

Ortiz, Samuel O., and Stacey A. Lella. "Cross-Cultural Assessment." *Encyclopedia of School Psychology*. 2005. SAGE Publications.

The term *culture*, as Matsumoto (1994) defines it, refers to the collection of values, attitudes, traditions, behaviors, and language that are specific to a particular group of individuals who purposely transmit those characteristics from one generation to the next. Therefore, the term *culturally different* can refer not only to those individuals who come from a culture other than the United States, but also to those who may be a member of a particular U.S. subculture (e.g., those who live in poverty, the deaf community, Greek Americans, etc.).

With respect to cross-cultural testing, there are four main issues with which school psychologists should be knowledgeable:

1. The cultural loading and linguistic demands of standardized, norm-referenced ability tests
2. Norm sample representation and stratification of different cultural groups
3. Effects of cultural differences on performance on tests
4. Cross-cultural dynamics involving the examiner and the examinee

CROSS-CULTURAL ASSESSMENT

As the population of the United States becomes more diverse, the need for school psychologists to be competent in assessing individuals from cultures other than the U.S. mainstream has progressively intensified. *Assessment* is a broad term encompassing many different kinds of procedures and techniques. Cross-cultural testing is the use of standardized tests of intelligence and cognitive abilities with individuals who are culturally different and often linguistically different as well.

CULTURAL LOADING AND LINGUISTIC DEMANDS OF ABILITY TESTS

Intelligence and cognitive ability tests have long been criticized as unfair, inappropriate, and invalid for the purpose of assessing individuals who are culturally different. In general, they have been attacked as being culturally *biased*. In general, cultural bias has been defined historically as being psychometrically based. Such bias, therefore, has been sought with regard to item content, test sequence or difficulty, factor structure, prediction, and indices of reliability. For example, bias in item content might exist if tests that require knowledge of U.S. mainstream culture place culturally different individuals at a disadvantage because they have not been raised in the majority culture. Essentially, any aspect of the test that results in systematic differences in performance between one group and another, except for the actual ability being measured, could be construed as evidence of bias. Yet, when defined from the psychometric perspective, very few tests have ever shown evidence of bias. The reason may be that cultural differences in and of themselves simply do not alter the basic psychometric properties of a test, particularly not those related to reliability. However, the same cannot be said with respect to issues of validity.

Well-developed, modern tests of intelligence and cognitive ability may well be free of the types of psychometric-based cultural bias defined above; however, it cannot be disputed that such tests remain invariably loaded with cultural content and also require some type of linguistic or communicative ability. That is, all tests are a reflection of the very culture that gave rise to them and, by definition, will sample to varying degrees the knowledge and familiarity of that same culture possessed by any given individual. Test developers have long sought culture-free or culture-fair tests that eliminate cultural bias, but even tests that purportedly meet this goal contain elements with which individuals from some cultures would likely not be familiar (e.g., blocks and puzzles). Likewise, even so-called “nonverbal” tests require some form of linguistic or meta-linguistic communication and comprehension by the individual taking the test, albeit it may not require oral or spoken language. These factors are undeniably and inextricably linked to tests of ability developed in any country, and such tests are, therefore, inherently culturally loaded and linguistically demanding. If cross-cultural testing is to achieve

a greater degree of fairness, it must come in the form of understanding the manner in which cultural loading, not cultural bias, affects test performance.

Recently, Flanagan and Ortiz (2001) provided classifications of tests along the dimensions of cultural loading and linguistic demand that assist psychologists in selecting and interpreting tests that may be more appropriate for a given individual relative to his or her background. By viewing tests first from the perspective of cultural loading and linguistic demand, the basic question regarding whether performance is related more to differences in cultural or linguistic experiences versus differences in actual ability can be answered. That is, the validity of the obtained results can be directly evaluated, offering psychologists a systematic method for reducing the potential discriminatory aspects of cross-cultural testing with respect to issues of validity that ultimately underlie the fairness and equity of interpretations and conclusions.

