

New York State Association for Bilingual Education Journal v10 p6-12, Summer 1995**History of Bilingual Assessment and its Impact on Best Practices Used Today****Zina Yzquierdo Mclean**

Abstract: The population of minority students in the educational system is steadily increasing. The disproportionality of English language learners (ELL's) in special education programs is a concern to many educators across the country (LaCelle-Peterson, & Rivera, 1994). This article reviews the historical implications of bilingual assessment from the 1920's to the present. Contributions to the educational assessment of students from culturally and linguistically diverse students is examined. Traditional standardized tests may inadequately measure an English language learner's cognitive ability, thus making the child eligible for special education services (Hernandez-Chavez, Burt, & Dulay, 1978). Federal and state laws that protect culturally and linguistically diverse students are discussed. Standard procedures used for assessing students from culturally and linguistically diverse students is presented.

History of Assessing Bilingual Students

Educational placements and entitlements for students have been determined by standardized testing for over 100 years (Tyack, 1974). Educational concerns of students from culturally and linguistically diverse (CLD) backgrounds appeared to be unimportant during the turn of the century. During this historically formative period of education in the United States, debates about a national language ensued. As the number of students increased and the limits of public budgets decreased, testing was used to group students for differential instruction based on the relatively meager training of teachers. Intelligence tests were widely used as a measure of educational input to sort pupils so they could be efficiently educated according to their future roles in society (Cremlin, 1961). Testing proved a convenient instrument of social control for superintendents in the late 19th century who sought to use tests as a mean for creating the "one best system" of education (Tyack, 1974). Proponents of IQ tests argued that the use of testing as a tool for tracking students would enhance social justice.

During the 1920's, the era of mass testing among various groups of people began. In 1921, Yerkes convinced army leaders of the importance of creating an intelligence test to categorize the militia and distinguish the "simpletons" from the "cream of the crop" (Gould, 1993). He and a group of psychologists created two intelligence tests. The Scale Alpha for men who were literate and the Scale Beta for men who were illiterate. Yerkes and his colleagues proceeded to test over 1.75 million men (Tyack, 1974). Poor testing conditions consisted of overcrowded rooms, poor lighting, and large rooms with poor audibility. The men tested were diagnosed with an average mental age of 13 years. Many immigrants who took the test were disregarded because they did not speak English and were unfamiliar with the American culture. However, some men who were illiterate were administered the reading section of the test and were given a score of zero because of their inability to read. Despite this inappropriate use of tests these men were not given an opportunity to retake the appropriate test for nonreaders. Consequently, Yerkes categorized many immigrants as "simpletons" for receiving zeros on the test because they were unable to speak English, could not read, and were unfamiliar with the American culture (Tyack, 1974).

Low scores for southeastern Europeans on the Army Alpha and Beta tests confirmed White Anglo Saxon

Protestant beliefs that immigrants were inferior, giving powerful arguments to those congressmen who voted to discriminate against them in the immigration-restriction laws of the 1920's (Tyack, 1974). The Restriction Act of 1924, imposed harsh quotas against immigrants. Data from the Army's tests were used to "prove" that Jews, Hungarians, Italians, and Russians were "feeble-minded", and therefore, justified the lower immigration quotas for those groups (Kamin, 1974). "America must be kept American", said Calvin Coolidge, as he signed the bill allowing quotas for various immigration groups (Tyack, 1974). Only 2% of people from the southern and eastern areas of Europe were allowed into the United States in contrast to the high percentages from the British Isles and Northwestern Europe.

Regardless of the improper use of the Alpha and Beta tests, these assessment tools were the first written IQ tests in the United States to gain respect. They implemented hereditarian ideologies of testing and ranking for children. Hereditarians explained the differences in intelligence among groups were due to inherited intellectual inferiority (Herrstein & Murray, 1994). It is interesting to note this belief was contrary to the opinion of Binét who proclaimed that intelligence was too complex to be measured with a single number (Gould, 1993). Binét and Theodore (1916) argued that intelligence could be augmented by good education and was neither fixed nor an inborn quality.

