# Chapter 4

# Standards and Fairness

## Chapter Questions

- Are professional guidelines available for the assessment of individuals from diverse cultural, linguistic, ethnic, economic and other backgrounds?
- Why is it problematic to use the phrase, "valid test?"

The majority of problems that occur in applied clinical assessment are due not to inherent flaws in the tests, but to the inappropriate use of tests, and misinterpretation of their results, by clinicians (Anastasi, 1992). Test misuse is primarily due to substandard practice by clinicians, just as most auto accidents are caused by driver error and not by the car per se. Even a widely used and accepted test can become a tool for disserving a client. And these cases of misuse are common, and include misuses ranging from incorrect scoring to interpretations of scores that have not been shown to be valid by several independent research studies (Eyde et al., 1993). Consequently, psychological assessment practice has long been governed by peer-developed guidelines and standards that have proliferated and become more explicit and sophisticated as the field matures (AERA, APA, NCME, 1999). This chapter is devoted to providing an executive summary of some of the major publications in this area, especially those developed by relevant learned societies. It also provides guidance for practice based on some of the most widely cited ethical principles, test standards, regulations, and recent treatises that give suggestions for assessing diverse clientele. (A self examination for enhancing retention of these issues is given in Box 4.1.)



P.J. Frick et al., *Clinical Assessment of Child and Adolescent Personality and Behavior*,
DOI 10.1007/978-1-4419-0641-0_4, © Springer Science+Business Media, LLC 2010

**Box 4.1**

**Ethics and Standards Self-Examination Checklist**

Periodic completion of this checklist may serve as a quick reference for the clinician to cue adherence to optimal practice methods.

**Principle/Guideline Questions**

1. Do I have adequate training to use the tests/methods that I plan to use?
2. How might the individual's background-cultural, linguistic, social, economic, or otherwise-affect the planning of my evaluation or the interpretation of my results?
3. Are the tests that I am using validated for the specific purposes that I have in mind?
4. Are there particularly unreliable scales that I should refrain from interpreting?
5. Have I received informed consent and assent prior to initiating the evaluation?
6. Will I provide feedback to the client or to the others concerned, such as the child's parents, teachers, or pediatrician?
7. Did I adequately protect patient privacy?
8. Do I have written permission to share confidential information with concerned parties?
9. Whom do I need to assist with in this examination-a translator, patient, social worker, community member, etc.?
10. Have I consulted a professional colleague regarding questionable issues as needed?

## Use and Misuse

There are many elements of competent test usage. According to one empirical study of test usage, there are exactly 86 of them (Eyde et al., 1993). These competencies range from "accepting responsibility for competent use of the test to "not making photocopies of copyrighted materials" to "restricting test administration to qualified personnel." In their unique casebook, Eyde and her colleagues assembled 78 case studies of test "misuse," accompanied by instructional questions and documentation of each violation of the 86 elements. This work represents a unique effort to document the relationship between clinician behavior and assessment practices that makes it a recommended reading for students of assessment. Furthermore, the vignettes provide ample evidence that guidelines and standards are necessary for promoting optimal assessment practices.

## APA Ethical Principles of Psychologists

Virtually every professional organization adopts some ethical responsibility for its members. An initial step is the development and dissemination of ethical principles for the members of the organization. Many organizations also adjudicate ethical complaints against members made by the public or others.

Psychology has a long history of involvement in test development and assessment practice, resulting in the frequent use of the term *psychometrics*. Hence, the ethical standards published by the American Psychological Association (APA) are among the most well known sets of ethical principles promulgating standards for assessment practice. This section presents some of the relevant APA principles and provides sample applications of their use. These standards provide helpful guidance regarding the restriction of test use from a professional context only, requirements for evidence or scientifically-based test interpretation, restriction of test use to qualified persons, adherence to relevant testing guidelines and standards of practice, proper and full explanation of test results, and maintaining the security of test items and other content that may make the test useless if released to the general public.

