Running Head: ALTERNATIVE ASSESSMENT OF CULTURALLY DIVERSE

Alternative Assessment Options for Students with Disabilities who are

Culturally and Linguistically Diverse

L. Elena Kelley

University of Nevada, Reno

Completed for Comprehensive Exams

April 28, 2008

Abstract

This review of the literature addresses the issue of assessing students with disabilities who are culturally and linguistically diverse (CLD). An examination of data showing disproportionate representation of students with disabilities who are CLD establishes a case for using alternative forms of assessment. Problems with some forms of traditional or commonly used assessments are also addressed. A discussion of four types of alternative assessments— comparable standardized assessment, dynamic assessment, curriculum-based assessment, and performance assessment—provides examples of research showing promise for use with students with disabilities who are CLD. Benefits and drawbacks to each of the four types are described in each section. Recommendations and the need for further research are also discussed.

Alternative Assessment Options for Students with Disabilities who are

Culturally and Linguistically Diverse

In 2006, 26% of the U.S. population identified themselves as a non-White race or mixed race. Fourteen point eight percent of the population was identified as members of Hispanic or Latino ethnic groups of any race (U.S. Census Bureau, 2006). These data show increases from both 1990, when only 19.7% of U.S. Census respondents were reportedly from non-White races and 9% from Hispanic ethnic groups, and 2000 in which 24.8% were from non-White races or mixed races, and 12.5% from Hispanic ethnic groups (U.S. Census Bureau, n.d.). It is estimated that 19.7% of all people in the US now speak a language other than English at home (U.S. Census Bureau, 2006). It follows that the children entering U.S. schools come with a myriad of racial, ethnic, and linguistic backgrounds. The National Center for Education Statistics (NCES) Common Core of Data state level statistics for the 2005-2006 data school year showed that 42.5% of all students enrolled in pre-kindergarten through twelfth grade in the US were identified as a race other than White (Institute of Education Sciences & U.S. Department of Education, n.d.). These data represented a 32.6% increase in the number of non-White students from the 1995-96 to 2005-06 school years. That same academic year (2005-06), of the 48.9 million students enrolled in public schools, 8.6% were identified as English language learners (ELL; Institute of Education Sciences & U.S. Department of Education, n.d.).

In addition to the changing racial, ethnic, and linguistic landscape of U.S. schools, students also have varying ability levels, strengths, and needs. For example, in 2005-06 nearly 6.7 million (13.6%) students were identified as having individual education plans (IEP) for special education services (Institute of Education Sciences & U.S. Department of Education, n.d.). This represents a 47% increase in students with IEPs from 1995-96. In the fall of 2002, the

number of infants and toddlers receiving early intervention services contributed an additional 264,893 children to those receiving specialized services (United States Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, 2006).

When looking at the distribution of students in special education who were also ELLs, research by Hopstock and Stephenson (2003) found that during the 2000-01 school year, 12.4% of all students on a national level were in special education. Of these, only 7.9% were ELLs. However, when state level data were examined, the percentage of ELLs ranged from 0.0 to 17.3%. When specific disabilities were considered, ELLs with high-incidence disabilities (i.e., mild mental retardation [MMR], learning disability [LD], and emotional disturbance [ED]) were nationally under-represented in all areas. Data for students from non-White backgrounds show similarly variable trends in overall under and overrepresentation in special education (Chinn & Hughes, 1987; Parrish, 2002). This information suggests that students from racial, ethnic, and linguistic backgrounds that differ from White, native English speakers face difficulties with being accurately assessed and placed in special education (Harry & Klingner, 2006). Because of the need to ensure that all students receive an appropriate education, this review of the literature will examine the complex issue of assessing students from culturally and linguistically diverse (CLD) backgrounds who have disabilities. The purposes of this review are to: (1) establish the need for appropriate assessments for students with disabilities who are also CLD, (2) examine research on the major types of alternative assessments of students with disabilities who are also CLD, and (3) provide a summation of the benefits and drawbacks of the selected alternative assessments.

Limitations

This literature review has two important limitations. First, the concept of culture has no single definition. As such, research pertaining to *cultural diversity* tends to vary in its focus (e.g., race, ethnicity, gender, etc.) depending upon the researchers conducting the study. This concept will be discussed in more detail following this paragraph. Second, while the relatively new practice of response to intervention (RTI) warrants significant attention, especially as it pertains to the identification and placement of students from diverse backgrounds, a thorough discussion of its place in the assessment process is beyond the purposes of this paper. Although some references will be made to RTI throughout, no specific attention is given to it.

Regarding the first limitation, it is important to note that, although data on race, ethnicity, and language can help facilitate an understanding of the level of diversity in the US and in U.S. schools, these descriptors are limited in scope when it comes to the actual breadth of cultural and linguistic differences (Arzubiaga, Artiles, King, & Harris-Murri, 2008). Students being served in U.S. schools can represent a broad spectrum of socioeconomic levels, geographic regions, urban or rural locations, religious backgrounds, learning styles, gender roles, dialectic variations, and native, migrant, or immigrant statuses (National Council for the Accreditation of Teacher Education, 2007). Even among immigrant groups, reasons for coming to the US (e.g., political refugee, economic opportunity, educational opportunity, family, adventure, etc.) can play an important role in a child's educational experiences. Furthermore, the cultural and linguistic diversity of students, even those from similar backgrounds, assumes many forms, from explicit, easily identifiable language differences to subtle, unconscious behavioral expectations.

As a concept, culture represents a complex web of beliefs, values, experiences, and behaviors that are simultaneously embedded and dynamic. In essence, "one's own culture

provides the 'lens' through which we view and bring into focus our world…[and] ways of seeing, thinking, and feeling about the world which in essence define normality for us" (Avruch & Black, 1993, p. 133). Due to its abstract nature, establishing a universal definition of culture in which to guide educational decision-making has proven exceptionally elusive. One result has been a tendency to equate culture with race or ethnicity (Flanagan & Ortiz, 2001). Because cultural diversity consists of a vast array of both student and teacher differences, researchers primarily focus on only a few differences (e.g., race, ethnicity, language, ability, or gender) as they pertain to their particular research question. While this approach allows for an in-depth examination of particular aspects of cultural diversity, it does limit the type of information available, rendering culture in more simplistic terms.

One aspect of culture that educators in particular must also be aware of is that children have unique learning needs. For students who are CLD, this includes practices focused on assisting ELLs in gaining important academic language skills in English (Cummins, 1984, 1991; Echevarria, Vogt, & Short, 2004) and providing a multicultural perspective in the classroom (Banks, 2004; Garcia, 2004; Klingner & Edwards, 2006). In special education, this process is guided by a student's IEP or individualized family service plan (IFSP; U.S. Department of Education, 2004). The IEP or IFSP provides the basic structure for servicing a student's specific learning needs and governs the implementation of techniques used by educators. Students with disabilities who are also CLD should receive combined services that attend to not only their learning needs, but also their individual cultural and linguistic needs (Gersten & Baker, 2003; Mueller, Singer, & Carranza, 2006). While this paper will attempt to address the wider range of cultural and linguistic diversity found among students, it is important to remember that the

studies described herein only include a subset of the actual breadth of culture and language present in today's schools.

<div align="center">The Problem of Disproportionate Representation</div>

Much of the current literature on students with disabilities who are also CLD focuses on the overrepresentation of minority students in special education based on race (Artiles, Rueda, Salazar, & Higareda, 2005; National Organization on Disabilities, 2001). However, the proportion of students with disabilities who are CLD tends to vary greatly based on the level of the report (e.g., region, state, district, or school) and the group targeted (Artiles et al., 2005). In a study of California school districts, Artiles et al. compared four groups of students: ELLs with limited English proficiency (L2), English proficient learners, White learners, and ELLs with both limited native (L1) and limited L2 proficiency. Risk index data from this study suggested that ELLs with both limited L1 and limited L2 were at a higher risk of being placed in the disability categories of mental retardation (MR), speech and language impairments (SL), and LD at both elementary and secondary levels (with the exception of MR at the elementary level for which no data was available).

In a study of states' differences, Parrish (2002) examined different ethnic and racial groups' representations in special education. They also measured the proportion of minorities overall in the entire state. Parrish found that a minority group in a state with a high proportion of that group tended to have a much greater probability of being identified as having MR than in a state with a low proportion of that group. Thus, depending on the level of measurement and the group studied, disproportionate representation can manifest quite differently. Regardless of the degree to which students who are CLD are represented in special education, "most scholars agree that disproportionate representation is a problem" (Artiles et al., 2005, p. 283), a sentiment

widely echoed by others (Artiles & Trent, 1994; Chinn & Hughes, 1987; Gersten & Baker, 2003; Harry & Klingner, 2006; Harry & National Association of State Directors of Special Education, 1994; Hosp & Reschly, 2004; Losen & Orfield, 2002; Reschly, 1997; Rhodes, Ochoa, & Ortiz, 2005; Rueda & Windmueller, 2006).