The popularity of Yerkes's assessments grew into the mid-1920's. Yerkes received a steady stream of requests from educational institutions for the use of the army methods of psychological testing for assessing children (Tyack, 1974). Mass testing of students in educational systems erupted. The army test was revised for school use and 400,000 copies of the "National Intelligence Test" were sold in six months. Early advocates of IQ tests felt they should be used to segregate students by ability, to aid in vocational guidance, to detect unusually able or "retarded" students and to diagnose learning problems (Tyack, 1974).

The National Association of Directors of Educational Research during the mid-1920's observed that educators quickly utilized the new group intelligence tests as a means of sorting children for instruction. The same tests were frequently used to compare the efficiency of teachers or of the school systems than to guide changes in the curriculum (Tyack, 1974). The tests were also useful to justify a child's poor grade to irate parents. In this situation, the school personnel showed parents the test results which were considered irrefutable.

Disturbing conclusions were drawn from the results of intelligence testing. In an Arizona copper mining city in 1922, the superintendent was disturbed by the students' inability to learn and the high rate of retardation of pupils among the school district (Tyack, 1974). Tupper, the superintendent, hired an assessment evaluator from Stanford to administer group intelligence tests to children across grades 2-8. The evaluator found that the children were performing at their mental level. Half of the pupils were from a Mexican background, therefore, the evaluator determined students' low functioning ability did not lie in the teachers' manner of teaching, but in the students' genes. The solution created was to develop a special vocational curriculum for the Mexican students in segregated classrooms (Tyack, 1974).

As the 1920's decade came to a close, measuring intellectual capacity became part of the conventional wisdom of not only educational staff and administration, but of the general public (Tyack, 1974). The use of intelligence tests was widely accepted. Results from these evaluations were considered irrefutable and complete.

Public school administrators began to question homogenous groupings of students in schools during the 1930's. Nevertheless, in 1932 three-fourths of 150 large cities across the United States used IQ tests for ability grouping. Intelligence tests established the mental ability of the pupil and resulted in a placement where the pupil belonged (Tyack, 1974).

By the 1940's, testers recognized that standardized tests presented special problems for children from foreign language backgrounds (Havighurst, 1948). In addition, indicators became evident that middle class children tended to stay in school longer than immigrant lower class children which directly impacted their

performance on intelligence tests. The testers in the 1940's felt that identifying situations that were shared among all cultural groups would provide the basis for measurement of comparative abilities (Davis, 1948). Evaluators began to warn that their tests would not be fair to children from foreign language backgrounds (Havighurst, 1948). Test designers believed that American children had certain common experiences because they shared a common American culture. Indeed test experts believed that all American children could recognize the American flag, apple pie, basketball and the current movie stars. Nonetheless, evaluators felt that a test which included naming apple pie, basketball and movie stars would be too easy. The challenge remained in design ing tests that used difficult words and categories and remained part of the experiences of all cultures. Test designers attempted to select items that concerned practices common to the cultural and socioeconomic groups. Davis (1948) reported that evaluators tried to control or equalize the cultural factors in test problems legitimately compare mental abilities of different individuals. Therefore, test developers recognized that common symbols for all socioeconomic and cultural groups would have to be utilized, but the method to achieve this culture free test still remained.

Federal and state cases interceded in protecting the rights of CLD students, requiring these students to receive culturally bias-free tests. During the mid-1950's, the case of *Brown v. Board of Education of Topeka* (1954) played a major role in the passage of the Title VII Civil Rights Act of 1964 and the Education for All Handicapped Children Act (EAHCA, 1975). *Brown v. Board of Education* began an intense inquiry into the process of providing equal opportunities for all students across racial boundaries. School desegregation and civil rights outlawed socially segregated systems of public education as inherently unfair and unconstitutional. The 1960's brought the passage of Title VII of the Civil Rights Act of 1964 and banned discrimination on the grounds of race, color, or national origin in any program or activity receiving federal financial assistance.

