CHAPTER 22

PSYCHOLOGICAL ASSESSMENT OF MINORITY GROUP MEMBERS

Antonio E. Puente

Little doubt is evident in either the professional or the lay literature regarding the contribution of psychology to the understanding of abnormal behavior. The doubt that is expressed is sparse and often politically motivated. This acceptance has arisen largely because empirically-based psychological principles are the foundation for assessment strategies in general psychopathology.

To those involved with the understanding of psychopathology in general, and measurement strategies of abnormal behavior in particular, one source of continuing concern is individual differences. Indeed, individual differences and diversity are often viewed as impediments to the development of general principles of behavior.

Beyond the theoretical and practical concerns raised by individual differences are the issues of confounding or measurement errors. 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., Halstead-Reitan) to assist in the discrimination or classification of specific diagnostic groups (i.e., those with organic brain dysfunction). Practitioners who do question their validity may not be examining the correct criterion variables. Regardless, expert diagnosticians are in agreement about the potential for classification of behavior through testing. In contrast, few would agree that affiliation with specific demographic groups (nondiagnostic) would be of great value in diagnostic classification. Presumably, this assumption is based on the concept that psychopathology (or for that matter abilities or any other behavioral variable) is relatively free from the contamination of these potential confounds (Westermeyer, 1987a). Thus, this perspective suggests that other (nondiagnostic) 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 observation.

Few attempts have been made to understand the behavior of individuals in minority groups, possibly because it did not matter. Brislin (1988) and others have cogently argued that psychologists for too long have categorically shown poor understanding of behavioral traits and patterns of individuals who do not belong to groups associated with mainstream America. This limited perspective of the nature of behavior was first addressed by the late 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. These trends persist a decade after the historical report in 1978 by the President’s Commission on Mental Health. Among other observations, this group concluded not only that minority groups are not adequately served but that too few professionals are available to deal with the issues relevant to nonmajority concerns.

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, Algazin, Mercer, and Yeysleyke (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. While it is unclear that “fairer” tests will produce less overrepresentation, it is certain that this trend persists.

The overrepresentation of minority groups in handicapped conditions has, in turn, resulted in negative stereotypes. 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 data base 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 enquiry 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 minority groups.

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 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 that presently exist.

This chapter is intended for North American audiences. Numerous limitations in the available data set, whether clinical or otherwise, would make a more geographically ambitious approach unworkable. 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 minority group members.

An initial step in understanding members of minority groups is to define such groups. According to a system often cited by psychopathology textbooks, 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 variables represent the most robust method as they should be free of bias due to the use of empirical behavioral methods to determine group composition. As a consequence, it would appear logical that minority group composition is determined as much as possible on behavioral and not other variables because we are essentially interested in behavior. Such an approach reduces confusion and highlights what is important—the behavior of the minority group member.

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, Brinlin’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., Jewish, 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 the area of psychological assessment, race has been the most widely studied of the previous variables. Sex, and to a lesser degree, ethnicity have been considered as potential variables. However, culture, gender, 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). However, as much as possible the named variables should be addressed in the psychological assessment of minority groups.

Regardless of the variable used, 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 2,000. Census figures suggest, for example, that by the year 2025 there may be more Hispanics than traditional Caucasian (white, Anglo-Saxon) Americans.

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 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 Hispanic differs from Caucasians. Thus, within-minority group analysis will eventually become as important as minority versus majority group comparisons.

HISTORICAL FOUNDATIONS

The origins of mental testing can be traced to Galton’s Inquiries into Human Faculty and Its Development in 1883 (Boring, 1950). In order to assess potential human defects (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, 1981).

The discrimination of acceptable and nonacceptable 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, 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 U.S. 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 systematic assessment 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” (pg. 68). Nevertheless, he concluded later 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 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 if not directly, at least as an undercurrent, in Goddard’s work and later in Robert Yerkes’ 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 (feebleminded) 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 non-whites tended to score lower, prompting Yerkes (1923) to
write in *Atlantic Monthly* about non-inherited racial differences. This conclusion readily supported the racist conclusion 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. Noyer and Victoria Woodhull) established a center for American eugenics in Cold Spring Harbor with financial support from the Carnegie Institution (Leahy, 1987). 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 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 "feebleminded" 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. For example, Spanish-speaking Hispanics were often placed in handicapped classes based solely on verbal IQ scores.