While it is important to ensure that students are not misplaced in special education, it is equally important to remember that the delay of referral may compound academic difficulties by preventing students from receiving necessary services (Donovan & Cross, 2002; Wagner, Francis, & Morris, 2005). Many factors can influence the referral, placement, and services provided to students who may have disabilities. Historically, determinations of eligibility for students with high-incidence disabilities (i.e., MMR, SL, LD, and ED) have been based on IQ and achievement tests, classroom observations, and behavioral checklists (Donovan & Cross, 2002). The actual assessments used, the fidelity with which they are employed, and the interpretation of results, however, have produced a wide range of eligibility practices that vary dramatically from location to location and student to student (Harry & Klingner, 2006). Once identified as needing special education, continued progress monitoring must occur to ensure students are working towards the goals in their IFSPs or IEPs. Because no legally mandated, single method for monitoring a student's progress exists, techniques used by educators represent an incredibly vast range of assessment practices. Because of this variation in practice, assessment of students with disabilities who are CLD has become a topic of growing interest to researchers in the fields of both special education and English for speakers of other languages (ESOL).

One area of emerging research is on the use of alternative assessments with students with disabilities who are also CLD (Barrera, 2003; Beaumont, De Valenzuela, & Trumbull, 2002; Donovan & Cross, 2002; Hafner & Ulanoff, 1994; Laing & Kamhi, 2003; Maoz, 2000; Notari-

Syverson, Losardo, & Lim, 2003; Saenz & Huer, 2003; Wagner et al., 2005). An important distinction must be made, however, between *alternate* versus *alternative* assessments. *Alternate* assessment is a method of measuring the performance of students who are unable to participate in standard district or state exams (Spinelli, 2006; Thurlow, Elliott, & Ysseldyke, 2003). *Alternative* assessments refer to a battery of both standardized and non-standardized tests used to refer, place, and teach students who may need special education services (Kea, Campbell-Whatley, & Bratton, 2003). Such assessments are typically used to make decisions regarding referral and placement of students, but may also assist in developing individualized instructional programs (Laing & Kamhi, 2003).

Need for Appropriate Assessments

For over thirty years, U.S. laws have governed the rights of individuals with disabilities in education. Court cases such as *Larry P. v. Riles* (1972/1974/1979/1984/1986) and consent decrees such as *Diana v. California State Board of Education* (1970/1973) have provided important legal backing to support the assertion that students should be assessed appropriately for special education. *Larry P.* was the first court case to draw attention to the overrepresentation of African American students in programs for students with mental retardation. This case established the need for modern day diversity sampling during norming procedures for IQ and achievement tests. In addition, *Diana*, attended to the issue of language in testing procedures and instituted the call for linguistically appropriate assessments. These cases inculcated a sense of seriousness in the education community for ensuring that students who are CLD are not erroneously assessed for special education.

The Individuals with Disabilities Improvement Education Act of 2004 (IDEIA; U.S. Department of Education, 2004) continues to provide legal support for the aforementioned

landmark cases. IDEIA requires that all children, regardless of background or special services designation, be provided with a free and appropriate education. Moreover, IDEIA specifies that students deemed eligible for special education services must not be so designated due to culture or language differences or a lack of opportunity to learn. Determining the extent to which English language acquisition and/or culture interact with a student's learning, however, is not easily ascertained. Traditional assessments (i.e., IQ and achievement tests) may not take into account language and culture differences and may misrepresent a student's true abilities, especially when the assessors are unfamiliar with the student's particular cultural or linguistic background (Rhodes et al., 2005). In addition, federal definitions of disabilities, especially high-incidence disabilities, have a great deal of latitude for educators to choose and interpret assessments (U.S. Department of Education, 2004).

The concern about appropriate placement and services for students with disabilities who are also CLD is made even more apparent in the laws on assessing students from diverse populations. IDEIA § 300.304 clearly states that a child must be assessed using a variety of assessment tools and strategies, not one sole assessment should determine eligibility, and the instruments used be technically sound (U.S. Department of Education, 2004). In addition, IDEIA requires the use of instruments that do not have racial or cultural bias as well as ones that are administered in the language in which the student is most proficient. Furthermore, experts in the field of assessment argue that students with limited English proficiency "should be assessed in the language that permits the most valid inferences about the quality of their academic performance" (Thurlow et al., 2003, p. 113). The No Child Left Behind (NCLB; U.S. Department of Education, 2001) mandate to share this information publicly has further motivated

educators and researchers to find assessment techniques that are more congruous with the diversity present in schools today.

Despite the well-publicized laws and statistics on cultural and linguistic diversity in U.S. schools, Rhodes, Ochoa, and Ortiz (2005) are careful to point out that "legal requirements establish minimal standards of practice. They are not aspirational in nature, nor do they provide a sufficient safeguard to ensure appropriate and accurate assessment of each student, even if followed in a prescriptive fashion" (p. 43). That being said, traditional IQ and achievement tests used to measure a student's intelligence and static content knowledge infrequently address cultural and linguistic distinctions. This can be attributed to a myriad of elements including: (1) differences in practice regarding what to assess, (2) use of assessments with questionable validity, (3) linguistic complexity of tests, (4) culturally biased or culturally loaded assessments, and (5) a failure to distinguish between difference and disability.

First, one common problem identified by the research comes from confusion about what to assess. For example, Sideridis (2007) explored the eligibility practices for eight different countries and found a wide range of considerations for learning disabilities. Among other things (e.g., environmental factors, hearing problems, attention problems, self esteem), six countries included socio-emotional factors in their determination of learning disabilities. In the US, IDEIA clearly states that the identification of learning disabilities should not be due to these factors (U.S. Department of Education, 2004). Thus, educators in the US have traditionally used IQ and achievement tests to determine eligibility for special education. However, Schrag (2000) found that the interpretation of these tests differs greatly from state to state. While some states use a standard score, others employ a regression formula. In addition, the amount of discrepancy (i.e., the difference between a child's ability and achievement) that will qualify a child for special

education varies among states. Thus, a student eligible for special education services in one state may not be eligible in another (Donovan & Cross, 2002; Schrag, 2000). While the discrepancy model of identification has dominated the field of special education for over 30 years, the 2004 reauthorization of IDEIA now allows states to choose the use of alternative types of assessment in determining a student's eligibility (U.S. Department of Education, 2004). The flexibility in the assessment process, while ideal in some respects for addressing the needs of a diverse population, can also potentially pose serious problems for families who move from state to state or district to district.

A second factor in the assessment practices of students who are CLD and who may have disabilities is the use of instruments that do not accurately ascertain a student's true content or language knowledge in either English or their native language (Abedi, 2006; MacSwan & Rolstad, 2006). In keeping with the letter of the law, it must be determined that a student's qualification for special education is not predicated on language differences or a lack of opportunity to learn (U.S. Department of Education, 2004). Indeed, the use of such assessments would be inappropriate due to the fact that "their biases make them illegal for use with this population" (Roseberry-McKibbin & O'Hanlon, 2005, p. 180). Cummins (1984) proposed that there are two major types of language proficiencies: basic interpersonal communication skills (BICS) and cognitive academic language proficiency (CALP). Therefore, assessments should provide a measure of both BICS and CALP. However, few language measures actually assess CALP (Roseberry-McKibbin, 2002), an important skill ELLs need for success in school. Of those instruments that do test academic language, some researchers contest the appropriateness of assessing academic knowledge concurrently with language proficiency (Mahoney & MacSwan, 2005).

On a survey of states' and territories' ELL identification processes, Mahoney and MacSwan (2005) found that 14 (out of 52) U.S. states/territories used primary language oral proficiency tests, 16 used primary language reading/writing tests, and 38 and 36 used English oral and reading/writing tests, respectively. Of these, three emerged as most common: the Woodcock-Muñoz Language Survey (WMLS in Spanish and English; cited in Mahoney & MacSwan, 2005), the Language Assessment Scales (LAS in Spanish and English; cited in Mahoney & MacSwan, 2005), and the Idea Proficiency Test (IPT in Spanish and English; cited in Mahoney & MacSwan, 2005). Some researchers attest that tests such as the WMLS accurately assess students' academic language proficiency (Rhodes et al., 2005). Others contest that commonly used language assessments are not only inaccurate, but also theoretically flawed (MacSwan & Rolstad, 2003; MacSwan, Rolstad, & Glass, 2002). Two studies, in particular, exemplify how these assessments may erroneously identify ELLs as non-proficient in either their native language or in English: Pray's (2005) research on commonly used English language proficiency tests and MacSwan and Rolstad's (2006) study of naturalistic language samples versus primary language proficiency tests.

