Howard University

From the Selected Works of Kay T. Payne, Ph.D.

Summer June, 1983

Culturally valid testing--a proactive approach

Orlando Taylor
Kay Payne, Howard University

Available at: https://works.bepress.com/KayTPaynePhD/3/
Culturally valid testing: a proactive approach

Orlando L. Taylor, PhD  
Graduate Professor  
Department of Communication Arts and Sciences  
Howard University  
Washington, DC

Kay T. Payne, PhD  
Research Associate  
Department of Communication Arts and Sciences  
Howard University  
Washington, DC

Since the enactment of Public Law (PL) 94-142, the Education for All Handicapped Children Act of 1975, it has been federally mandated that all test materials and procedures used for the evaluation of handicapped children be selected and administered in such a manner that they are not racially or culturally discriminatory. If this stipulation were enforced today to its fullest intent, most school systems in the United States would be in violation of the law. Eight years after PL 94-142 promised to be the salvation for all handicapped children, little has been done to improve tests and other evaluation procedures for handicapped children, especially those with communicative handicaps, to make the tests linguistically and culturally valid.

Results of Discriminatory Testing

One of the results of the use of discriminatory tests in the school setting, including
tests of speech and language function, is the disproportionate and inappropriate placement of many children from culturally and linguistically diverse populations in special education classes. Many of these children do not suffer from permanent handicaps; rather, their performance on standardized tests is related to differences in language and dialect, values, outlook, world view, aspirations, and rules of social behavior. Once improperly identified as handicapped, these children are typically not afforded their full educational opportunities.

Discriminatory testing, of course, is not just a school issue. Faulty assessments of a person's communicative behavior at any age level and in any type of a clinical situation can lead to faulty management. In some cases, persons are inappropriately placed in therapy. In other cases, incorrect judgments are made with respect to the setting of priorities for clinical goals and objectives. Also, persons may be denied therapy by a well-meaning clinician who mistakenly interprets a mild communication disorder as a communication difference largely because available assessment procedures are not sufficiently sensitive to cultural and linguistic differences.

RESEARCH AND RULINGS ON NONDISCRIMINATORY TESTING

Non-discriminatory testing has been a topic of interest to speech and language professionals since the late 1960s, when Taylor, Stround, Moore, Furst, and Williams (1960), among others, argued that clinical judgments should not be based on linguistic norms and instruments that are inappropriate for the cultural group of a given client. The notion that tests and other assessment procedures should not be linguistically or culturally discriminatory was subsequently upheld by several important court decisions, including Larry P. v. Riles,1 Diana v. California State Board of Education,2 and Mattie T. v. Holladay,3 and by PL 94-142.

To ensure that tests and evaluation procedures used to identify handicapped students are neither culturally nor linguistically biased, PL 94-142 includes the following stipulations, among others:

- Testing and evaluation materials and procedures must be selected and administered so that they are not racially or culturally discriminatory.
- Testing and evaluation materials and procedures must be provided and administered in the language or other mode of communication in which the child is most proficient.
- The tests administered to a child with motor, speech, hearing, visual, or other communication disability, or to a bilingual child, must accurately reflect the child's ability in the area tested, rather than the child's impaired communication skill or limited English language skill.
- Tests and other materials must be properly and professionally evaluated for the specific purpose for which they are used and administered by qualified personnel in conformance with instructions provided by the producers of the tests.

Given the present state of the art in speech and language tests, it can be concluded that there are few, if any, standardized measures that can provide a completely valid and unbiased evalu-
There are few, if any, standardized measures that can provide a completely valid and nonbiased evaluation of handicapping conditions for linguistically and culturally diverse populations.

tion of handicapping conditions for linguistically and culturally diverse populations. For example, most standardized tests, if used in accordance with the publisher's instructions, would be culturally discriminatory because no specific norms have been established for various cultural groups and because representatives of diverse cultural groups were not used in the standardization sample. In addition, language tests are typically designed to measure facility with standard English, which makes it difficult to accurately identify language disorders in those students whose major mode of communication is a non-standard-English dialect.

Much of the literature concerning the issue of test bias has been concerned with the identification of sources of cultural and linguistic bias in specific tests (Hoofer, Politzer, & Taylor, 1974; Roberts, 1970; Wolfram, 1974). Many authors have suggested ways in which sources of bias in standardized tests may be eliminated, but relatively few have outlined specific means by which clinicians can provide nondiscriminatory assessments as required by law in the absence of nonbiased or culturally valid instruments and procedures. (A major exception to this trend is the work of Seymour & Miller-Jones, 1981.)

