E. L. (1981). Testing linguistic minorities. *American Psychologist*, 36, 1078-1085.
- Ortega-Esteban, J., Ledesma-Sanz, A., Lopez-Sanchez, F., & Prieto-Adanez, G. (1983). Profit of the academically successful student in the Spanish universities. *Scienia Pegagogica Experimentalis*, 20, 62-82.
- Padilla, A. M. (1985). *Acculturation and stress among immigrants and later generation individuals*. Spanish-Speaking Mental Health Research Center: Occasional Paper, 20, 11-60.

- Paspalanova, E., & Shtetinski, D. (1985). Standardization of the CF 2A Intelligence Test of Cattell for Bulgarian Population. *Psikhologiya Bulgaria*, 1985, 12--22 (translation).
- Perez-Arce, P., & Puente, A. E. (1996). Neuropsychological assessment of ethnic minorities: The case of assessing Hispanics living in North America. In R. Sbordone & C. Long (Eds.) *Ecological validity in neuropsychological testing*. Delray Beach, FL: St. Lucie Press.
- Phinney, J.S. (1996). When we talk about american ethnic groups, what do we mean? *American Psychologist*, 51, 918-927.
- Prasse, D. (1979). Federal legislation and school psychology: Impact and implication. *Professional Psychology*, 9, 592-601.
- Puente, A.E., & Salazar, G. (1998). Assessment of minority and culturally diverse children. In A. Prittera & D. Saklofske (Eds.), *WISC-III. Clinical use and interpretation*. New York: Academic Press.
- President's Commission on Mental Health (1978). *A report to the President from the President's Commission on Mental Health*. Washington, DC: U.S. Government Printing Office.
- Ravishankar, V. (1982). A correlational study of Cattell's personality factor B. and I. Q. as measured by his culture free test. *Indian Psychological Review*, 22, 9-11.
- Reschly, D. J. (1984). Aptitude tests. In G. Goldstein & M. Hersen (Eds.), *Handbook of psychological assessment*. New York: Pergamon Press.
- Reschly, D. J. (1988). Larry P. Larry P. Why the California sky fell an IQ testing. *Journal of School Psychology*, 26, 199-205.
- Restandardization Committee of the University of Minnesota Press. (1989). MMPI-2 Minnesota Multiphasic Personality Inventory. *Critical Items*, 4 (2), 1-4.
- Reynolds, C. R., & Brown, R. T. (1984). Bias in mental testing. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum Press.
- Reynolds, C. R., Chastain, R. L. Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristic of IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology*, 25, 323-342.
- Roberts, R. J., & Hamsher, K. D. (1984). Effects of minority status on facial recognition and naming performance. *Journal of Clinical Psychology*, 40, 539-545.
- Ross-Reynolds, J., & Reschly, D. J. (1983). An investigation of the item bias on the WISC-R with four sociocultural groups. *Journal of Consulting and Clinical Psychology*, 51, 144-146.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority group. *Journal of Consulting and Clinical Psychology*, 47, 919-927.
- Sandoval, J., & Mielle, M. P. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology*, 48, 249-253.
- Scarr, S. (1988). Race and gender as psychological variables: Social and ethical issues. *American Psychologist*, 43, 56-60.
- Schratz, M. M. (1978). A developmental investigation of sex differences in spatial (visual-analytic) and mathematical skills in three ethnic groups. *Developmental Psychology*, 14, 365-380.
- Schultz, D.P., & Schultz, S.E. (1996). *A history of modern psychology* (6th ed.). New York: Academic Press.
- Slate, N. (1983). Nonbiased assessment of adaptive behavior: Comparison of three instruments. *Exceptional Children*, 50, 67-70.
- Spitzer, R. L., Endicott, J., & Fleiss, W. (1967). Instruments and recording forms for evaluating psychiatric status and history: Rationale, method of development and description. *Comprehensive Psychiatry*, 8, 321-343.