The 1970's introduced major court cases that influenced the testing of culturally diverse students. In the court case *Diana v. California State Board of Education* (1970), nine Mexican-American children used Spanish as their primary language. These students were placed in special education programs after receiving assessments that were conducted in English. The court ruled the school districts of California were to test children in their primary language and to use non-verbal tests as well as extensive supportive data. The *Lau v. Nichols* (1974) litigation was initiated on behalf of 1,800 Chinese students in California. The state segregated Chinese students into separate "Oriental" English only schools (Baron, 1991). The Civil Rights movement had done little to improve the linguistic education of California's language minority students, who continued to be ignored by the administrators and teachers (Baca & Cervantes, 1989). The *Lau v. Nichols* Supreme Court decision ordered schools to provide education for all students, whether or not they spoke English. The courts indicated that equality of treatment did not exist merely by providing students with the same facilities, text books, teachers, and curriculum, for students who do not understand English effectively. The decision did not expressly endorse bilingual education, but it did legitimize and give impetus to the movement for equal educational opportunities for non-English speakers (Baca & Cervantes, 1989). Fortunately, the case raised public consciousness about the need for equitable education for CLD students.

Factors that Impacted Current Assessment Practices

The American Education Research Association, American Psychological Association, and the American Council on Measurements developed *The Standards for Educational and Psychological Testing* in 1985, which set clear and explicit regulations concerning bilingual testing. Some of the standards set by the Educational and Psychological Testing are as follows:

1. English tests are confounding for bilingual students, therefore, alternative forms of testing must be elicited.
2. Language background must be taken into account for all phases of assessment.

3. Tests developed without accounting for language differences have limited validity.
4. Cognitive measures do not translate from one language to another.
5. Ability to speak English in a naturalistic situation may not predict ability to learn academic material in English.
6. Assessment of non-native speakers will take more time and observations.
7. Special training for bilingual communication in testing is profitable and beneficial.

The Testing and Educational Standards advocate that evaluators become familiar with the adaptations necessary to provide an equitable testing situation for CLD students. Neglect of these factors may result in an invalid and conceivably harmful set of evaluative recommendations for CLD students in special education (Figueroa, 1990).

The high incidence of students from minority groups in special education may be due to ineffective use of assessment materials. Assessment tools designed for one target audience cannot be transported directly to another cultural setting; in other words, there are no pure "culture fair" or "culture blind" materials (Puckett & Black, 1994). Poor performance on tests is explained by a number of factors including bias of the test, the students' lack of test-taking skills, and their inability to comprehend the information requested on the test because of limited proficiency in English (Puckett & Black, 1994). Content bias has been a long-standing criticism of standardized tests of intelligence for culturally different groups in the United States. Samuda, Kong, Cummins, Pascual-Leone & Lewis (1989) explained that relying on content validity of a test is not sufficient. Content validity measures whether or not the test assesses the objectives that it was designed to measure. Evaluators assume that students have been exposed to the information the test measures.

Problem areas may be identified in the English language learner (ELL); however, these problems may not indicate a language-learning impairment (Hamayan & Damico, 1991). Many students exhibit language or communicative difficulties as a normal aspect of the process of second language acquisition, resulting in misdiagnosis of speech, language, or learning disorders (Cummins, 1984; Kayser, 1987). Children from culturally diverse backgrounds may have developed cognitive competencies substantively different from their more affluent peers from the mainstream culture. Exploration of the sociocultural context of the child's development may yield important insights regarding specific aspects of the environment (Armour-Thomas, 1992). Culture provides the substance and content for attitudes, thought, and action; it allows for a culture-specific idiosyncratic representation of knowledge among its peoples; it determines the kinds of cognitive strategies and learning modes that individuals use for solving complex problems within their society.

Using "Best Practices" in Assessing Students from Culturally and Linguistically Diverse Populations

Assessments of English language learners (ELL) must utilize a different set of evaluative criteria than that used for monolingual English speakers (Hamayan & Damico, 1991). The evaluator must take into account such distinct obstacles as cultural and interactive differences that may make the CLD student appear to have a disability (Westby, 1985). Evaluators must take into account the cultural and linguistic variables during the assessment process. The more familiar and culturally appropriate the content of the stimulus materials, the more likely children demonstrate behavior that accurately represents their real abilities (Armour-Thomas, 1992).

Standard Procedures used in Assessing Bilingual Students

New forms of assessment must be developed based on theoretically defined criteria of cognitive competence derived from the cultural perspective of the individuals to be tested (Armour-Thomas, 1992).