It can be challenging to maintain the security of test content given the access to records by many sources, and numerous requests from parents, lawyers, or patients themselves to see the actual record forms used for the evaluation. Some internet sites provide sample items that are analogous to items found in popular tests, such as the MMPI-2. However, for the most part, test item content can be shielded from would-be test takers. When faced with questions about test security, it has commonly been considered a good practice to a) explain the problems associated with release if items have the ability to practice psychological assessment with others, or b) agree to release test record forms only to other qualified professionals, to interpret them appropriately for the person making the request. These two responses are an oversimplification of the various request types and potential responses, suggesting that consultation with colleagues will be considered wise under these circumstances.

## Test Standards

The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) discuss in great detail the many issues introduced by the ethical principles. In fact, this volume should be part of every psychologist's library, and it would serve as a useful adjunct to this or any other textbook dealing with clinical assessment.

This latest version of the test standards is ambitious, and includes chapters on validity, reliability, test development, scales and norms, test administration, scoring and reporting, supporting documentation for tests, fairness, rights of test takers, testing individuals from diverse linguistic backgrounds, testing individuals with disabilities, responsibilities of test users, psychological testing, educational testing, employment testing, and program evaluation and public policy. These standards are well-articulated and thorough, which portends that they are influential in court proceedings and forensic work, formation of public policy, and, hopefully, in the assessment training of psychologists and other users of tests (Kamphaus, 1998).

One of the vital points presented by the test standards is that it is incorrect to use the phrase "the validity of the test," because it cannot be concluded that a particular test is valid for all children under all assessment situations (AERA, APA, NCME, 1999). The validity of a test should be gauged properly in relation to every assessment situation in which it may be put to use. Clinicians who assess children's personalities must, therefore, learn how to use more than one personality test well in order to validly assess children, families, institutions, or systems. A small sampling of some of the important aspects of psychological test use and validation are given in the following excerpts from the test standards.

### Evidence-Based Interpretation

Given the scope and complexity of the standards, only a few of them can be summarized here. Of course, there is always some loss of content when one summarizes an original source. Therefore, again, the reader is advised to read the original standards. Also, the Test Standards should be a required study for all doctoral programs in psychology, as is currently the case for relevant ethical standards.

While it is understood that it is simplistic to say that a test is valid or not valid, it is, therefore, also incorrect to consider a particular test interpretation to be valid for all children being evaluated. First, test developers must present a rationale for each intended interpretation for their test

(AERA, APA, NCME, 1999). This rationale must include documentation of validity evidence and theory relevant to the interpretation in a "comprehensive summary" (Standard 1.1; AERA, APA, NCME, 1999). In addition, although test developers are required to produce such evidence, individual psychologists have to evaluate the quality of the evidence, as it relates to specific circumstances.

With regard to validity, the following standards are also offered (paraphrased from AERA, APA, NCME, 1999).

1. Uses and interpretations of test scores, intended populations of application, and the construct(s) assessed must be clearly stated by the test developer. (Standard 1.2)
2. Users should be cautioned not to interpret tests in a manner inconsistent with the available evidence. (Standard 1.3)
3. Psychologists must justify and collect evidence for an interpretation of a test that is not justified. (Standard 1.4)
4. Validity of study samples should be described in detail. (Standard 1.5)
5. Procedures for selecting the test content should be specified. (Standard 1.6)
6. If experts are used in the test design, their roles and credentials should be described in detail. (Standard 1.7)
7. A rationale should be offered for interpretation of items or item subsets when such interpretations are advised by the developer. (Standard 1.10)
8. The relationships among scores within a test should be supported with evidence. (Standard 1.11)
9. A rationale and scientific evidence should be provided for all interpretations of score differences and profiles. (Standard 1.12)
10. The conditions under which validity evidence was gathered should be described. (Standard 1.13)
11. The relationship between a test and scores on other measures should be theoretically consistent. (Standard 1.14)
12. When statistical adjustments, such as those for restriction of range, are made, both adjusted and unadjusted values should be given. (Standard 1.18)
13. If a test is used to recommend alternative treatments, evidence of differential treatment outcomes should be provided when feasible. (Standard 1.19)
14. When unintended consequences of test use occur, test invalidity should be ruled out as a cause for such consequences. (Standard 1.24)

With regard to psychologists, the following standards are given.