KEY DEFINITIONS AND CONCEPTS

At this juncture, it is appropriate to define some basic concepts pertaining to the issue of nonbiased or culturally valid testing. For purposes of this article, culture is defined as the set of institutions, rituals, values, world views, artifacts, and rules of behavior (including language) used by a group of people for the purpose of relating to their environment. The concept of culture is not isomorphic with race, although members of a given culture are often of the same race. In fact, members of the same race need not share the same culture. Variations typically exist within a culture. They are usually associated with factors such as age roles, region of settlement, gender roles, social status, economic status, and amount of formal education. Given this framework, it is reasonable to assert that cultural bias in testing is not just a race, class, or minority group issue. Tests might well be biased in the United States, for example, against the majority middle-class white population in a rural southern state if the linguistic and cultural presuppositions of the test are based on an urban, northern model.

It is also necessary to differentiate between assessment, testing, and measurement, since all three terms are typically used in speech-language evaluations and diagnostic processes. Assessment usually refers to the process by which data about a person are gathered and critically evaluated in an attempt to obtain an accurate view of the person and his or her adaptation to the environment. Testing, one of several approaches to assessment, is
the use of specific tests or defined test procedures for the purpose of generating a score or rating for an individual. Measurement is the process of generating objective scores, subjective ratings, or other quantitatively values or information that can be used in the assessment process.

In the ideal case, the clinical process for treating speech and language disorders begins with an assessment of communicative behaviors. This assessment may include test and non-test procedures that attempt to determine deviations of the client’s communicative performance from the expectations of factors, such as age, gender, education, socioeconomic class, region, cultural or racial group, and external factors, such as audience, topic, situation, and context. Culturally valid, or nonbiased, assessment is a data collection process wherein testing, measurement, and evaluation are conducted using instruments and procedures that discriminate only in those areas for which they were designed (i.e., normal versus pathological behavior) and do not discriminate unfairly either for or against a client for cultural reasons or because of social variations within a culture based on such factors as age, gender, socioeconomic class, and dialect.

**TYPES OF BIAS IN SPEECH AND LANGUAGE ASSESSMENTS**

Bias in the collection of clinical data of any type may take many forms. In addition, bias can exist even in procedures that are thought to be objective, such as observation and collection of spontaneous language samples. The principal forms of bias are:

- situational bias;
- bias in directions or format;
- value bias; and
- linguistic bias.

**Situational bias**

As far back as 1938, Morris discussed the pragmatic or functional dimension of language as a social device. Interest in the interactional dimension of language has escalated considerably in the past decade, as demonstrated by the work of Searle (1969), Dore (1974), Halliday (1975), and Bates (1976). This dimension of language has been incorporated into clinical models of communicative disorders (cf. Bloom & Lahey, 1978; Lucas, 1980).

The pragmatic dimension of language embraces the notion that language, whether verbal or nonverbal, is always generated within the framework of intentions and semantic references of speakers and the interpretations of listeners. The encoding process of the speaker and the decoding process of the listener are both based on culturally specific linguistic and communicative rules deemed to be appropriate for the topic, audience, situation, and intention (or perceived intention) of the communicative event.

In this context, then, one could view all assessments of communicative behavior as social occasions, even those in which an individual is observed in self-play or self-communication. Thus, the clinician is placed in the position of eliciting, recording, and evaluating language produced by the client in the framework of some set of communication rules.

Mismatches can occur between the clinician and the client with respect to the
social rules of language interaction. Some of the major areas of mismatching include rules of who may speak to whom, appropriate elicitation procedures, appropriate language behaviors, what behavior serves as communication, and rules of production and interpretation.

Because of these interactional rules, silence—which may be a communicative device—may be misinterpreted as a sign of a disorder when it is actually considered appropriate by the speaker. When given a standardized test in which an examiner from another cultural group or gender seeks responses to obvious questions, the respondent may choose to give no response or an incorrect answer that is perceived to be what the examiner wants to hear. In any case, the respondent will probably produce what is perceived as appropriate within the framework of the topic, situation, and audience. Examiner misinterpretation, misunderstanding, or rejection of the individual's output can lead to faulty assessment of language growth and development.