Assessment procedures must uncover performance differences caused by process inefficiencies or inaccuracies arising from impoverished or unstable knowledge representations. Procedures that allow probing of responses and feedback provide opportunity for response justification by the student. This enables diagnosticians to distinguish more readily differences caused by problems in cognition from those developing out of sociolinguistic incompatibilities between the examiner and examinee. Armour-Thomas(1992) recommends core academic requirements, including innovative theories of intelligence and courses on cross-cultural counseling for diagnosticians. Observation and structured interview training should be provided for diagnosticians to use qualitative assessment procedures. Qualitative information may be used to compare language use in various classroom situations, to indicate settings in which the student uses the first or second language, and to observe the degree of communication skills in each setting (Mehan, 1981). Culturally relevant data should be provided in report writing. Fieldwork and internship experiences in culturally diverse settings are a necessary component for personnel involved in the initial screening, assessment, and placement of LEP students. Consultative services must be provided for teachers that emphasize the instructional relevance of students' strengths and weaknesses. Schools with LEP students should designate one or more bilingual specialists to participate in all staffing decisions for CLD exceptional students (Ochoa, Pacheco, & Omark 1983). LaCelle-Peterson and Rivera (1994) emphasizes that assessments must be equitable, test procedures must not penalize students who are unfamiliar with the test language, items on the test must be scrutinized for gender, cultural or other bias. Validity is verified when the norming population of the test is similar to the student receiving the test.

Equitable accountability of assessment depends on the background knowledge of the examiner, a flexible assessment system. Time adjustments allowed for a task, varying the mode of expression (e.g. written or oral), and providing measures of the student's proficiency in the native language of instruction are examples of some adaptations that may be made (LaCelle-Peterson & Rivera, 1994). Information on language learning should also include measures of the student's proficiency in academic and social uses of English.

Darling-Hammond (1994) specified that the evaluator must be familiar with the language and culture of the LEP student. Family and health information should be obtained in the native language and reflect on the home culture norms. If appropriate, testing should occur in English as well as in the native language. Formal assessments should be adapted and documentation of adaptations should be made by the diagnostician. Standardized scoring procedures should be used only when no adaptations are made on the test. The amount of time students are exposed to the English language is important for the evaluation of the bilingual student (Figuroa, 1990). Curriculum-based assessment may be used where students are tested over material that is taken directly from classroom texts or presented in the classroom (Kauffman, 1993).

Diagnostic procedures must uncover performance differences due to process inefficiencies or inaccuracies from differences due to impoverished or unstable knowledge representations within a given domain of knowledge. Assessment procedures must be sensitive to the sociolinguistic patterns that LEP children bring to the testing situation. The ideas that intelligence comprises multiple qualities suggest that multiple measures and contexts would be required to assess intelligence and the various experiential realms in which these qualities develop (Gardner, 1991).

"Best practices" are ensured by law. The Bilingual Education Act mandates that schools conduct language screenings in both the first and second language of all new limited English proficient (LEP) students at the beginning of each school year to determine their need for language related services. Unfortunately, in many districts, such best practices are not monitored.

Summary

In summary, historical conditions determined assessment procedures that formed homogenous grouping for students based on their language limitations. Yerkes revised the Army test he developed in the 1920's for school use. Early advocates of IQ tests felt that testing should be used to segregate students by ability, to aid in vocational guidance, to detect unusually able or "retarded" students and to diagnose learning problems

(Tyack, 1974).

By the 1950's, federal and state laws ruled that discriminatory testing for LEP students was unconstitutional. Laws ensured that CLD students must receive bias-free tests. The Civil Rights movement banned discrimination on the grounds of race, color, or national origin. The 1970's brought major court cases such as Diana v. California State Board of Education (1970), and Lau v. Nichols (1974) which influenced the testing of culturally diverse students.

Assessing LEP students occurs when students' abilities are examined according to their individual capability of using their array of cognitive potentials in different domains of knowledge (Gardner, 1991). LaCelle-Peterson (1994) emphasizes that assessments must be equitable, and test procedures must not penalize students who are unfamiliar with the test language. Items on the test must be scrutinized for gender, cultural or other bias.

Author Recommendations

The population of minority students in the educational system continues to increase. Resolutions to effective assessment of minority students in special education programs must be alleviated. School districts that provide language screening to students with culturally and linguistically diverse backgrounds must use caution when identifying student levels of language proficiency. School districts must develop a fail-safe intervention that provides equitable testing for all students. Students may be identified as English proficient because they have adequate social interaction skills. However, these students may have a distinct lower proficiency when academic ability is measured. Referring teams must be cautioned when recommending standard English assessments. Students may speak English but have cultural considerations that require adaptations to the assessment process.