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 for 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. This chapter attempts to build on this new-found scientific interest in an effort to determine the needs, limitations, and directions associated with the psychological assessment of minority populations in North America.

### ASSESSMENT METHODS

This section of the chapter will focus on specific assessment methods, examining in turn standard measures, culturally sensitive methods, and behavioral assessment methods. As feasible, each section will cover a variety of tests or assessment strategies including application (and/or translation), norms, limitations and cautions, and suggestions for use.

#### Standard Measures

**Interview.**

The interview, whether structured or unstructured, remains not only the initial step of any psychological assessment but 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, including several presented in this chapter, seem generally well suited for use with minority populations, especially since they are often based on objective diagnostic
PSYCHOLOGICAL ASSESSMENT OF MINORITY GROUP MEMBERS

criteria (e.g., Research Diagnostic Criteria). Several of these methods are found in Table 22.1.

Although many 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, gender differences, and general cultural beliefs—are often misunderstood by even the most sensitive clinician (Westermeyer, 1987a).

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 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 be fluent in one but not necessarily both languages, or worse yet, not understand the culture in question. Distortion or misconception further impairs data gathering, especially with severely disorganized patients.

In order to bridge the gap between patient and psychologist, rapport should be established. Westermeyer (1987b) suggested that 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 is classified. 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 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, 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.

Intellectual.

Tests which attempt to measure the construct of intelligence are not only the most commonly used psychological tests (see chapters 4, 5, and 6) but also the most vehemently criticized. The literature is replete with controversies about the efficacy of the construct of intelligence and its measurability, and strong and often emotional arguments have been levied against tests of intelligence by members of minority groups. 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-Revised (WISC-R) is one of the most popular psychometric tests of intelligence. Nevertheless, conflicting and nonconclusive evidence has resulted from use of intelligence tests with minorities. For example, in one thorough review of the literature, the race of the examiner did not seem to affect the

<table>
<thead>
<tr>
<th>Table 22.1. Several Structured Interview Applicable to Minority Populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief Psychiatric Rating Scale</td>
</tr>
<tr>
<td>Inpatient Multidimensional Rating</td>
</tr>
<tr>
<td>Mental Status Schedule</td>
</tr>
<tr>
<td>Present State Examination</td>
</tr>
<tr>
<td>Structural Clinical Interview (DSM III)</td>
</tr>
<tr>
<td>International Classification of Disease Interview (for ICD-10)</td>
</tr>
<tr>
<td>Mini-Method Station Examination</td>
</tr>
<tr>
<td>References</td>
</tr>
<tr>
<td>Overall &amp; Gorham (1962)</td>
</tr>
<tr>
<td>Lorr &amp; Lorr (1969)</td>
</tr>
<tr>
<td>Spitzer, Endicott, &amp; Flenn (1967)</td>
</tr>
<tr>
<td>Wing (1970)</td>
</tr>
<tr>
<td>Spitzer &amp; William (1983)</td>
</tr>
<tr>
<td>World Health Organization (in press)</td>
</tr>
<tr>
<td>Folstein, Folstein &amp; McHugh (1975)</td>
</tr>
</tbody>
</table>
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-American 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.

Lampl 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 (1986-87) 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). 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 Weschler Adult Intelligence Scale-Revised (WAIS-R). 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) reanalyzed the data for the 1981 standardized sample of the WAIS-R and reported a 14% 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," Hiltonsmith and colleagues (1984) reported that these subjects actually scored lower on the Beta than on the WAIS-R.

Achievement, Aptitude, and Interest

Achievement tests are widely used in educational as well as clinical settings. However, as with many other psychometric instruments, relatively few data on minorities are available. In chapter 7 of the 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 United States 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
PSYCHOLOGICAL ASSESSMENT OF MINORITY GROUP MEMBERS

procedures have been applied to reduce such biases, the reliability and validity of these tests may be in jeopardy (Linn & Drsgow, 1987a). 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, Drsgow 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. Drsgow and Hulin concluded that both linguistic and cultural measurement equivalence must be addressed in measures of vocational interest.

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 aptitude. 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.

Personality.

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 interpretation. According to preliminary analyses by Exner and ScFra (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.