Directions or format bias

Tests and other assessment procedures are frequently conducted in much the same manner as play or classroom activities. Familiarity with the framework of the assessment procedures is an advantage for children who find the testing framework consistent with their educational or leisure-time routines. Unfamiliarity with the testing framework can be a serious detriment.

Because different cultural groups have different child-rearing practices, it is reasonable to suspect that some tests are biased against some groups of children because the tests contain faulty presuppositions about the appropriateness of certain directions or assessment formats. For example, a test that requires children to tell stories about abstract or imaginary topics that have little or no relationship to their everyday lives may discriminate against many Chicano children, who operate with a more field-dependent orientation to cognition than other children do (Ramirez & Castaneda, 1974). These children might remain silent or produce very little language or atypically short sentences in the test situation, which could cause the examiner to underestimate their linguistic skills.

The issue of bias in test directions is a more complex matter, since in addition to unfamiliarity with the testing situation, directions involve such linguistic factors as content, length, and syntactic complexity. Syntactic complexity is frequently observed in multiple-choice questions. Consider, for example, the following direction: "None of the following is true, except..." Many test takers can be confused by the syntactic construction of this direction because it requires decoding of a sentence containing an exception to a negative case. The result is that the test taker is unsure whether the direction requires
the selection of an alternative that is true or one that is false.

Value bias

Value bias can occur when a respondent is required to indicate a preference for a set of stimuli or is requested to indicate what a person should do in certain situations. For example, in the Wechsler Intelligence Scale for Children-Revised, one item requires the subject to indicate what should be done if one finds a stamped, sealed, and addressed letter in the street. This item reflects an assessment of knowledge of or belief in a certain ethic, rather than an assessment of cognitive growth.

Even if presumed to be culturally fair, timed tests may discriminate unfairly against children of cultures that emphasize contemplation as a preferred mode of thinking such as many Native American cultures. Or a group play activity that involves friendly individual competition may discriminate against field-dependent children, who typically prefer noncompetitive group activity. In general, it can be asserted that test items are discriminatory when a correct response requires knowledge or acceptance of a value that may be unfamiliar or unacceptable to a respondent.

Linguistic bias

The best literature on test bias focuses on measures that discriminate in the area of speech and language function. Bias is most obvious in tests that purported to assess linguistic development but, in reality, assess knowledge of the phonological, semantic, syntactic, or pragmatic rules of a given language. Thus, the examination of a child for proficiency in English when his or her first or preferred language is Vietnamese or Spanish is obviously discriminatory.

A more complicated problem is presented when the respondent speaks a dialect of English that is reflective of his or her native geographic region, cultural group, or home language. Because dialects are often associated with identifiable racial, cultural, or social groups, it is sometimes mistakenly presumed that a client who is a member of any one of these groups communicates only according to the rules of the group. For example, it is often presumed that a black American from a black inner city community speaks what is commonly known as Black Vernacular English. This presumption may or may not be true. In fact, the subject may speak one of the many variations of languages spoken within the black American community.

Thus, bias can occur when the examiner thinks that the assessment procedure has to be altered to take into account the presumed dialect of the client. Notwithstanding individual differences, there are many social dialects spoken in the United States. Taylor (1965) reports that these dialects are associated with a number of cultural and historical factors, and that some variations within dialects are associated with such social factors as region, gender, socioeconomic class, education, and peer group associations. Dialects are not isomorphic with race; rather, they emerge as a result of an intricate interplay among political, social, cultural, historical, economic, and educational factors, even within races.
MINIMIZING CULTURAL BIAS IN ASSESSMENT PROCEDURES

Harbin and Brantley (1976) have developed a comprehensive model for identifying and minimizing potential sources of child evaluation bias. This model is presented in Figure 1. Although this model was designed for general education and clinical evaluations, it is appropriate for speech and language evaluation.

The model is especially useful because it focuses on every phase of the evaluation process that might be subject to bias: (a) the referral source, (b) the examiner, (c) the tests and other assessment procedures, (d) the interpretation of the child’s performance, and (e) the placement decision.

The reader is encouraged to carefully examine each of the issues presented in the seven steps of the Harbin and Brantley model to determine if cultural biases are present in any of his or her evaluation procedures.