Pray (2005) examined the construct validity of the WMLS (cited in Pray, 2005), the LAS-O English (cited in Pray, 2005), and the IPT English (cited in Pray, 2005) with native English speakers. Results from that study revealed that none of the students assessed demonstrated English fluency according to the WMLS; most (85%) were classified as fluent by the IPT English. All participants received a designation of fluent according to the LAS-O English. These data provide evidence that these assessments lack concurrent validity, and in the cases of the WMLS and IPT English construct validity. If students who are native English

speakers do not test in the fluent range, it is unlikely that non-native speakers will be able to demonstrate fluency on such assessments.

MacSwan and Rolstad (2006) found naturalistic language samples painted a significantly different portrait of students' native Spanish language abilities than either the LAS-Oral-Español or the IPT-Spanish I-Oral. While a detailed analysis of the naturalistic samples portrayed most students as fluent in their native language, both the LAS-Oral-Español and the IPT-Spanish I-Oral classified the majority as less than fluent (73% and 91%, respectively). The results of these standardized tests are in stark contrast to decades of language acquisition research suggesting that children learn language (i.e., structure, system, and use) effortlessly and without instruction (Chomsky, 1965; MacSwan & Rolstad, 2006). Nonetheless, sometimes students who do not score as proficient in either their native language or English are dubbed "non-non" (MacSwan & Rolstad, 2006, p. 2305) to emphasize their supposed lack of proficiency in both languages.

Test results indicating a lack of proficiency in English and/or one's native language can have a significant impact on decisions regarding eligibility and placement into special education (Artiles et al., 2005). Artiles et al. (2005) found that "compared to English Proficient students, ELLs with limited L1 and L2 were three times more likely to be labeled LAS [language and speech impaired] and over four times more likely to be designated LD…" (p. 293). The nature of this problem may in part be due to the inability of assessment tools in accurately assessing native and English language proficiency. When such decisions are based on assessments with questionable validity, the placements assigned to students can be not only flawed, but also illegal if they result in inadequate or inappropriate services.

Third, in addition to issues with construct and concurrent validity, Abedi (2006) also found that when testing students in English, assessment features such as linguistic complexity

(e.g., long phrases, euphemisms, relative clauses) can act as nuisance variables, confounding the results in ways that suggest the students might have learning disabilities when in fact they do not. It is interesting to note that research on changing such features has had mixed results with ELLs (Abedi & Hejri, 2004; August & Hakuta, 1997). This ambiguity in the relevance of various linguistic test features has been understudied for ELLs, and is virtually non-existent for students who are culturally diverse. The lack of empirical, evidence-based research in this area is coupled with the unsettling realization that the body of research on ELLs alone suggests that "there is, in effect, a high likelihood of being diagnosed as LD as a result of being bilingual" (Figueroa, 2005, p. 164).

Fourth, when cultural differences become part of the mix, determining which assessments will accurately represent a student's knowledge, skills, and abilities can be challenging. The theory of *culturally loaded* versus *culturally biased* (Flanagan & Ortiz, 2001) proposes that even a well-normed assessment (i.e., with low cultural bias) carries with it the cultural perspectives of its authors (i.e., high cultural loading). Whether a test purports to measure language, achievement, or IQ, the context within which test items are created can never be wholly separated from the developer's cultural background (Rhodes et al., 2005). What constitutes degrees of intelligence and achievement for one culture may be unharmonious with those of another culture. While some test developers have gone to great lengths to include diverse racial, ethnic, age, and disability groups (McGrew & Woodcock, 2001), tests normed on U.S. census data still tend to have an unintentional bias towards the majority population, namely monolingual, Euro-American students (Saenz & Huer, 2003). As Rhodes et al. (2005) states, "it cannot be overstated that *stratification in the norm sample on the basis of race is not equivalent to stratification on the basis of culture*" (p. 158, emphasis in original).

Two studies that illustrate the concept of cultural loading are Williams, Turkheimer, Schmidt, and Oltmanns's (2005) use of the Padua Inventory (cited in Williams et al., 2005) for assessing obsessive compulsive behaviors and Hagie, Gallipo, and Svien's (2003) study of commonly used assessments with Lakota Sioux children and adolescents. Williams et al. found that there were significant differences between Black and White participants' responses to the Padua Inventory. Based on differential item functioning analysis, the authors suggested that what may constitute an obsessive or compulsive behavior on the Padua Inventory may be a function of racial or cultural preference. Likewise, Hagie et al. found that Lakota Sioux children showed depressed scores for expressive language and adolescents showed sharp declines in scores for technology and discussion of typical emotions. When cultural norms were evaluated, however, the authors found that nonverbal communication was widely used in the community studied, technology was often unavailable, and that children were expected to be quiet out of respect for their elders. Cultural considerations in both studies should preclude a diagnosis of disorder.

A fifth and final point illustrating the need for alternative assessments in the identification and service of students with disabilities who are also CLD relates to the distinction between disability and difference. IQ tests have shown that some racial and ethnic groups routinely score lower than Whites. A study of the Woodcock-Johnson III battery (Woodcock, McGrew, & Mather, 2001) by Edwards and Oakland (2006) found evidence that there are mean IQ differences between African Americans and Caucasian Americans who took the test. This disparity between groups was explained as being "within the expected range given previous research findings of mean IQ differences between ethnic groups" (Edwards & Oakland, 2006, p. 362). The use of such unquestioned beliefs can, as discussed previously, ultimately have

deleterious effects on a student's opportunity to an appropriate education. As Artiles and Trent (1994) point out:

> …the notion of disability is concerned with atypical functioning or educational performance due to biological, psychological, and/or social factors. The level of functioning for individuals with disabilities falls in the lower portion of the normal distribution curve. The notion of disability exists because we have established parameters to judge when a person functions anatomically, physiologically, intellectually, and/or psychosocially within the limits of what is considered typical. On the other hand, cultural diversity is not defined—at least theoretically—by a standard parameter of functioning. Although it is also concerned with the idea of difference, it is not—unlike the disability construct—inherently linked to the notion of deviance. (pp. 424-425)

Alternative Assessments

The need for alternative assessments for students with disabilities who are also CLD comes none too soon as issues of disproportionate representation, discrimination, and accountability continue to plague local education agencies (i.e., school districts; Barrera, 2006; Figueroa, 2005; Harry & Klingner, 2006). Some local educational agencies assert the importance of culturally responsive practices by extending accommodations to not only students receiving special education services, but also ELLs on district and statewide assessments. These accommodations underscore the seriousness of providing inclusive situations with accessible information for students with language and ability differences. Unfortunately, these accommodations do not extend to students on the basis of cultural difference when language and ability are not factors. "Indeed, there are no tests currently available that have norm samples in

which differences in experiential background (i.e., acculturation) have been systematically controlled" (Rhodes et al., 2005, p. 158)

Certain researchers in the fields of special education, ESOL, and multicultural education have recognized the importance of finding appropriate alternative assessments for evaluating ELLs (Barrera, 2003, 2006; Cho, Hudley, & Back, 2003; Laing & Kamhi, 2003). Some have looked at existing standardized tools, such as in Tsai, McClelland, Pratt, and Squires' (2006) study of the 36-Month Ages and Stages Questionnaire (ASQ; Squires, Potter, & Bricker, 1999) with Taiwanese children. Other researchers, such as Barrera (2003, 2006) and Laing and Kamhi (2003), have explored the use of non-standardized methods for monitoring progress and informing instruction for ELLs diagnosed with learning disabilities.

Still others have dedicated themselves to developing practical classroom instruments (Collier, 2001), comprehensive books and question checklists (Collier, 2001; Spinelli, 2006), and manuals such as Rhodes et al.'s (2005) *Assessing Culturally and Linguistically Diverse Students: A Practical Guide*. National organizations have also latched onto the notion that these types of resources are necessary. The Council for Exceptional Children and the National Association for Bilingual Education (2002) co-authored a manual for administrators to help assess students in meaningful, sensitive, and culturally appropriate ways. Collier's (2001) book, *Separating Difference from Disability*, proposed using a combination of instruments to derive data from classroom observations, family interactions, and specific linguistic features of the student's native language.