NORM-SAMPLE REPRESENTATION

Norm-referenced measures are used to evaluate an individual's performance relative to other people (i.e., the norm sample). To make such comparisons, test developers strive to obtain a norm sample that is as representative of the general population as possible with respect to important variables that generally are associated with performance—age, grade, geographic location, socioeconomic status, and so forth. Whenever comparisons are made, it is assumed that the individual being tested is comparable or very similar to those in the norm sample in terms of his or her background experiences, particularly acculturation and language development. Salvia and Ysseldyke (1991) describe the underlying assumption of comparability relative to acculturation as the collection of background experiences; and formal and informal learning opportunities that involve familial, social, educational, cultural, and linguistic experiences. Only when an individual's background experiences are comparable, not necessarily identical, to the background experiences of the individuals who comprise the norm sample can it be said that comparisons regarding performance are valid.

It has long been the practice of test developers to include a heterogeneous mix of races in their norm sample development and routinely provide representation of the various major racial/ethnic groups that exist in the United States according to their proportion

in the general population. Although this makes some sense on the surface, it begs the question regarding whether differences in ability actually exist as a function of race or ethnicity. But test developers tend to defend inclusion of different racial and ethnic groups primarily as a way of controlling for the effect of culture on test performance. Unfortunately, race and ethnicity are not the same thing as culture. They are related, of course, but to equate them is a mistake and does little to directly address the variable that actually affects test performance—background experiences. People from different races and ethnicities may or may not share a common culture. For example, Puerto Ricans have a mixed racial heritage that includes native origins (the Taíno), African origins (from the slave trade), and Anglo-European origins (Portugal, Spain, and France). Thus, to equate culture with race or ethnicity misses the point because it is not race or ethnicity that results in group differences in performance, but rather the extent to which the acculturation and background experiences of the individual taking the test are similar to that of the individuals included in the norm sample. The validity of results obtained from the use of tests with culturally different individuals is, therefore, questionable, as it may likely be more a reflection of the extent to which the individual has learned information pertinent to U.S. mainstream culture rather than actual intelligence or ability.

Because acculturation varies directly and proportionally according to an individual's length of experience and exposure to U.S. mainstream culture, use of standardized, norm-referenced tests with culturally different individuals may become more an exercise in evaluation of acculturation than intelligence or ability. This is true even with tests that include norms sampled from other countries, because diverse individuals in the United States have backgrounds that are bicultural and bilingual compared to the monocultural and monolingual experiences of individuals raised and educated in their native countries.

EFFECTS OF CULTURAL DIFFERENCE ON TEST PERFORMANCE

Culture may interact with test performance indirectly through acculturation and linguistic variables as described or it can directly affect test performance. Huerta-Macias (1995) provides an example wherein she notes that her daughter had a negative reaction to “none of the above” as a response choice in a

fourth-grade multiple-choice test. Her daughter's lack of familiarity with the concept of “none of the above” illustrates that seemingly innocuous content may well carry implicit cultural expectations—in this case, the learning of particular test-taking strategies. Individuals may well know the concepts being tested but be unfamiliar or uncomfortable with the manner in which they are being measured, which adversely affects their performance. Children from culturally and linguistically diverse backgrounds may have difficulty with test taking because of a lack of exposure to the culture and testing situation as well as a lack of understanding of the importance of the test and the lack of practice or preparation. Similar limitations in exposure or experience with materials, content, or language (as a function of differences in cultural experiences) can also affect an individual's ability to answer questions or complete tasks that currently exist on tests of intelligence or cognitive ability. For example, Native American children are often taught to value reflective thought and deliberation as well as group harmony and a pluralistic community philosophy. This contrasts with some of the cultural implications of tests that tend to rely on speed of processing and responding, and individual excellence and competition. Western European cultural variables that tend to emphasize academic or intellectual achievement and individuality may create significant difficulties for individuals from diverse cultural backgrounds who have not been so inculcated or who, for whatever reasons, are simply disinclined to adopt such values. In general, individuals who are unfamiliar with and have had limited exposure to the mainstream U.S. society, the public education system, and the embedded cultural values therein are likely to be less adept when being tested than students who have been raised and indoctrinated in the mainstream U.S. culture.