In contrast, limitations of test adaptability are more readily accounted for with objective measures. Clearly, the best example 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 normals used in the original MMPI sample were Caucasian, middle-class, rural, blue-collar workers, with an eighth-grade education. However, Dahlstrom, Diehl, and Lachar (1986), and Lachar and Dahlstrom (1986) have suggested that even when important demographic variables are taken into account (e.g., race and socioeconomic status), approximately 12% to 13% 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 excellent 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, and Asian-American-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). Whether this implies that a representative number of minorities included in the normative sample remains to be seen.

Neuropsychological.

It is often assumed that brain functions are not affected by non-neurological variables. To determine the current status of that assumption, the table of contents was reviewed in three major textbooks used to teach introduction to neuropsychology textbooks—Lezak (1987), Golden (1981), and Wedding, Horton, and Webster (1986). The review 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, Franzén (1989) 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 haphazard. On a more theoretical note, Dragay (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 Harmsner (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 non-samples, African-Americans and Mexican-Americans exhibited more errors than did white subjects. 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 often 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 of 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 (HRNB by Melendez; Luria-Nebraska by Puente and colleagues) and are presently being used in other cultures. The Luria-Nebraska has been successfully adapted for Chinese-speaking su-
jects. Of the two, the Halstead-Reitan may prove, at least initially, to be more adaptable since the focus is less on language function than in 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, Meijis, Nachshon, and Au- rand (1982). Differences were noted on a variety of cognitive tests (e.g., Stanford-Binet) but only for 4- and 8-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.

Culturally Sensitive Measures

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. 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 (Zeidler, 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 thrust is promising in theory, additional research both in the United States and abroad will have to occur prior to its wider clinical acceptance.

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 Coloured 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 Luria-Nebraska 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.

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 nonobserv-
able 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. The reader is directed not only to these chapters but to an excellent book by Hayes and Nelson (1986) for clarification of procedures and purposes of behavioral assessment.

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 the 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 nonretarded 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. Letherman, Williamson, Moody, and Wozniak (1986) examined the effects of race of rater on the rating of the 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 white raters.

While intuitive support exists for the use of behavioral assessment with non-mainstream populations, 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.

THEORETICAL ISSUES

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 assessment of mental function (Eells, 1951). Although the reasons for doing so are not entirely clear, some workers in the psychometric field generalized his findings to other populations, namely African-Americans. This generalization, an incorrect one, launched a wave of poorly developed and executed studies on bias in testing.

One of the most controversial figures in mental bias research is Arthur 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 probably due to 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.

In a more objective manner, 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.

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 than others. The following are two samples of the many available.

Eells (1951): [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. (pg. 54)
Jensen (1980): Psychometric bias is a set of statistical attributes conjointly of a given test and two or more specified subpopulations (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 (overinterpretation, sexism, selection model, and atmosphere). He concluded that in 1978 the research was promising, but the results were still disappointing.

Among the more current research findings, an excellent example is Dragows'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 support 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" (italics added) (p. 529). 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.

Despite these opinions, many questions still need to be formulated and answered. Until then, as Reynolds and Brown have concluded, the verdict on test bias is still not in.

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 were 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 Otéro-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 (Burnam, Telles, Karas, & Hough, 1987) as well as cultural awareness and ethnic loyalty (Padilla, 1985) are critical factors in the acculturation process. Frad 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.

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 exist. To deny the obvious would be foolish. Certain minority groups perform differently, at times worse, than majority groups on specific items, tasks, or tests. Such apparent 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 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 procedures and that all candidates, regardless of demographic affiliation, are treated equally. An alternative to this approach is equal outcomes, which implies that selection should match population demographics. Regardless of the approach and the data, the 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.

To assist policy makers, 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. 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 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. Similar trends exist in academic ranks, and presumably in clinical settings as well. 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. 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 minority group members looks bleaker than its past.

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 even at the level of the Supreme Court. 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 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.

The obvious outcome is a field lacking in adequate data. The data that are available are clouded not only by a host of methodological problems but by researchers’ gross misunderstanding of minority group members (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. 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, caution should be used in the application of present knowledge of the psychological assessment of minority group members.

REFERENCES


Gynther, M. D. (1981). In the MMPI an appropriate