In addition, this current review of the research on alternative assessments has yielded some empirical studies that have shown promise in assessing students who are CLD who have high-incidence disabilities. A review of six articles (Kea et al., 2003; McCloskey & Athanasiou,

2000; Notari-Syverson et al., 2003; Shang, 1998; Spinelli, 2008; Wagner et al., 2005) and one

book (Spinelli, 2006) on assessment techniques consistently identified the following four major

types of alternative assessments as the most promising models for students who are CLD: (1)

comparable standardized assessments (CSA), (2) dynamic assessments, (3) curriculum-based

assessments, and (4) performance assessments. In addition, database searches of nearly 400

articles on assessment and alternative assessment for students with disabilities, students who are

CLD, and students with disabilities who are also CLD revealed the majority of the research to be

on these four types. The remainder of this paper will discuss these four major types of alternative

assessments as well as provide some of the benefits and drawbacks of each. The Appendix

includes a brief description of each type with some potential uses, benefits and drawbacks.

Comparable Standardized Assessment

As discussed previously, educational diagnosticians have typically relied primarily on

standardized, norm-referenced measures to assess behavioral and educational characteristics of

children (Spinelli, 2006). Comparable standardized assessments provide a similar structure for

testing students who are CLD for potential disabilities. Wagner et al. (2005) asserts that a CSA

"should assess the same domain, at identical levels, and with identical precision" (p. 10). In order

to achieve this, a distinction must be made between what is merely a translated version of the test

as opposed to a comparably normed, culturally appropriate rendition.

Attempts to create culturally and linguistically appropriate assessments have resulted in

various CSAs. Francis and Carlo (cited in Wagner et al., 2005) developed a Spanish version of

Wagner, Torgesen, and Rashotte's Comprehensive Test of Phonological Processing (CTOPP;

Wagner, Torgeson, & Rashotte, 1999). The new version, called the Test of Phonological

Processes in Spanish (TOPP-S), shows promise as a comparable tool for assessing phonological

awareness in native Spanish speakers in kindergarten through third grade. Test development included in-depth analyses of the types and structures of each question of the CTOPP so that comparable phonological domains could be designed for the TOPP-S. Because of the unique phonological structure of each language, simple translations of the CTOPP would have been invalid (Wagner et al., 2005). While still in its validation phase, the TOPP-S yielded a reliability of .83 ($N = 100$), which is approaching Bracken's (1987) proposed acceptable value of .90. When compared to the CTOPP, the correlation on subtests for both versions was .69 ($N = 1000$ [CTOPP], $N = 1000$ [TOPP-S]) indicating a strong relationship for phonological processing assessment between the two.

In another study on CSAs, Tsai et al. (2006) used a translated and back translated version of the 36-Month ASQ (Squires et al., 1999) with Taiwanese preschoolers. In addition to the translation and back translation, the researchers used a panel of experts to examine the translated version for cultural appropriateness. This process revealed that all items appeared sound except for an image depicting a left hand holding scissors. Because of the cultural preference in Taiwan for right-handedness, it was suggested that this image be changed. The researchers also asked parents and teachers who participated in the study to indicate whether or not they felt the assessment was culturally appropriate. Ninety-nine percent agreed that it was. Finally, their sample included both students with no history of developmental delay and those with documented delays. Their data showed significant agreement between parent and teacher measures, with all previously identified students being identified on the modified ASQ. This study provides evidence that this tool appropriately assesses children aged 34 to 38 months.

An important point of consideration from the Tsai et al. (2006) study was that of cultural appropriateness. When students who are different from the majority population are assessed with

CSAs, the fit of the instrument should accommodate for a student's level of acculturation to the majority culture. Acculturation is defined as "the adaptation to a new culture, language, and interaction environment" (Collier, 2001, p. 2). In a study of Korean American adolescents, Cho et al. (2003) examined students' scores on Reynolds and Kamphaus's Self-Report of Personality (SRP) scale of the Behavioral Assessment Scale for Children (BASC; cited in Cho et al., 2003). The researchers compared these scores to students' levels of acculturation on the Suinn-Lew Asian Self Identity Acculturation scale (Suinn, Ahuna, & Khoo, 1992). Results of the comparison showed that the participants' levels of acculturation were not significantly correlated with their mean scores on the SRP except for on the Self-Reliance subscale (i.e., the more assimilated to mainstream U.S. culture the students were, the more confident they felt). Cho et al. posited, however, that this could have been due to the relatively small and homogeneous sample size ($N = 51$). An interesting result of this study showed that after eliminating certain items perceived to contain cultural bias toward Korean American adolescents from the SRP (as identified by the data analysis program used), the data suggested that the SRP yielded valid and reliable results for measuring social and emotional adjustments of Korean American students. As with Tsai et al, this study demonstrated that when adjustments are made for cultural appropriateness, CSAs can act as effective and accurate tools for assessing students who are CLD.

Benefits of CSA. When used as recommended by the publishers, standardized, norm-referenced assessments can provide valid, reliable registers of IQ, achievement, and behavior of the majority culture on which they were normed (Donovan & Cross, 2002). By renorming a test and comparing a student's score to local instead of national norms, these tests may even give practical, quantifiable data about the intellectual or language status of a child who is CLD (Saenz

& Huer, 2003). CSAs that are modified for cultural (Cho et al., 2003; Tsai et al., 2006) and linguistic (Wagner et al., 2005) appropriateness can result in meaningful data that matches existing theories of behavior and language proficiency for specific groups. Used in conjunction with other alternative measures, CSAs add to a compendium of information necessary for making sound decisions regarding placement and services. In addition, translated versions of these tests are widely available (Rhodes et al., 2005).

Drawbacks of CSA. Standardized tests present a norm-referenced comparison group that may not represent the diversity found in the target group (Rhodes et al., 2005; Valdivia, 1999; Wagner et al., 2005). Also, established norms no longer apply when a test is modified (Saenz & Huer, 2003). When CSAs are used in isolation, inappropriately compared to majority norm standards, or as determining factors in decision-making processes, they may prove unreliable when used with students who are CLD. This could be due to multiple factors, including construct bias (MacSwan & Rolstad, 2006), cultural loading (Flanagan & Ortiz, 2001), and/or disproportionate representation of non-majority groups in normalized samples (Laing & Kamhi, 2003; Notari-Syverson et al., 2003). Furthermore, Figueroa (2002) argues that modified versions of standardized tests in another language may not account for the complex array of syntactic, contextual, lexical, or semantic variations.

Dynamic Assessment

Typified by its ability to assess a student's cognitive learning potential as opposed to prior knowledge, the dynamic assessment model (also called mediated learning) has become a popular tool for use with students who are CLD with possible or actual disabilities (Barrera, 2003, 2006; Kea et al., 2003; Laing & Kamhi, 2003; Notari-Syverson et al., 2003). Based on Feuerstein's (1979) Learning Potential Assessment Device, dynamic assessments afford testers

the opportunity to observe a student's learning potential and self-regulatory behaviors (Notari-Syverson et al., 2003). Dynamic assessment focuses on the process of learning, as opposed to just a single response (Missiuna & Samuels, 1989). As described by Laing and Kamhi (2003), three main approaches to dynamic assessment exist that can be used together or independently from each other: test-teach-retest; task/stimulus variability; and graduated prompting. Because most of the literature on dynamic assessment reviewed for this paper fits into the categories of dynamic assessment as outlined by Laing and Kamhi, these same groupings are used here.

Test-teach-retest. This application of dynamic assessment involves a pre-test of student ability in a specific skill such as note-taking (Barrera, 2003, 2006), explicit teaching of the skill, and a post-test of the skill after instruction (Kea et al., 2003; Laing & Kamhi, 2003; Notari-Syverson et al., 2003). Evaluators then examine pre- and post-test scores to determine a child's responsiveness to instruction. Barrera (2003) looked at the note-taking abilities of 38 Mexican American high school students before and after two-weeks of instruction on writing notes in journals. Comparing three groups—bilinguals without disabilities, students with disabilities rated as high limited English proficient (LEP), and students with disabilities rated as low LEP— Barrera (2003) found that ELLs with disabilities scored significantly higher on their post-tests than pre-tests. In addition, they closed the gap between themselves and high achieving bilinguals without disabilities on three out of four measures of note-taking ability (the exception being spelling). Follow-up research with 114 Mexican American students (Barrera, 2006) also provided data that teachers' blind ratings of the notes followed the expected pattern of student groupings and were significant for 12 of 17 possible ratings. These results demonstrated that dynamic assessment can help teachers to reliably identify ELLs with disabilities versus ELLs without disabilities.