EXAMINATION OF CULTURAL VARIATIONS

An effort to reduce cultural bias in tests must begin with the examiner, irrespective of the subject’s cultural group membership and even when the examiner and child are of the same general cultural group because there may be intracultural variations that are due to such factors as age, gender, region, and socioeconomic class. Saville-Troike (1977) and Watson, Omark, Grovell, and Heller (1980) are among those who support an ethnomethodological approach to the elicitation and interpretation of language samples. Relying heavily on theory and data from cultural anthropology, these authors assert that as a prerequisite for addressing human needs, the speech-language pathologists must become familiar with many aspects of the culture of the populations they serve. Indeed, clinicians should examine their own cultures with respect to their expectations.

The following is a list of some of the topics the clinician might seek knowledge about for a given cultural group:

- cultural values;
- preferred modes of communication;
- nonverbal communication rules;
- rules of communication interaction;
- Who communicates with whom? When? Under what conditions? For what purposes?
- child-rearing practices;
- rituals and traditions;
- perceptions of punishment and reward;
- What is play? Fun? Humorous?
- social stratification and homogeneity of the culture;
- rules of interaction with nonmembers of the culture;
- preferred forms of address;
- preferred teaching and learning styles;
- definitions of handicapped in general and communicatively handicapped in particular; and
- taboo topics and activities, insults, and offensive behavior.

Test modification

Although many clinicians are reluctant to alter standardized test procedures, there are some professionally ethical techniques that can be employed to modify
Although many clinicians are reluctant to alter standardized test procedures, there are some professionally ethical techniques that can be employed to modify these procedures so as to minimize bias.

these procedures to minimize bias. Before any test is administered, the clinician might complete a checklist similar to the one presented in the boxed material to determine whether an assessment instrument is potentially biased against persons from a specific cultural group. If the answer to any question in the checklist is no, appropriate corrective remedies should be instituted prior to the administration of the test.

The clinician should also determine the average mental or chronological age equivalency for each test item for subsequent use in determining the likely effect of potentially biased items on a respondent’s total and subtest scores. For example, if the score on a vocabulary test increases by approximately 3 months for each correct response, eight potentially biased items might reduce a respondent’s score by 54 months.

In conducting item analyses, the clinician should also consider the appropriate language or dialect for the cultural group in question. One way to determine whether a test is biased against dialect speakers is to conduct an item analysis of the instrument in which the phonological, semantic, syntactic, and pragmatic as-

Checklist for Determination of Potential Discrimination of an Assessment Instrument

1. Do I know the specific purpose for which this test was designed?
2. Has the test been validated for this purpose?
3. Are any limitations of the test described in the manual?
4. Do I know the specific information about the group on whom the test was standardized (socio-cultural, sex, age, etc.)?
5. Are the characteristics of the student being tested comparable to those in the standardization sample?
6. Does the test manual or research literature (or my own experience) indicate any differences in test performance across cultural groups?
7. Do test items take into account differences in values or adaptive behaviors?
8. Does the test use vocabulary that is cultural, regional, colloquial, or archaic?
9. Does the test rely too much on receptive and expressive standard English language to measure abilities other than language?
10. Is an equivalent form of the test available in any other language?
11. Am I aware of what the test demands of (or assumes about) the students in terms of:
   - reading level of questions or directions;
   - speed of response;
   - style of problem solving;
   - "test-taking" behavior, and
   - format?
12. Will students with specific physical or sensory handicap be penalized by this test or by certain items?
13. Has an item-by-item analysis been made of the test item by item to determine the linguistic and communicative features of the group for which it is to be used?
sumontions of normalcy are compared with the linguistic assumptions of the client’s home community.

Wolfram, Williams, and Taylor (1972) and Wolfram (1976) have offered a model for conducting item analysis of these types. Using Black Vernacular English and Appalachian English as a base, Wolfram analyzed items from the Illinois Test of Psycholinguistic Abilities (ITPA) Grammatical Closure subtest to demonstrate differences between the correct response for each item and the alternate correct response for speakers from each of these dialect communities (see Wolfram, 1979, Table 1, p. 11). Wolfram et al. (1972) did a similar analysis for several other frequently given tests of articulation, auditory discrimination, and grammatical acquisition.

Once potential sources of bias have been identified, the clinician has four alternatives: (a) Change the stimuli into a parallel form that is likely to be appropriate for the cultural or linguistic group, (b) change the scoring to permit dialect alternatives to be considered correct, (c) multiply the number of potentially biased items by the average age equivalency of the items and add the product to the score obtained by persons from the cultural group, or (d) establish new test norms for the targeted population by obtaining typical response profiles and scores from random samples of normal persons of various age groups in the targeted population. Alternatives a and b are perhaps the easiest and most feasible choices for the clinician; however, Alternatives c and especially d offer the most dramatic opportunities for effecting change.