1. Psychologists should limit their assessment practice to the use of tests that are qualified by training. (Standard 12.1)
2. Psychologists should refrain from making biased interpretations that serve as special interests. (Standard 12.2)
3. Tests should be selected only if they are suitable for the characteristics of the patient. (Standard 12.3)
4. Evidence must be provided if it is suggested that an interpretation can be made based on combinations of test scores. (Standard 12.4)
5. Tests used in combination to make a diagnosis must show adequate validity (sensitivity and specificity), and the psychologist must meet the user qualifications required to interpret the tests involved. (Standard 12.5)
6. Psychologists should choose tests for differential diagnosis only if evidence shows that the test can differentiate between clinical samples of interest, not just between a clinical sample and the general population. (Standard 12.6)

7. When a test is used to help make a diagnosis, the diagnostic category must be carefully defined. (Standard 12.7)
8. Psychologists must ensure that psychometrists who work under their supervision are adequately trained. (Standard 12.8)
9. Psychologists should describe the testing procedures in a language understandable by the patient. (Standard 12.10)
10. Confidentiality of the results should be maintained consistent with legal and ethical requirements. (Standard 12.11)
11. Psychologists should use a setting and equipment necessary to obtain valid results. If this setting is not available, the test should also be administered under optimal conditions when possible for comparison purposes. (Standard 12.12)
12. Psychologists should be familiar with the reliability and validity of evidence of each test they use, and they should provide a logical and coherent analysis of the results that support their inferences. (Standard 12.13)
13. Qualitative information, such as background information and observations, should be considered when making test interpretations. (Standard 12.14)
14. The quality of actuarial or computer-based interpretations and the norms on which they are based should be evaluated for their quality. (Standard 12.15)
15. Psychologists should not imply that a relationship exists between test results and prognoses or treatment outcomes unless evidence is available for patients who are similar to the patient being evaluated. (Standard 12.16)
16. Interpretations that suggest how a patient will perform on other measures or outcomes should be supported by evidence of criterion-related validity. (Standard 12.17)
17. Psychologists should base their interpretations on several sources of data, test results, or other evidence, and they should be cognizant of the theory, empirical evidence, and limitations of each test used. (Standard 12.18)
18. Construct irrelevant factors (e.g., motivation, response sets, health factors, suboptimal testing conditions) should be considered as an alternative explanation for a set of test results. (Standard 12.19)
19. Normally, psychologists should discuss the results with the patient in a language that he or she can understand. (Standard 12.20)

One reason for delineating these two sets of standards is to give the reader an appreciation of their scope, which is truly impressive. Awareness of the scope of the standards should alert psychologists that a compliance with these standards requires dedicated effort and self monitoring.

A few themes in these standards deserve elaboration. First, interpretation must be evidence-based and, when it is not, the fact should be made known to all consumers of the results. We suggest that psychologists ask themselves a few questions about a test interpretation in order to encourage compliance with the standards. If we conclude, for example, that a child has depression, our self-monitoring questions could include:

1. On what assessment results do we base this interpretation?
2. Do the structured interviews and self-report measures have adequate evidence of reliability, sensitivity, and specificity?
3. Is this scientific evidence based on adolescent samples with similar demographics (e.g., sex, age, ethnicity, geographic region, language, etc.)?
4. Is this evidence based on studies that use clinical control samples in addition to comparisons with the general population?

5. Are the scales and interviews free of construct-irrelevant variance (e.g., "threatens others," which is not a core symptom of depression but is an indicator of aggression)?
6. Is there an evidence of construct under-representation (e.g., scale does not include any "vegetative" symptoms of depression)?
7. Are the procedures for developing diagnostic interview and scale content described and are reasonable?
8. Were the interview results, test findings, patient background/history, and other evidence integrated into a coherent rationale for the diagnosis?
9. Could there be an alternate explanation for the conclusions drawn by the construct-irrelevant variance (e.g., child was coached on the content of some test items)?
10. What is the definition of *depression (e.g., DSM-IV* diagnostic criteria)?
11. Is a classification or diagnostic decision made only to receive insurance reimbursement or other remuneration, financial or otherwise?
12. Have these results been shared with the client in a manner that he or she can fully understand?
13. Has confidentiality of the clients' results been maintained?

Of course, a different set of questions may be posed for non-diagnostic interpretations such as prognosis, treatment, or program evaluation.