CROSS-CULTURAL DYNAMICS BETWEEN EXAMINER AND EXAMINEE

The vast majority of research involving examiner–examinee effects in testing has relied upon racial differences between the two. Again, this is quite distinct from the dynamics that might result as a function of difference in culture, such as might occur with a Mexican examiner and a Cuban child—both possess an Hispanic ethnic heritage but have different cultural backgrounds. In general, when examiner–examinee effects have been studied with respect to

racial difference, very little evidence has been uncovered to suggest that performance is significantly or adversely affected. Perhaps the reason for such findings relates again to the manner in which such effects were defined (i.e., racially instead of culturally). Variables that relate to cultural differences such as examiner's perceptions, prejudices, and stereotypic beliefs about the examinee's race can have a negative effect on rapport with the student, and subsequent performance on tests, particularly if the cultural difference is linked to a racial difference.

Management of potential bias related to examiner-examinee effects rests primarily on securing training, education, and experience in serving diverse cultural populations. Although the opportunity to obtain such experiences may be rather limited, psychologists nonetheless must possess specific skills and knowledge that help to guard against the more egregious forms of errors in assessment that can stem from preconceived notions, racial bias, or cultural stereotypes. Such qualifications should include competence in selecting and using tests appropriate for the individual, knowledge of and sensitivity toward the examinee's culture and its effect on establishing rapport and the subsequent test performance, and knowledge of first and second language acquisition and their relationship to cognitive and academic development. In addition, where feasible, the ability to assess the individual in his or her native language is extremely valuable.

SUMMARY

The preceding discussed the primary issues relevant to the practice of cross-cultural testing. Modern, well-developed tests of intelligence and cognitive ability are rarely biased from a psychometric point of view, but remain potentially discriminatory and unfair for several reasons. Tests developed and normed in the United States will always be imbued with content and concepts that reflect mainstream U.S. values. Moreover, the degree to which an individual is expected to possess these values or be familiar with such content is a function of the length of time spent living and being exposed to that culture and language. Thus, individuals whose cultural and linguistic backgrounds and experiences differ significantly from those of the individuals on whom the test was normed will necessarily perform at a lower level than age- or grade-related peers directly in proportion to the degree of difference. The more *different*

an individual is with respect to cultural experiences (or linguistic experiences), the more his or her performance on tests of intelligence or cognitive ability will be compromised.

To mitigate the potential biasing effects that can arise in cross-cultural assessment, it is necessary for psychologists to understand the assumption of comparability and the inherent dangers of comparing the performance of culturally different individuals with those on whom the test was normed. In addition, psychologists should be knowledgeable regarding several background and experiential issues of the individuals they seek to test, including:

- Their level of acculturation and knowledge of culture-based information
- Their exposure to and familiarity with testing situations
- The implied concepts required by specific tests and tasks
- Culture-specific values that may affect aspects of response style, speed, or even motivation

Psychologists should also seek to develop the qualifications required for conducting competent cross-cultural assessments including skill in using less biased or nondiscriminatory methods, as well as an awareness of how their own beliefs and prejudices may affect the testing situation and examinee performance. Professional ethics mandate that psychologists be cognizant of when an assessment or evaluation may be beyond their level or scope of competence and know when to elicit the supervision of a more experienced or qualified professional.

—Samuel O. Ortiz and Stacey A. Lella

See also Academic Achievement; Bias (Testing); Intelligence; Multicultural Education; Race, Ethnicity, Class, and Gender; Reliability

REFERENCES AND FURTHER READING

- Flanagan, D. P., & Ortiz, S. (2001). *Essentials of cross-battery assessment*. New York: Wiley.
- Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8–11.
- Matsumoto, D. (1994). *Cultural influences on research methods and statistics*. Pacific Grove, CA: Brooks/Cole.
- Salvia, J., & Ysseldyke, J. E. (1991). *Assessment* (5th ed.). Boston: Houghton Mifflin.