Naturalistic observations

Perhaps the best alternative to standardized tests as an approach to collecting speech and language data from an individual is the use of focal child assessment techniques with individual children and scan sampling techniques with groups of children (Watson et al., 1980). The advantage of these techniques is obvious: They offer opportunities for collecting data with fewer presuppositional constraints. They also permit the demonstration of code-switching behavior (i.e., the conscious or unconscious art of using different rules of language behavior as appropriate for specific roles, contexts, or audiences). These approaches have become increasingly popular with the escalation of interest in pragmatic and semantic analyses.

In conducting naturalistic observations, the clinician should take steps to ensure that the settings, tasks, protocol, participants, and so forth, used in the observational situation are compatible with the communicative and interactional rules of the person being observed. Also, and perhaps most important, naturalistic settings, which may include participants, must be sufficiently varied to permit the individual to use the widest possible range of his or her communicative repertoire. Analysis of data derived from these samples should include linguistic, pragmatic, and contextual parameters and, where possible, be corroborated by members of the respondent’s speech community.

Payne (1982) has conducted some preliminary research in the Virgin Islands that demonstrates the advantage of using parents and other members of the speech
community to corroborate diagnostic judgments of professional speech-language pathologists. In addition, her findings suggest that speech-community members are a valuable source of information on communicative norms and perceptions of disordered communication.

Payne states that the clinician should visit the client's speech community to acquire information that would be helpful in the diagnosis and treatment of speech and language disorders. The objectives of this visit should include:

* gaining a deeper insight into dialectal and stylistic patterns of variation that are within the range of acceptable speech and language behavior;
* finding differences in speech-community norms that may foster different perceptions of speech and language disorders from those held by the clinician; and
* discovering language functions and language uses that may be useful as elicitation techniques in assessment as well as in therapy.

Other assessment procedures

Seymour and Miller-Jones (1981) have offered some additional considerations for resolving bias problems in assessment. From the reference point of black-English-speaking children, the authors cite a large and impressive body of literature to support the notions that (a) a bidialectal referent should be used as the standard with which the language of these children is evaluated, (b) the children's cognitive styles and orientations should be considered in the assessment of language; and (c) standardized tests should not be used for diagnostic purposes in the assessment of language and cognition.

As an alternative assessment strategy, obtaining language samples in naturalistic settings is supported by Seymour and Miller-Jones (1981). In addition, the authors make a strong case for the use of criterion-referenced tests that use no psychometric norms. According to Glover (1963), a criterion-referenced test translates the test score into a statement about the behavior of an individual. For example, a content-reference scale, which is a type of criterion reference, summarizes the areas of the examination in which the subject showed strengths and weaknesses. Seymour and Miller-Jones state that the criterion of language tests of this kind should be the sequential order of language and cognitive development within the subject's own culture and linguistic group.

To date, there are no commercially developed criterion-referenced language tests available to speech-language pathologists; however, a strong case can be made that none are necessary, given the wide range of linguistic and communicative behavior that exists in the United States as well as within many ethnic and cultural groups. For example, a criterion-referenced test for black children that uses the linguistic behaviors of inner-city Washington, D.C., children as a point of reference may discriminate against black children from another part of the country or from another social class.

An individual clinician might use the criterion-referenced approach to assessment by developing his or her own in-
house instrument. To do this, the clinician might conduct a survey of the linguistic behaviors of normal individuals at a number of age levels, including adults, from the region or cultural groups that represent the clinical population. These behaviors could then be used as the point of reference for assessing the clinical population through the use of tasks designed to elicit these behaviors. Or the clinician might see the client who is receiving treatment as his or her own point of reference by comparing the client's periodic performance on tasks that reflect the linguistic and communicative behaviors of the client's community.

* * *

Many of the approaches advanced to resolve the problem of test bias involve compensatory strategies within the total assessment situation. These strategies have been called "proactive" because they propose constructive solutions to a real clinical problem. These strategies are just first steps that have been proposed to stimulate more thinking and research on the topics discussed.

FOOTNOTES

2. Diana v. California State Board of Education, C-79-37
   RFP (N. D. Cal., Jan. 7, 1979, and June 18, 1979).

REFERENCES

Halliday, M. A. Learning how to mean. In E.H. Lenzoberg