The issue of confidentiality is worth additional comment. Rights to privacy continue to be threatened due to changes in health insurance practices and increasing access to electronically stored information, among other factors (Alderman & Kennedy, 1995).

*Standards for Privacy of Individually Identifiable Health Information* have been issued by the U.S. Department of Health and Human Services (Federal Register: December 28, 2000, volume 65, number 250). These standards are remarkable in at least three ways: (1) They specifically include records of psychologists who provide "qualified psychologist services"; (2) Psychotherapy notes that are "separated from the rest of the individual's medical record" are generally excluded from release, including release to the patient; (3) A patient may be denied access to his or her medical record if

> "a licensed health care professional has determined, in the exercise of professional judgment, that the access requested is reasonably likely to endanger the life or physical safety of the individual or another person." (p. 82823)

These medical record release standards are, therefore, reflective of some of the vicissitudes of mental health care. The exclusion of psychotherapy notes from the medical record shows a high regard for the sensitivity of such information. For the purposes of this text, however, it should be noted that psychotherapy notes are not de-fined to include the "results of clinical tests," "diagnosis," and so on.

With regard to the other aspects of the medical record that may include psychological test and assessment results and interpretations, a psychologist may also deny access if the "protected health information" makes reference to another individual who may suffer harm if information is released. Finally, release may be denied to the patient's personal representative, if the provider thinks that harm may occur. Of course, a patient may ask for a review of any denial request and certain exclusions (e.g., criminal activities) from this denial provision are stipulated. However, these regulations appear to give psychologists some discretion regarding the release of patient records.

## BIAS and Cultural Competence

### Test Bias

The perception that bias is inherent in psychological tests has spurred many challenges and accusations (Kamphaus, 2001). A review of some of the major technical issues follows.

#### Mean Score Differences

Most definitions of test bias do not usually consider the issue of mean score differences as a meaningful test of bias (Reynolds & Kaiser, 1990). Instead, important questions related to the validity of a personality test's inferences across groups forms a test of bias. In this approach to bias, an evidence of the *construct validity* for a personality test score inference differing across groups exist. Numerous studies have addressed these technical issues. For the purpose of this chapter, the definition of test bias offered by Reynolds and Kaiser is most appropriate.

> "Test bias refers in a global sense to *systematic* error in the estimation of some true value for a group of individuals. The key word here is *systematic;* all measures contain error, but this error is assumed to be random unless shown to be otherwise." (p. 624)

Given this definition, one would expect score differences to indicate genuine differences in behavior or personality, if the test works the same way (i.e., measures the same constructs in a valid way) for various groups. An interesting finding with regard to ratings of child behavior is that few mean differences exist between cultural and linguistic groups even for tests developed primarily for use in the United States. Crijnen, Achenbach, and Verhulst (1997), for example, found remarkably small differences between groups for parent ratings using the Child Behavior Checklist. They studied the results for 13,697 children and adolescents from 12 cultures including China, Israel, Sweden, German, Jamaica, and the United States. Their results have a striking similarity across culture, including similarities in cross-sectional changes associated with age. They also noted that sex differences were invariant across cultures:

> With no significant exceptions, boys obtained higher externalizing scores but lower internalizing scores than girls. This gender difference in the kinds of problems that parents report might thus be a "cultural universal"… (p. 1276).

Several aspects of this study were corroborated for four cultural groups, for both parent and teacher ratings in a study by Kamphaus et al. (2000). This investigation evaluated the differences between U.S. Anglo, U.S. African American, U.S. Hispanic, and Colombian (Medellin) samples for both parent and teacher ratings from the BASC. Kamphaus et al. also found that differences were small between groups for both parent and teacher ratings, and sex differences were consistent regardless of cultural/linguistic groups.

Taken together, these studies, and many others, revealed that differences among racial, ethnic, cultural, and linguistic groups for behavior problem measures are relatively small, a trend that stands in contrast to other realms of testing (i.e., intelligence and academic achievement) with large differences (Kamphaus, 2001). And, while sex differences are relatively small for cognitive measures (i.e., intelligence), they are greater for child behavior problem measures. While interesting, however, these studies of mean score differences do not serve as clear indices of bias.