Task/stimulus variability. This method of dynamic assessment involves the presentation of tasks embedded in contextualized stimuli (Laing & Kamhi, 2003). In other words, evaluators reformat the presentation of the test material in order to accommodate the sociocultural or linguistic differences of the subjects. For example, Fagundes, Haynes, Haak, and Moran (1998) conducted a study of African American and Caucasian American five-year olds from low socioeconomic backgrounds. Students received testing in the context of thematic activities versus a standardized setting. The findings demonstrated that the African American children performed comparably to Caucasian Americans when test tasks occurred as part of thematic, contextualized settings. Conversely, when presented via the standardized format, the African American children had significantly lower scores than their Caucasian peers, especially as the difficulty levels increased. Moore-Brown, Huerta, and Uranga-Hernandez (2006) further showed how mediated learning situated around the students' particular responses can help identify potential disabilities or learning difficulties. They examined three case studies of Hispanic elementary students and found that specific mediated learning opportunities revealed strengths and weaknesses of students not available from the traditional battery of standardized, norm-referenced tests.

Graduated prompting. Although no recent studies were found on graduated prompting, this technique was mentioned in at least three articles that discussed dynamic assessment (Fuchs, Fuchs, Compton, Bouton, Caffrey, & Hill, 2007; Laing & Kamhi, 2003; Saenz & Huer, 2003). Graduated prompting proposes a tiered approach to eliciting responses from a student (Laing & Kamhi, 2003). This technique most closely resembles computer-based standardized aptitude tests that pose progressively more difficult questions, worth more points, until the participant misses a question. When a question is missed, the difficulty and points gradually decrease until the participant answers correctly or reaches the lowest level of questioning. In a similar vein as these

computerized assessments, graduated prompting allows evaluators to use specifically delineated tiers of prompts to guide students toward the solutions of each question. Increases in prompting result in decreased scores until a zero score level is reached.

Benefits of dynamic assessment. Implementation of dynamic assessments can occur throughout the course of a lesson in any classroom. Barrera (2003, 2006) illustrates this in his studies on curriculum-based dynamic assessment in which note-taking skills were integrated into the current curriculum. Assessing a student's growth before and after instruction should occur naturally in most classrooms as it complies with generally accepted best teaching practices. In addition, the scenario of teaching new information followed by immediate application directly mimics most employment situations, thus rendering this methodology as not only a valid assessment tool for special services in school, but also a potential model for real-world situations. Fuchs et al. (2007) recommend the use of dynamic assessment for schools employing the RTI model of eligibility because of its ability to identify not only students at risk of school failure, but also to provide instructional intervention strategies. Thus, both instructors and evaluators alike partake in the assessment process. Moreover, the focus on process in dynamic assessment helps to remove much of the cultural bias found in traditional standardized assessments (Roseberry-McKibbin & O'Hanlon, 2005).

Drawbacks of dynamic assessment. Not all instructors incorporate pre-tests into their curricula. To do so would propose that these teachers undergo a major philosophical shift that examines gains in learning and skills in contrast to static knowledge. In addition, educators who might feel that it takes too much time (Saenz & Huer, 2003). There is also a lack of research on the validity and reliability of dynamic assessment (Saenz & Huer, 2003). Yet another concern is that Feuerstein's (1979) original model proposed that modifications taught to students should be

contingent on his/her own learning needs, but many models of dynamic assessment involve pre-determined, scripted teaching guides. An early study by Missiuna and Samuels (1989) on 43 preschoolers provided evidence that the method of instruction, contingent versus scripted, made a difference. Children assessed using the contingent method showed significantly higher post-test gains than the scripted group. Their study demonstrated that not all dynamic assessments are created equal. Therefore, when determining which method to use, it is important to consider not only the child being assessed, but also purposes and practical aspects of the assessment.

Curriculum-Based Assessment

Gaining notable popularity among educators, curriculum-based assessments (CBA) "center on measuring students' mastery of goals, objectives, and criteria embedded in the school-adopted curriculum" (Kea et al., 2003, p. 33). Curriculum-based measurement (CBM) is a type of CBA and has emerged from Deno's (1985) research on ways to monitor students' progress and provide teachers help in making instructional decisions (Wiley & Deno, 2005). CBAs are usually non-standardized assessments in which teachers use classroom-based tasks to determine students' capabilities (Barrera, 2006) and instructional needs within a school-adopted curriculum (Kea et al., 2003; Rhodes et al., 2005). Scoring of CBAs is often based on a compilation of work that is measured according to appropriate scoring guidelines (Barrera, 2006). CBM, sometimes called formative assessment, follows a more structured approach than CBA by using standardized administration and scoring. CBM "simultaneously yields information about standing as well as change and about global competence as well as skill-by-skill mastery" (Fuchs & Fuchs, 2002, p. 66). CBMs are scored against a validated set of skills that may or may not be derived from a student's actual curriculum. Criterion-referenced assessment is a particular subset

of CBA and CBM that measures a student's performance as compared to a specific criterion, as opposed to a skill set.

Research on CBA and CBM with students who are culturally and linguistically diverse is a rapidly growing field. Programs such as AIMSweb (Shinn & Shinn, 2002) have documented results from several thousands of students showing that CBM is a valid and reliable tool for improving students' skills in both math and reading (National Center on Student Progress Monitoring Technical Review Committee, 2007). A literature review by Fuchs and Fuchs (2002) described several studies in which CBM data repeatedly showed strong psychometric properties, gains in student learning, evaluative effectiveness for interventions, and ability to identify students who did not respond to instruction.

Results for ELLs have also shown promise. Graves, Plasencia-Peinado, Deno, and Johnson (2005) used CBM with first grade ELLs from multiple language backgrounds who all qualified for free and reduced lunch programs. The purpose of the study was to ascertain how well English proficiency correlated with oral reading fluency. The researchers found that scores on a standardized language proficiency assessment given at the beginning of kindergarten only weakly correlated with CBM data for oral reading fluency at the end of first grade ($N = 134$). While the time gap in this study is clearly a limitation, the results nonetheless suggest that English language skills at the beginning of kindergarten are not necessarily good predictors of reading skill in first grade.

In another study, Wiley and Deno (2005) found that using CBMs to measure oral reading fluency in third and fifth graders significantly predicted state achievement test scores for ELLs. They also found that maze tasks do not significantly predict achievement scores ($N = 69$), even when combined with oral reading fluency. The results of this study, combined with those of

Graves et al. (2005) suggest that using CBMs with ELLs can provide important feedback regarding students' reading abilities. While the Graves et al. and Wiley and Deno studies focused only on ELLs, not necessarily those with disabilities, they nonetheless provided evidence that CBM is a tool that is appropriate for use with students who are linguistically diverse.

Of particular note is a study by Fuchs, Fuchs, and Hamlett (1989) comparing teachers and students in three groups: dynamic goals CBM, static goals CBM, and a control group. Students in the study included 26 minorities, 46 students with LD, 12 students with emotional disturbance, and 2 students labeled educable mentally retarded. Their data showed that teachers in the dynamic goals CBM group increased their goals more frequently and ended up with more ambitious goals at the end of the study. In addition, students in this group achieved better results than students in the static goals CBM group. This study provides an excellent illustration of the potential of combining dynamic assessment properties with CBM.

Benefits of CBA. Both CBA and CBM offer the opportunity to directly apply assessed skills towards instructional goals, apply to the classroom environment, and are virtually unlimited in nature (Rhodes et al., 2005). Assessment can occur as frequently as needed (Wiley & Deno, 2005) and in relation to a specific set of curriculum skills (Notari-Syverson et al., 2003). They supply both qualitative and quantitative data (Shang, 1998), are sensitive to growth (Wiley & Deno, 2005), and may be individualized according to cultural, linguistic, or educational background (Rhodes et al., 2005). Moreover, repeated administrations of alternate forms of CBM allow educators to see a student's progress over time. The standardized nature of CBM has also made it an especially useful tool for schools using the RTI model of eligibility for special education.

Drawbacks of CBA. Drawbacks of both CBA and CBM are the risk of teaching to the test (Rhodes et al., 2005; Shang, 1998), variability among teacher-made assessments, and lack of comparison to curriculum outside of the specific school, district, or state (Rhodes et al., 2005). In addition, published CBMs may utilize standards and skills that do not match a student's actual curriculum (Rhodes et al., 2005). Finally, if the skills being assessed are too contextually removed from the actual curriculum, validity for students with disabilities, especially those who are also CLD may be compromised.