References

American Educational Research Association, American Psychological Association, American Council on Measurement. (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

Armour-Thomas, E. (1992). Intellectual assessment of children from culturally diverse backgrounds. *School Psychology Review*, 21(4), 552-565.

Baca, L., & Cervantes, H. (1989). *The special education bilingual interface*. Columbus, OH: Merrill.

Baron, D. (1991). English in a multicultural America. *Social Policy*, 21(4), 5-14.

Binét, A., & Theodore, S. (1916). *The development of intelligence in children* (Rev. ed.). Baltimore: Williams & Wilkins.

Brown v. Board of Education, 347 U.S. 483 (1954).

Cremlin, L. (1961). *The transformation of the school: Progressivism in American education, 1876 -1957*. New York: Vintage Books.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. San Diego, CA: College Hill Press.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.

Davis, A. (1948). What are some of the basic issues in the relation of intelligence test to cultural

background? In K. Eells, A. Davis R. J. Havighurst, V. E. Herrick, & R. W. Tyler (Eds.), *Intelligence and Cultural Differences* (pp. 22-28). Chicago: University of Chicago Press.

Diana v. California State Board of Education, No. C-70, RFT, (N. D. Cal. 1970).

Education for All Handicapped Children Act of 1975, Pub. L. No. 94-142, § 1400-1420, Stat. 773 (1975).

Figueroa, R. (1990). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 93-106). Washington, DC: National Association of School Psychologists.

Gardner, H. (1991). *The unschooled mind*. New York: Basic Books.

Gould, S. J. (1993). *The mismeasure of man*. New York: W. W. Norton.

Hamayan E. V., & Damico, J. S. (1991). *Limiting bias in the assessment of bilingual students*. Austin, TX: Pro-Ed.

Havighurst, R. J. (1948). What are the cultural differences which may affect performance on intelligence tests? In K. Eells, A. Davis, R. Havighurst, V. E. Herrick & R.W. Tyler (Eds.), *Intelligence and cultural differences* (pp. 16-21). Chicago: University of Chicago Press.

Hernández-Chávez, E., Burt, M., & Dulay, H. (1978). Language dominance proficiency testing: Some general considerations. *NABE Journal*, 3, 41-54.

Herrstein, R. J., & Murray, C. A. (1994). *Bell Curve*. New York: Free Press.

Kamin, L. (1974). *The science and politics of IQ*. New York: John Wiley.

Kauffman, J. M. (1993). *Characteristics of emotional and behavioral disorders of children and youth*. New York: Merrill.

Kayser, H. G. (1987). A study of three Mexican American children labeled language disordered. *NABE Journal*, 12, 1-22.

LaCelle-Peterson, M. W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64(1), 55-75.

Lau v. Nichols, 414 U.S. 563 (1974).

Mehan, H. (1981). Ethnography of bilingual education. In H. Trueba, G. Gunthrie, K. Au (Eds.), *Culture in the bilingual classroom* (pp. 36-55). Rowley, MA: Newbury.

Ochoa, A. M., Pacheco, R., & Omark, D. R. (1983). Addressing the learning disability needs of limited-English proficient students: Beyond language and race issues. *Learning Disability Quarterly*, 6(Fall) 416-423.

Puckett M. B., & Black, J. K. (1994). *Authentic assessment of the young: Celebrating development and learning*. New York: Macmillan.

Samuda R. J., Kong, S.L., Cummins, J., Pascual-Leone, J., & Lewis J. (1989). *Assessment and placement of minority students*. Kingston, Toronto: Intercultural Social Sciences Public.

Title VII of the Civil Rights Act of 1964 42 U.S.C. § 2000e to 2000e-17 (1988 & Supp. III 1991).

Tyack, D. B. (1974). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.

Westby, C. E. (1985, November). *Cultural differences in adult-child interacproficiency testing: Some general considerations: Assessment and intervention implications*. A mini-seminar presented at the American Speech-Language Hearing Association convention. Washington, DC.

[Return to Table of Contents, Vol 10.](#)

The HTML version of this document was prepared by NCBE and posted to the web with the permission of the author/publisher.

[go to HOME PAGE](#)

www.ncela.gwu.edu