### Content Validity Bias

Content validity was one of the first areas of investigation of test bias. This search for bias is understandable given that a

frequent bias concern is usually directed at item content that seems inappropriate or, perhaps, even offensive to a group of individuals. Again, a very helpful definition of content validity bias can be taken from Reynolds and Kaiser (1990):

> "An item or subscale of a test is considered to be biased in content when it is demonstrated to be relatively more difficult for members of one group than for members of another in a situation where the general ability level of the groups being compared is held constant and no reasonable theoretical rationale exists to explain group differences on the item (or subscale) in question." (p. 625)

Numerous procedures have been proposed for assessing bias in individual items, but the logic behind item bias detection techniques is fairly simple (Kamphaus, 2001). The fundamental aspect of most statistical methods that assess bias across cultural or gender groups is to match the groups on an overall score level, which is the first step in the procedure. If, for example, one was looking for gender bias in a pool of personality test items, one would first match boys and girls on their overall test score, be it standard or raw score. So, if one wanted to evaluate biased items in the MMPI-A, for example, one would first statistically group the cases, with perhaps all the boys and girls with f-scores above 90 on a particular scale as one group, those between 80 and 89 as another group, those between 70 and 79 as a third group, and so on (it should be noted, however, that this is not the exact procedure used by most item bias techniques but an oversimplification of such procedures). Subsequently, some statistical test of significance is applied to check if, within these various score groups, there are still significant differences in response to the items of one group or another.

This discussion relates to another item bias detection technique: judgmental bias reviews. The procedure used by some publishers is to have groups of individuals review the items carefully. This procedure ensures that members of a number of cultural groups review the items to determine not only the potential bias, but also the items that may be inappropriate for various cultural groups. There is, however, much disagreement between judgmental reviews of items and statistical analyses of bias. It appears that statistical analyses of bias are more reliable (Reynolds and Kaiser, 1990). In an investigation of judgmental bias reviews for intelligence test items, Sandoval and Mille (1979) compared the ratings of 45 WISC-R items by 38 African American, 22 Mexican American, and 40 undergraduate students. This study found that minority and non-minority judges did not differ in their ability to identify the culturally biased items. The conclusions of Sandoval and Mille were that: (1) Judges are not able to detect items that are more difficult for a minority child than for a Caucasian child, and (2) the item selection for minority children by judges of ethnic background did not show any difference.

Item bias, however, may be subtle and difficult to detect. Canino and Bravo (1999) cite an example of a problem with content equivalence on the Diagnostic Interview Schedule for Children (DISC; Shaffer et al., 2000). They found that, initially, it was difficult to translate items regarding seasonal depression into Spanish. Later it was found that even successful translation of the symptoms was of no value because seasonal depression never occurred for the children in sunny Puerto Rico.

### Construct Validity Bias

A workable definition of construct validity bias by Reynolds and Kaiser (1990) reads as:

> "Bias exists in regard to construct validity when a test is shown to measure different hypothetical traits (psychological constructs) for one group or another, or to measure the same trait but with differing degrees of accuracy" (p. 632).

The most popular method used for the study of construct validity bias is factor analysis. Numerous researchers have used similar procedures. The central characteristic of these procedures is to conduct factor analyses separately for various cultural and gender groups, and determine if a similar factor structure is yielded for each group. The most popular procedure for assessing agreement between factor structures across groups is a coefficient of congruence, which is interpreted similar to a correlation coefficient.

Lachar and Gruber (1994) provide an example of this method for the Personality Inventory for Youth (see Chap. 6). They conducted factor analyses separately by gender and ethnicity, and then compared the factors yielded separately for the groups. Their findings were similar to those for ability tests (Kamphaus, 2001) in that correlations between the obtained factors were uniformly *high*, in the low 0.90 at their worst.

### Predictive Validity Bias

The final type of bias that has received a great deal of attention is predictive validity bias. A working definition of predictive validity bias is

> "A test is considered biased with respect to predictive validity if the inference drawn from the test score is not made with the smallest feasible random error or if there is constant error in an inference for prediction as a function of membership in a particular group". (p. 638) (Reynolds & Kaiser, 1990)

The issue of predictive or criterion-related validity is that these coefficients should not differ significantly across cultural or gender groups. One of the typical procedures in this research literature is to compare the predictive validity coefficients across groups. A study might compare the ability of a depression measure to predict future adjustment for various groups, for example.