- Spitzer, R. L., & Williams, J. B. W. (1980). Classification of mental disorders and DSM-III. In H.I. Kaplan, A. M. Freedman, & B. J. Sadock (Eds.), *Comprehensive textbook of psychiatry-III*. Baltimore: William and Wilkins.
- Sternberg, R.J. (1996). *Successful intelligence: How practical and creative intelligence determine success in life*. New York: Simon & Schuster.
- Sternberg, R.J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, 52, 1030-1037.
- Sternberg, R.J., & Williams, W.M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? *American Psychologist*, 52, 630-641.
- Stepanile, C. (1982). Contributo per una taratura italiana de test culture fair di Cattell. *Bollettino & Applicata* (161-164), 81-86 (translation).
- Sue, S., & Zane, N. (1987). The role of culture and cultural techniques in psychotherapy. *American Psychologist*, 42, 37-45.
- Suinn, R. M., Richard-Figueroa, K., Len, S., & Vigil, P. (1987). The Suinn-Law Asian Self Identity Acculturation Scale: An initial report. *Educational and Psychological Measurement*, 6, 103-112.
- Suzuki, L.A., & Valencia, R.R. (1997). Race-ethnicity and measured intelligence. *American Psychologist*, 52, 1103-1114.
- Terman, L. W. (1916). *The measurement of intelligence*. Boston: Houghton-Mifflin Company.
- Triandis, H. C. (1982). Acculturation and biculturalization indices among relatively acculturated Hispanic

- youths. *Revista Interamericana de Psicologia*, 16, 140-149.
- U.S. Bureau of the Census (1992). *Survey of Business Owners and Self-Employed Pearsons: Form MB-1*. Washington, D.C.: Department of Commerce.
- Velazquez, R.J., Gonzales, M., Butcher, J.N., Castillo-Canez, I., Apocada, J.X., & Chavira, D. (1997). Use of the MMPI-2 with Chicanos: Strategies for counselors. *Journal of Multicultural Counseling and Development*, 25, 107-120.
- Violato, C. (1984). The effects of Canadianization of American-biased items on the WAIS and WAIS-R information subtests. *Canadian Journal of Behavioral Science*, 16, 36-41.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition: Manual*. San Antonio: The Psychological Corporation.
- Wedding, D., Horton, A. M., & Webster, J. (1986). *The neuropsychology handbook*. New York: Springer.
- Weiss, J. (1987). The Golden Rule bias reduction principle: A practical reform. *Educational Measurement Issues and Practice*, 6, 23-25.
- Weschler, D. A. (1960). *Escala de Inteligencia de Weschler-Adultos*. New York: Psychological Corporation.
- Westermeyer, J. (1987a). Cultural factors in clinical assessment. *Journal of Consulting and Clinical Psychology*, 55, 471-478.
- Westermeyer, J. (1987b). Clinical considerations in cross-cultural diagnosis. *Hospital and Community Psychiatry*, 38, 160-165.
- Whitworth, R. H., & Gibbons, R. T. (1986). Cross-racial comparison of the WAIS and WAIS-R. *Educational and Psychological Measurement*, 46, 1041-1049.
- Williams, C. L. (1987). Issues surrounding psychological testing of minority patients. *Hospital and Community Psychiatry*, 38, 184-189.
- Williams, T. S. (1983). Some issues in the standardized testing of minority students. *Journal of Education*, 165, 192-208.
- World Health Organization (1973). *International pilot project on schizophrenia*. Geneva, Switzerland: Author.
- Yerkes, R. N. (1923, March). Testing the human mind. *Atlantic Monthly*, pp. 358-370.
- Zeidner, M. (1987). Test of the cultural bias hypothesis: Some Israeli findings. *Journal of Applied Psychology*, 72, 38-48.
- Zubin, J. (1965). Cross-national study of diagnosis of the mental disorder: Methodology and planning. *American Journal of Psychiatry*, 125, 12-20.
- Zukerman, M. (1990). Some dubious premises in research and theory on racial differences: Scientific, social, and ethical issues. *American Psychologist*, 45, 1297-1303.