Performance Assessment

This relatively large group of alternative assessments can include, but is by no means limited to, play-based assessment, direct observations, writing samples, ethnographic assessment, collaborative projects, and portfolios. Performance assessment "focuses on the students' abilities to produce a product or otherwise apply classroom learnings to real or simulated situations" (Shang, 1998, p. 270). Sometimes called authentic assessment, performance assessment gives children an opportunity to demonstrate and apply knowledge and can include tasks such as stacking blocks, telling stories, and drawing pictures in the case of young children (Notari-Syverson et al., 2003) or projects, tests, experiments, and portfolios for older students (Im, 2000; Kea et al., 2003). Performance assessments should be a direct measure of learning by eliciting specific behaviors of interest to the instructor (Tombari & Borich, 1999). While many valuable types of performance assessment exist, this section will only describe portfolios and observations as examples. An in-depth discussion of various types of performance assessments can be found in Tombari and Borich (1999).

While research is still greatly lacking on performance assessment for students with disabilities who are also CLD, a few studies describe ways that these tools can be used in the

classroom. One of the more well-documented forms of performance assessment is the use of portfolios. These purposeful collections of a child's work document her progress over time (Kea et al., 2003; Notari-Syverson et al., 2003; Tombari & Borich, 1999) and might include video or audio tapes, anecdotal notes, progress notes, tests, or pictures. While they can include an expansive array of artifacts, portfolio contents should be carefully planned, chosen, and evaluated.

In their study of a middle school English as a second language classroom, Smolen, Newman, Wathen, and Lee (1995) documented a teacher's use of a two-tiered portfolio approach. The teacher's goal with the portfolios was to improve metacognitive reading skills in her students. Students wrote goal statements at the beginning of the week, reflected on them at the end of the week, and fully participated in the collection of and justification for the presence of various artifacts in the portfolio. Qualitative examination of students' work showed that they were not only employing class-negotiated reading strategies, but also understanding the impact involved in choosing to use those strategies.

Gottlieb (1995) takes this approach one step further by proposing a developmental scheme for portfolios called the CRADLE approach. The CRADLE approach stands for: Collecting, Reflecting, Assessing, Documenting, Linking, and Evaluating (see Gottlieb, 1995). In this approach, students and teachers work together to develop a list of portfolio contents that will best reflect students' abilities and learning. Reflection on both individual and class levels is facilitated by the instructor. "The centerpiece of this portfolio type [reflective portfolio] is the students' perceptions, interpretations, and strategies utilized in acquiring knowledge" (Gottlieb, 1995, p. 13). Gottlieb also argues that validity and reliability of portfolios can be established

through the use of goals aligned with curriculum (content and ecological validity) as well as by multiple teacher ratings using a rubric (inter-rater reliability).

Another type of performance assessment, direct observation, is exemplified by Gersten and Baker's (2003) study of their researcher-developed English-Language Learner Classroom Observation Instrument (ELLCOI). The purpose of this tool was to observe classroom reading instruction for evidence of how teachers were addressing the needs of ELLs. Used in first grade classrooms, the results provided evidence that students in classrooms where reading instruction was rated higher had better adjusted reading scores at the end of first grade. A subsequent study by Graves et al. (2005), however, showed that teacher ratings on the ELLCOI did not substantially affect oral reading fluency in first graders. This instrument has obvious potential, but needs further study to determine its usefulness in predicting reading skills.

Benefits of performance assessment. Because performance assessments are "ideally suited to assess knowledge, deep understanding, and problem-solving strategies, they can also be used to assess a learner's work habits and social skills such as cooperation, sharing, and negotiation" (Tombari & Borich, 1999, p. 146). For families in which the parents or caregivers do not speak English, performance assessments such as portfolios allow them to participate in the collection of materials pertinent to the child (Spinelli, 2008). This body of work can then be passed along and contributed to by other teachers in subsequent years, an especially beneficial feature for migrant families who wish to give teachers a quick means for getting to know their child (McCloskey & Athanasiou, 2000; Notari-Syverson et al., 2003). Portfolios potentially offer detailed insight into a student's academic abilities, emotional welfare, interests, and cognitive processes (Tombari & Borich, 1999). Moreover, for students who are CLD and are being

evaluated for potential disabilities, invaluable information from performance assessments can guide the referral process.

Drawbacks of performance assessment. The primary drawback of performance assessments is that "they are difficult to design and even more difficult to evaluate" (Shang, 1998, p. 270). In addition, the lack of reliability and validity is frequently cited as a problem (Saenz & Huer, 2003). In addition, performance assessments such as portfolios can be cumbersome and difficult to maintain and transport, especially when they contain a compilation of several years of work for each student. Thus, performance assessments should be conducted in addition to other forms of assessment when considering students for special education services.

Conclusion

The emphasis on what teachers actually teach imbues any test with a greater sense of relevance. In contrast, a one-size fits all approach often ceases to address the specific cultural and linguistic diversity evident within even the smallest school district. As teachers adjust their curricula to address this, instruction can stray persistently further from the established assessment. The existence of resources supporting multimodal evaluations of diverse learners does not preclude the importance of informed, thoughtful, and collaborative decision making. As Figueroa and Newsome (2006) point out, "failure to use the available corpus of regulatory, professional, and research-based knowledge on how to test bilingual children cannot be assuaged by simply hiring bilingual school psychologists or by using interpreters" (p. 213). Indeed, educators, psychologists, and diagnosticians must be cognizant of the variety with which students come to school (Barrera, 2000; Hagie et al., 2003; Overton, Fielding, & Simonsson, 2004).

Recommendations

Child study teams (CST) provide a successful model for addressing the concerns of students who are CLD who may have disabilities. These teams should incorporate a variety of personnel such as school psychologists, counselors, special education, ESOL, and mainstream educators, administrators, and parents (Rhodes et al., 2005). The collaboration integral to high-functioning CSTs can be difficult to attain, however, when educational professionals lack the training and skills necessary to support service delivery to students with disabilities who are also CLD (Harry, 2002; Roache, Shore, Gouleta, & Butkevich, 2003). Klingner and Artiles (2003) assert that a child study team is only as effective its members. They stress that too many CSTs do not pay due attention to language scores or issues, classroom ecology, intervention strategies, classroom observations, or cultural values. Moreover, research by Roache et al. (2003) and Sutton et al. (2003) states that many teachers feel ill-prepared to contend with sociocultural issues from the onset of their careers, thus precipitating the need for ongoing professional development in culturally responsive teaching. Despite some issues with CSTs, they still offer a valuable method by which to facilitate the appropriate assessment and referral of students who are CLD who may need special education services.

Similarly, Ortiz, Wilkinson, Robertson-Courtney, and Kushner (2006) describe the use of teacher assistance teams (TAT) and student assistance teams (SAT) to provide ongoing, appropriate services to ELLs with disabilities. TATs, consisting of four to six general and specialty area (e.g., ELL, special education) teachers, focus on the teacher requesting assistance. Together, TATs may use assessments to delineate a student's academic difficulties, share information about a child's behaviors exhibited in other situations, or discuss possible instructional or behavioral interventions. An SAT may be a part of a TAT, but tends to follow a

model more similar to that of child study teams. SATs may involve a wide range of professionals and are often instrumental in eligibility decisions. Unlike CSTs, however, SATs continue to meet in order to monitor progress and suggest specialized interventions. In either case, the focus is on team problem solving and providing the most culturally and linguistically appropriate educational experiences for children.

Future Research

The need for further empirical research on students with disabilities who are also CLD is great. In Sutton et al. (2003), the researchers searched Office of Special Education (OSEP) projects for programs focusing on assessments for learners in rural settings, yet another manifestation of cultural diversity. After contacting the project coordinators of several programs, they found only four projects that centered on students in rural areas. A cursory search of current grants (2005-2007) on the OSEP website (2006) revealed 51 sponsored projects that included the search term "English language learner." This offers hope that more researchers are conducting studies to address the needs of at least one group of students with disabilities: ELLs. Sutton et al.'s (2003) findings, however, illustrate the scarcity of OSEP funded projects for specific cultural groups. The lack of studies on some groups (e.g., religion, sexual orientation, and rural versus urban location), coupled with the relatively nascent research addressing concerns related to appropriate assessment, placement, and services of CLD learners suggests a critical need for increased research in this area.

Alternative assessments offer a viable solution for meeting the diverse needs of this nation's changing demographic landscape. Research on specific subgroups of students who are CLD (e.g., ELLs, Korean American, Mexican American, African American, etc.) has shown promise in helping educators to accurately and appropriately assess students' true capabilities.

However, cultural and linguistic diversity encompasses a much broader range of differences than are being addressed in the research. More studies still need to be done on cultural factors that go beyond race, ethnicity, language, and ability. Even among the research that does exist on the assessment of students who are culturally and linguistically diverse, *best practices* is still an ambiguous term, leaving educators to wonder if they are on the right track.

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record, 108*(11), 2282-2303.