If different predictive validity coefficients were obtained for two or more groups, the results would be called *slope bias*. In order to understand the concept of slope bias, it is helpful to recall how correlation coefficients are learned in introductory statistics courses. Such procedures are typically taught by having the students collect data on two variables and plot the scores of a group of individuals on these two variables. This plot results in a scatter plot. Then students compute a correlation coefficient and draw a line of best fit through the scatter plot. This line of best fit is a visual representation of the slope. A correlation coefficient (predictive validity coefficient) of 0.90 would produce a slope that is very different from that obtained with a correlation coefficient of 0.30. Consequently, this form of bias in prediction is often referred to as slope bias.

### Summary Comments on Bias

While psychometric evidence of test bias can be found, little compelling evidence of bias is found for various groups residing in the United States (Figueroa, 1990). As a result, the focus has now changed to implicate test misuse as the major contributing factor to improper assessment of individual and groups of children. This misuse, however, includes more than individuals. Government, school district, or other entities may, for example, create unwise policies that inadvertently produce biased and untoward outcomes for children, such as imposing strict cut-off scores that affect assessment and conceptualization of the case.

## Fairness

The term *fairness* refers to "… the principle that every test taker should be assessed in an equitable way" (AERA, APA, NCME, 1999, p. 175). Some issues related to achieving this objective are discussed in this section. The renewed focus on test use comes

at a time when psychologists are seeking to improve test use for various linguistic and cultural groups. Cultural plurality has posed a challenge to assessment and diagnostic practice since the early days of the mental tests. The testing movement was forced early on to change tests and testing practice in the United States because of the tremendous influx of new immigrants. Between 1901 and 1910, over nine million immigrants entered the United States-more immigrants than the combined populations of New York, Maryland, and New Hampshire in 1900 (French & Hale, 1990). One component of the initial appearance of the Wechsler scales as an alternative to the Stanford-Binet monopoly of the time was the fact that Wechsler included a performance scale that could be used with some success with non-English speakers (Kamphaus, 2001).

Little has changed since the days of mass migration to the United States. In many ways, psychologists have used the same strategies for dealing with clients from diverse cultures. A popular approach involves adapting existing assessment instruments. The Thompson adaptation of the TAT for adults of African American heritage during the 1930s is one of the early examples of such attempts. Psychologists with multi-cultural expertise should be able to adequately assess the needs of a child from a culture that may differ from their own, even if test instruments that are not specifically designed for the child's culture are the only ones available.

A study by Malgady and Costantino (1998) highlighted the need for developing new cultural competencies. They evaluated the effects of language and ethnicity of the clinician on diagnostic decision making, using Spanish-dominant adult patients of Puerto Rican and Dominican descent. These patients were then seen by board-certified psychiatrists and clinical psychologists for diagnostic interviews. Patients were matched on *DSM-IV* diagnosis and divided into four interview language and clinician assessment groups: English only/non-Hispanic clinician; Spanish only/Hispanic clinician; English only/Hispanic clinician; Bilingual/Hispanic clinician. Several of the Malgady and Costantino's results are noteworthy. First, they found no differences in the diagnoses or symptoms between psychiatrists and psychologists. Symptom severity was highest among patients interviewed in Spanish by a bilingual interviewer. Symptom severity was rated lowest when a patient was interviewed in English by an Anglo clinician. While concern has been expressed that a clinician who does not share ethnicity and language with a patient will pathologize (Cohen & Kasen, 1999), these results suggest that it is also possible that such a mismatch between patient and clinician could lead to failure in identifying psychopathology, which could exacerbate the symptoms of those denied access to treatment. Either way, research does suggest a need for broader training of clinicians in multi-cultural competencies.