Abedi, J., & Hejri, F. (2004). Accommodations for students with limited English proficiency in the national assessment of educational progress. *Applied Measurement in Education, 17*(4), 371-392.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*(3), 283-300.

Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education, 27*(4), 410-437.

Arzubiaga, A. E., Artiles, A. J., King, K. A., & Harris-Murri, N. (2008). Beyond research *on* cultural minorities: Challenges and implications of research as situated cultural practice. *Exceptional Children, 74*(3), 309-327.

August, D., & Hakuta, K. (1997). *Improving schooling for language minority children*. Washington, DC: National Academy Press.

Avruch, K., & Black, P. W. (1993). Conflict resolution in intercultural settings: Problems and prospects. In D. J. Sandole & H. van der Merwe (Eds.), *Conflict resolution theory and practice: Integration and application* (pp. 131-145). New York: Manchester University Press.

Banks, J. A. (2004). Multicultural education: Historical development, dimensions, and practice. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (2nd ed.). San Francisco: John Wiley & Sons.

Barrera, I. (2000). Honoring differences: Essential features of appropriate ECSE services for young children from diverse sociocultural environments. *Young Exceptional Children, 3*(4), 17-24.

Barrera, M. (2003). Curriculum-based dynamic assessment for new- or second-language learners with learning disabilities in secondary education settings. *Assessment for Effective Intervention, 29*(1), 69-84.

Barrera, M. (2006). Roles of definitional and assessment models in the identification of new or second language learners of English for special education. *Journal of Learning Disabilities, 39*(2), 142-156.

Beaumont, C., De Valenzuela, J. S., & Trumbull, E. (2002). Alternative assessment for transitional readers. *Bilingual Research Journal, 26*(2), 241-268.

Bracken, B. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 4*, 313-326.

Chinn, P. C., & Hughes, S. (1987). Representation of minority students in special classes. *Remedial and Special Education (RASE), 8*(4), 41-46.

Cho, S.-J., Hudley, C., & Back, H. J. (2003). Cultural influences on ratings of self-perceived social, emotional, and academic adjustment for Korean American adolescents. *Assessment for Effective Intervention, 29*(1), 3-14.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: The M.I.T. Press.

Collier, C. (2001). *Separating difference from disability*. Ferndale, WA: CrossCultural Developmental Education Services. (ERIC Document Reproduction Service No. ED456667)

Council for Exceptional Children. (2002). *Determining appropriate referrals of English language learners to special education: A self-assessment guide for principals*. Retrieved January 9, 2008, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a /fe/30.pdf

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

Cummins, J. (1991). *Empowering culturally and linguistically diverse students with learning problems*. Reston, VA: Council for Exceptional Children. (ERIC Document Reproduction Service No. ED333622)

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Diana v. State Board of Education, C.A. 70 RFT (N.D. Cal., Feb.3, 1970, 1973). (consent decree).

Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Echevarria, J., Vogt, M., & Short, D. (2004). *Making content comprehensible for English learners: The SIOP model*. Boston: Pearson Education.

Edwards, O., & Oakland, T. (2006). Factorial invariance of Woodcock-Johnson III scores for African American and Caucasian Americans. *Journal of Psychoeducational Assessment, 24*(4), 358-366.

Fagundes, D. D., Haynes, W. O., & Haak, N. J. (1998). Task variability effects on the language test performance of southern lower socioeconomic class African American and Caucasian five-year-olds. *Language, Speech, and Hearing Services in Schools, 29*(3), 148-157.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore, MD: University Park Press.

Figueroa, R. (2002). Toward a new model of assessment. In A. Artiles & A. Ortiz (Eds.), *English language learners with special education needs: Identification, assessment, and instruction* (pp. 51-63). Washington, DC: Center for Applied Linguistics.

Figueroa, R. (2005). Dificultades o desabilidades de aprendizaje? *Learning Disability Quarterly, 28*(2), 163-167.

Figueroa, R. A., & Newsome, P. (2006). The diagnosis of LD in English learners: Is it nondiscriminatory? *Journal of Learning Disabilities, 39*(3), 206-214.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York: Wiley.

Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness to intervention: A scripted protocol to identify young at-risk readers. *Teaching Exceptional Children, 39*(5), 58-63.

Fuchs, L. S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education, 77*(2), 64-84.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*(5), 429-438.

Garcia, E. (2004). Educating Mexican American students: Past treatment and recent developments in theory, research, policy, and practice. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (2nd ed.). San Francisco: John Wiley & Sons.

Gersten, R., & Baker, S. (2003). English-language learners with learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 94-109). New York: The Guildford Press.

Gottlieb, M. (1995). Nurturing student learning through portfolios. *TESOL Journal, 5*(1), 12-14.

Graves, A. W., Plasencia-Peinado, J., Deno, S. L., & Johnson, J. R. (2005). Formatively evaluating the reading progress of first-grade English learners in multiple-language classrooms. *Remedial & Special Education, 26*(4), 215-225.

Hafner, A. L., & Ulanoff, S. H. (1994). Validity issues and concerns for assessing English learners: One district's approach. *Education and Urban Society, 26*, 367-389.

Hagie, M. U., Gallipo, P. L., & Svien, L. (2003). Traditional culture versus traditional assessment for American Indian students: An investigation of potential test item bias. *Assessment for Effective Intervention, 29*(1), 15-26.

Harry, B. (2002). Trends and issues in serving culturally diverse families of children with disabilities. *Journal of Special Education, 36*(3), 131-138, 147.

Harry, B., & Klingner, J. (2006). *Why are so many minority students in special education? Understanding race and disability in schools*. New York: Teachers College Press.

Harry, B. (1994). *The disproportionate representation of minority students in special education: Theories and recommendations.* Alexandria, VA: National Association of State Directors of Special Education.

Hopstock, P. J., & Stephenson, T. G. (2003). *Special topic report #2: Analysis of office for civil rights (OCR) data related to LEP students*. Retrieved September 20, 2007, from http://www.ncela.gwu.edu/resabout/research/descriptivestudyfiles/

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education academic, demographic, and economic predictors. *Exceptional Children, 70*(2), 185-199.

Im, B.-B. (2000). Understanding and application of performance assessment. *Journal of Pan-Pacific Association of Applied Linguistics, 4*(1), 41-72.

Institute of Education Sciences, & United States Department of Education. (n.d.). *National center for education statistics*. Retrieved February 28, 2008, from http://nces.ed.gov

Kea, C. D., Campbell-Whatley, G. D., & Bratton, K. (2003). Culturally responsive assessment for African American students with learning and behavioral challenges. *Assessment for Effective Intervention, 29*(1), 27-38.

Klingner, J., & Artiles, A. (2003). When should bilingual students be in special education? *Educational Leadership, 61*(2), 66-71.

Klingner, J. K., & Edwards, P. A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41*(1), 108-117.

Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools, 34*(1), 44-55.

Larry P. v. Riles, 343 F. Supp. 1306 (N.D. Cal. 1972), 502 F.2d 963 (9th Cir. App. 1974), 495 F. Supp. 926 (N.D. Cal. 1979), *aff'd*, 793 F.2d 969 (9th Cir. App. 1984, 1986).

Losen, D. J., & Orfield, G. (Eds.). (2002). *Racial inequity in special education*. Cambridge, MA: Harvard Education Press.

MacSwan, J., & Rolstad, K. (2003). Linguistic diversity, schooling, and social class: Rethinking our conception of language proficiency in language minority education. In C. B. Paulston & R. Tucker (Eds.), *Sociolinguistics: The essential readings* (pp. 329-340). Oxford, England: Blackwell.

MacSwan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teachers College Record, 108*(11), 2304-2328.

MacSwan, J., Rolstad, K., & Glass, G. V. (2002). Do some school-age children have no language? Some problems of construct validity in the pre-LAS Español. *Bilingual Research Journal, 26*(2), 395-420.

Mahoney, K. S., & MacSwan, J. (2005). Reexamining identification and reclassification of English language learners: A critical discussion of select state practices. *Bilingual Research Journal, 29*(1), 31-42.

Maoz, S. (2000). Alternative assessment. *English Teachers' Journal (Israel), *(53), 91-93.

McCloskey, D., & Athanasiou, M. S. (2000). Assessment and intervention practices with second-language learners among school psychologists. *Psychology in the Schools, 37*(3), 209-225.

McGrew, K., & Woodcock, R. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside Publishing.

Missiuna, C., & Samuels, M. T. (1989). Dynamic assessment of preschool children with special needs: Comparison of mediation and instruction. *Remedial and Special Education (RASE), 10*(2), 53-62.

Moore-Brown, B., Huerta, M., & Uranga-Hernandez, Y. (2006). Using dynamic assessment to evaluate children with suspected learning disabilities. *Intervention in School and Clinic, 41*(4), 209-217.

Mueller, T. G., Singer, G. H. S., & Carranza, F. D. (2006). A national survey of the educational planning and language instruction practices for students with moderate to severe disabilities who are English language learners. *Research and Practice for Persons with Severe Disabilities (RPSD), 31*(3), 242-254.

National Center on Student Progress Monitoring Technical Review Committee. (2007). *Review of progress monitoring tools*. Retrieved April 7, 2008, from http://www.studentprogress.org/chart/chart.asp

National Council for the Accreditation of Teacher Education. (2007). *The NCATE unit standards*. Retrieved February 24, 2008, from http://www.ncate.org/documents/standards/ unitstandardsmay07.pdf

National Organization on Disabilities. (2001). *Employment rates of people with disabilities*. Retrieved April 8, 2007, from http://www.nod.org/index.cfm?fuseaction=Feature. showFeature&FeatureID=110

Notari-Syverson, A., Losardo, A., & Lim, Y. S. (2003). Assessment of young children from culturally diverse backgrounds: A journey in progress. *Assessment for Effective Intervention, 29*(1), 39-52.

Office of Special Education, & United States Department of Education. (2006). *OSEP grants and funding*. Retrieved April 7, 2008, from http://www.ed.gov/fund/grant/apply/osep/ funding.html

Ortiz, A. A., Wilkinson, C. Y., Robertson-Courtney, P., & Kushner, M. I. (2006). Considerations in implementing intervention assistance teams to support English language learners. *Remedial and Special Education, 27*(1), 53-63.

Overton, T., Fielding, C., & Simonsson, M. (2004). Decision making in determining eligibility of culturally and linguistically diverse learners: Reasons given by assessment personnel. *Journal of Learning Disabilities, 37*(4), 319-330.

Parrish, T. (2002). Disparities in the identification, funding, and provision of special education. In D. J. Losen & G. Orfield (Eds.), *Racial inequity in special education* (pp. 15-38). Cambridge, MA: Harvard Education Press.

Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal, 29*(2), 387-409.

Reschly, D. J. (1997). *Disproportionate minority representation in general and special education: Patterns, issues, and alternatives*. Des Moines, IA: Mountain Plains Regional Resource Center. (ERIC Document Reproduction Service No. ED415632)

Rhodes, R. L., Ochoa, S. H., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide*. New York: Guilford Publications.

Roache, M., Shore, J., Gouleta, E., & Butkevich, E. (2003). An investigation of collaboration among school professionals in serving culturally and linguistically diverse students with exceptionalities. *Bilingual Research Journal, 27*(1), 117-136.

Roseberry-McKibbin, C. (2002). *Multicultural students with special language needs: Practical strategies for assessment and intervention*. Oceanside, CA: Academic Communication Associates.

Roseberry-McKibbin, C., & O'Hanlon, L. (2005). Nonbiased assessment of English language learners: A tutorial. *Communication Disorders Quarterly, 26*(3), 178-185.

Rueda, R., & Windmueller, M. P. (2006). English language learners, LD, and overrepresentation: A multiple-level analysis. *Journal of Learning Disabilities, 39*(2), 99-107.

Saenz, T. I., & Huer, M. B. (2003). Testing strategies involving least biased language assessment of bilingual children. *Communication Disorders Quarterly, 24*(4), 184-193.

Schrag, J. (2000). *Discrepancy approaches for identifying learning disabilities*. Alexandria, VA: National Association of State Directors of Special Education. (ERIC Document Reproduction Service No. ED449595)

Shang, H.-f. (1998). Comparing and developing a framework for alternative literacy assessment. *International Journal of Lifelong Education, 17*(4), 265-273.

Shinn, M. M., & Shinn, M. R. (2002). *AIMSweb training workbook*. Eden Prairie, MN: Edformation.

Sideridis, G. D. (2007). International approaches to learning disabilities: More alike or more different? *Learning Disabilities Research & Practice, 22*(3), 210-215.

Smolen, L., Newman, C., Wathen, T., & Lee, D. (1995). Developing student self-assessment strategies. *TESOL Journal, 5*(1), 22-28.

Spinelli, C. G. (2006). *Classroom assessment for students in special and general education*. Columbus, OH: Pearson Education.

Spinelli, C. G. (2008). Addressing the issue of cultural and linguistic diversity and assessment: Informal evaluation measures for English language learners. *Reading & Writing Quarterly, 24*(1), 101-118.

Squires, J., Potter, L., & Bricker, D. (1999). *The ASQ user's guide for the ages and stages questionnaires: A parent-completed, child-monitoring system* (2nd ed.). Baltimore, MD: Paul H. Brookes Publishing.

Suinn, R. M., Ahuna, C., & Khoo, G. (1992). The Suinn-Lew Asian self-identity acculturation scale: Concurrent and factorial validation. *Educational and Psychological Measurement, 52*, 1041-1046.

Sutton, J. P., Nowacek, E. J., Hausman, A. M., Stoiber, R. M., Stoiber, K. C., & Tindal, G. A. (2003). Assessment of students from geographically diverse rural areas: Emerging research from four federally funded projects. *Assessment for Effective Intervention, 29*(1), 53-68.

Thurlow, M. L., Elliott, J. L., & Ysseldyke, J. E. (2003). *Testing students with disabilities: Practical strategies for complying with district and state requirements* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Tombari, M., & Borich, G. (1999). *Authentic assessment in the classroom: Applications and practice*. Upper Saddle River, NJ: Prentice-Hall.

Tsai, H.-L. A., McClelland, M. M., Pratt, C., & Squires, J. (2006). Adaptation of the 36-month ages and stages questionnaire in Taiwan: Results from a preliminary study. *Journal of Early Intervention, 28*(3), 213-225.

United States Census Bureau. (n.d.). U.S. census bureau. Retrieved March 25, 2008, from http://www.census.gov

United States Census Bureau. (2006). American community survey. Retrieved March 25, 2008, from http://factfinder.census.gov

United States Department of Education. (2001). *No child left behind*. Retrieved April 7, 2006, from http://www.ed.gov/nclb/landing.jhtml?src=pb

United States Department of Education. (2004). *Individuals with disabilities education improvement act of 2004*. Retrieved March 2, 2008, from http://idea.ed.gov

United States Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2006). *26th annual (2004) report to congress on the implementation of the individuals with disabilities education act*, *vol. 2*. Washington, D.C.: Author.

Valdivia, R. (1999). *The implications of culture on developmental delay*. Reston, VA: ERIC Clearinghouse on Disabilities and Gifted Education. (ERIC Document Reproduction Service No. ED438663)

Wagner, R. K., Francis, D. J., & Morris, R. D. (2005). Identifying English language learners with learning disabilities: Key challenges and possible approaches. *Learning Disabilities Research and Practice, 20*(1), 6-15.

Wagner, R. K., Torgeson, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO-Ed.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial & Special Education, 26*(4), 207-214.

Williams, M., Turkheimer, E., Schmidt, K., & Oltmanns, T. (2005). Ethnic identification biases responses to the Padua inventory for obsessive-compulsive disorder. *Assessment, 12*(2), 174-185.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Appendix

Summary of Alternative Assessments

Type	Description	Potential Uses	Benefits	Drawbacks
Comparable Standardized Assessments	Carefully modified versions of standardized, norm-referenced tests	Identification of learning or emotional disabilities; annual data collection	Comparison with national norms; widely available; have established theoretical foundations	Established norms may not apply; translated versions do not necessarily equate original test; cultural bias and loading may be present
Dynamic Assessment	Uses mediated learning to determine a student's learning potential, specific areas of need, and strengths	Targeting intervention areas; Assessing student gains with assistance; building skills	Compatible with classroom instruction; reduced cultural bias	Time and training may be an issue for some teachers; lack of research on validity and reliability
Curriculum-Based Assessment	Assesses particular skill set in relation to curriculum or set of criteria	Frequent progress monitoring; guiding instructional decision-making; assessing interventions	Efficient, easy to use; can be teacher made or standardized; not time consuming; databases exist for comparing student skills to expected criteria	May result in teaching to test; variability in teacher made assessments; possible disconnect from actual curriculum
Performance Assessment	Students demonstrate skill or knowledge by performing a task	Observe student growth in multiple areas; used as formal and informal assessment;	Works well with diverse learners; showcases students' actual abilities in context; allows families to participate	Can be cumbersome or time consuming; varies greatly depending upon circumstances; lack of research on reliability and validity for SWDCLD

Note: SWDCLD = Students with disabilities who are culturally and linguistically diverse