## Emic Versus Etic Perspectives

An *emic perspective* refers to behavior that is considered specific to a culture, whereas an *etic perspective* presupposes that much of the behavior and laws of psychology are applicable cross-culturally. Anastasi (1992) proposes that both perspectives are valid by theorizing that learned behavior may be culture-specific (emit) but that the "laws of learning" apply cross-culturally (etic). She hypothesizes further that hierarchical models of personality may be most useful in studying their behavior, as is the case for studies on intelligence when supporting evidence can be found for a "g" factor and for specific traits (e.g., spatial ability) at lower levels of the hierarchy. There is, for example, evidence that several temperament traits can be identified cross-culturally (Martin, 1988).

Inappropriate, ill-informed, or insensitive interpretations may also be made of

the "clinical" data. An examiner may conclude that a 13-year-old girl of Asian heritage is socially introvert, shy, and perhaps in need of assertive training because of her behavior during an interview with a male clinician. She may have been demure and made no eye contact. The examiner may draw such a conclusion despite the fact that she appeared friendly and outgoing when she was observed on the school playground and seemed to interact openly with her family members. This client may not, in fact, be pathologically shy; rather, she may be adhering to a prohibition against making eye contact with a male because of cultural values that suggest that this is a sexually seductive behavior (or an indication of a lack of respect) that is deemed inappropriate for her (Hasegawa, 1989). In this case, the clinician was simply ignoring relevant data, and the clinician's lack of familiarity with the child's culture resulted in an erroneous interpretation.

The clinician, however, must also remember the importance of individualizing interpretation. Within a cultural group, variability can be substantial (Zuckerman, 1990). It may be assumed by some that Vietnamese and Chinese children have similar values due to early Chinese domination and the inculcation of Vietnamese culture with Confucian ethics. There have also been other influences on this culture that may affect a child's behavior, including European Roman Catholicism, brought by the French conquest of 1958, the influence of U.S. culture from the Vietnam War, and Buddhist influences from neighboring Cambodia (Huang, 1989). Classifying children by race, culture, or language background is an appealing approach for researchers and clinicians alike that is fraught with errors, primarily due to the tendency to overgeneralize a particular group of people (Zuckerman, 1990).

Inclan and Herron (1985) cite the "culture of poverty" as another subculture that may affect a variety of groups. This "culture" is formed by a clash between those who have achieved material wealth and prosperity and those who struggle to achieve economic parity with little hope of doing so. Children reared in a culture of poverty possess identifiable characteristics: an orientation to present time, inability to delay gratification, impulsivity, sense of predetermined fate, resentment of authority, alienation and distrust of others, and lack of emphasis on rigor, discipline, and perseverance (Inclan & Herron, 1989). They note that some impoverished parents of adolescents may be assessed by a therapist as being too rigid and controlling their youngsters at a time when parents should be giving their children more freedom. It is possible, however, that poor parents may be all too familiar with the culture of poverty and may be seeking control, not for its own sake, but rather to ensure that their child or adolescent does not fall prey to the negative consequences of the behavior associated with that culture (Inclan & Herron, 1989).

These examples demonstrate the need for clinicians to develop an enlarged knowledge base in order to deal effectively with their referral population. Just as clinicians need to have knowledge of behavioral principles, psychometrics, child development, child psychopathology, and physiological psychology to conduct an evaluation competently, it is increasingly clear that they must know the history, culture, and language of their community extremely well in order to not use assessment procedures inappropriately, and to avoid making naïve and inappropriate interpretations.

## Guidelines for Assessing Children from Diverse Backgrounds

Resources for assessing children from diverse groups are now more readily available (e.g., Dana, 1999; Geisinger, 1992). Two developments that can assist practitioners are (1) the availability of guidelines

from blue ribbon panels and committees and (2) the increasing availability of formal measures of acculturation.

Numerous sets of guidelines provide specific advice for the psychologist who is unsure of what procedures to use in questionable situations. The *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations* give specific and helpful advice to the clinician seeking to carry out a competent evaluation of a child for whom cultural/social/linguistic issues loom large (see www.apa.org).

A good example of the guidance to be gained from such publications deals with the frequently occurring situation of a language difference between the examiner and the child or other family members. Guideline 6a suggests that a cascade of three options applies to the examiner faced with such a case: (1) Refer the child to a clinician who can communicate in the client's preferred language; (2) if this is not possible, use a translator who also possesses professional training; and, lastly, (3) one is advised to use a paraprofessional from the community to translate. Moreover, the next guideline, 6b, highlights the potential threat to validity of using a translator who has a dual relationship with the client (e.g., a grandparent).

## Assessing Acculturation

The previously discussed guidelines for considering cultural and linguistic issues hint at the need to more carefully assess an individual's level of adoption of the so-called "dominant culture," which, of course, could change from one neighborhood to the other. The guidelines indicate the need to collect information, such as the number of generations of residence within the dominant culture, number of years of residence, dominant language fluency, community resources, and so on. This data collection is an informal means of assessing level of acculturation. There are, however, more formal (some are quantifiable) methods for assessing acculturation. In fact, it has been argued that the ready availability of such measures warrants their routine use in assessment practice (Geisinger, 1992).

Marin (1992) defines the constructs relevant to assessing ethnic identity and acculturation. He cites three components of ethnic identity: (1) "birth and gestational history, (2) culture-specific behaviors and practices (e.g., language), and (3) culture-specific attitudes that include adherence to a culture's values and norms as well as in-group and out-group attitudes" (p. 236). The process of acculturation is defined as "… changes in individuals that are produced by contact with one or more cultural groups" (p. 237). Several instruments are now available for assessing the ethnic identification of individuals and the degree to which acculturation has taken place.

Dana (1993) provides a detailed compendium of measures of acculturation and identification with a particular culture. Some of these scales are listed below.

### African American Measures

Developmental Inventory of Black Consciousness (DIB-C; Milliones, 1980)

Racial Identity Attitude Scale (RIAS; Helms, 1986)

African Self-Consciousness Scale (ASC; Baldwin & Bell, 1985)

### Asian American Measures

Ethnic Identity Questionnaire (EIQ; Masuda, Matsumoto, & Meredith, 1970)

### Hispanic American Measures

Acculturation Rating Scale for Mexican Americans (ARSMA; Cuellar, Harris, & Jasso, 1980)

Children's Acculturation Scale (Franco, 1983)

Cuban Behavioral Identity Questionnaire (CBIQ; Garcia & Lega, 1979)

Hispanic Acculturation Scale (HAC; Marin et al., 1987)

Children's Hispanic Background Scale (CHBS; Martinez, Norman, & Delaney, 1984)

Cultural Life Style Inventory (Mendoza, 1989)

Multi-dimensional Scale of Cultural Differences (MSCD; Olmedo, Martinez, & Martinez, 1978)

Multi-cultural Experience Inventory (MEI; Ramirez, 1984)

Behavioral Acculturation Scale (BAS; Szapocznik, Scopetta, & Aranalde, 1978)

Bicultural Involvement Questionnaire (BIQ; Szapocznik & Kurtines, 1980)

## Conclusions

The child psychologist of today has to become steeped in various ethical, legal, professional, language and cultural issues, and standards of practice that face the profession. It is necessary for the practitioner to seek this knowledge through experiences during graduate school and beyond. Continuous professional development is especially important in order to achieve fairness in the assessment process. Knowledge of cultural, linguistic, technology change, and other effects on assessment remains in its infancy, thus portending considerable change in the future.

## Chapter Summary

Some of the APA ethics principles of relevance to assessment include

1. Evaluation, diagnosis, or intervention in a professional context, competence and appropriate use of assessments and interventions, limiting use of psychological tests to qualified professionals, and maintenance of test security.
2. The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) provide a rich resource for maintaining standards of practice.
3. While psychometric evidence of content-, construct-, or criterion-related validity test bias can be found, little compelling evidence of bias is found for various groups residing in the United States. As a result, the focus has now changed to implicate test misuse as the major contributing factor in improper assessment of children.
4. The renewed focus on test use comes at a time when psychologists are seeking to improve test use for various cultural and linguistic groups.
5. The emic perspective refers to behavior that is thought to be specific to a culture, whereas the etic perceptive presupposes the behavior theory and laws of psychology that are applicable cross-culturally.
6. The *Guidelines for Providers of Psychological Services to Ethnic, Linguistic, and Culturally Diverse Populations* give specific and helpful advice to the clinician seeking to carry out a competent evaluation of a child for whom cultural/social linguistic issues provide threats to test validity.
7. There are now formal and quantifiable methods for assessing acculturation.