Issues in the Assessment of Culturally and Linguistically Diverse Students with Special Needs

JANETTE K. KLINGNER, KAILONNIE DUNSMORE, AND PATRICIA EDWARDS

Recent research and policy changes highlight the need to examine the cultural and linguistic dimensions of assessment, particularly as part of a system for identifying students with disabilities. A considerable body of evidence substantiates that students of color, students living in poverty, and English language learners are disproportionately represented in special education (Artiles, Rueda, Salazar, & Higareda, 2005; Donovan & Cross, 2002). In addition, intervention research suggests that appropriate instruction can reduce the need for special education by accelerating students’ literacy development (Scanlon, Vellutino, Small, Fanuele, & Sweeney, 2005; Blachman et al., 2004; Taylor, Short, Shearer, & Frye, 2007). Changes in legal policies at state and national levels around special education identification procedures have forced far greater attention on the instructional context and on the appropriateness and quality of intervention services. This has led the education community to confront more directly the culturally situated nature of assessment tools and practices and the close link between assessment, instruction, and educational goals when evaluating students’ abilities.

In this chapter, we examine new directions for literacy assessment in a special education system that attends to the cultural basis of school-based cognitive development and learning (Rogoff & Chavajay, 1995). We describe recent changes in the assessment and identification of learning disabilities (LD), assessment in Response to Intervention (RTI) models, cultural and linguistic influences on test performances, problematic assessment procedures, and culturally and linguistically responsive assessment practices.

Changes in the Assessment and Identification of Learning Disabilities

The last decade has brought dramatic changes in how we assess students to determine whether they possess a learning disability (LD). In the almost 50 years since Samuel Kirk first used the term “learning disability” (Kirk & Bateman, 1962) and Barbara Bateman described LD as “an educationally significant discrepancy between [an individual’s] estimated intellectual potential and actual level of performance related to basic disorders in the learning process” (1965, p. 220), psychologists and other members of multi-disciplinary assessment teams have identified LD primarily by determining a student’s potential, typically by administering an intelligence test, and comparing that with the student’s academic achievement. Reading was by far the most common area of disability (about 80%). Yet, for various reasons, numerous researchers, professional groups, parents, and policy makers challenged the IQ-achievement discrepancy as the core of LD’s definition and focal point of identification procedures (see Aaron, 1997, for a review of research challenging the validity of the discrepancy formula). In this chapter, we discuss concerns about LD identification procedures that inadequately assess diverse student populations. We also describe current assessment practices and highlight some promising directions for developing assessment and instructional tools that acknowledge the cultural origin of human cognition (Rogoff, 2003) and literacy practices more specifically (Cole, 1996).

Problems with the Construct of Unexpected Underachievement as Assessed with IQ Tests

Until recently, one of the fundamental principles of LD has been the belief that there are real differences between individuals with a demonstrated discrepancy between IQ and achievement and those individuals with a “flatter” profile, or, in other words, who are low in both IQ and achievement. The presumption has been that the instructional needs of students with “true” LD are somehow different than those of students who are at the lower end of normal distributions in IQ and achievement. Yet, this does not seem to be the case. As far back as in 1992, Fletcher, Francis, Rourke, Shaywitz, and Shaywitz compared five groups of children with varying IQ-achievement profiles and determined that differences among groups were minimal. They concluded their report
by questioning the validity of segregating students according to IQ-achievement discrepancies.

Stuebing et al. (2002) conducted a meta-analysis of 46 studies to evaluate the validity of LD classifications based on IQ-discrepancy and the exclusion of students who are poor readers but who do not display this discrepancy. They found substantial overlap between IQ-discreant and IQ-consistent poor readers, and noted that differences could be largely explained by the selection criteria used to form groups. In other words, students' membership in a given IQ-discreant or IQ-consistent group did not predict how well they would respond to reading instruction. Stuebing et al. recommended doing away with IQ testing as part of LD identification procedures, and instead moving to a model that includes looking at how students respond to research-based interventions. Like other researchers who examined the validity of IQ-discrepancy classifications, Stuebing and her colleagues focused on reading and not other aspects of written or oral language.

Over this same time span, literacy intervention research was accumulating that indicated that appropriate instruction could reduce the number of students qualifying for special education, especially in reading. A number of studies found that adequate classroom instruction, which was differentiated in both material resources and instructional practices to support the individual learning needs of students, could effectively ensure that most students reached grade level achievement (Hiebert & Taylor, 2000; Cunningham & Allington, 2007; Taylor, Pearson, Clark, & Walpole, 2000). Other researchers provided evidence that specifically targeted interventions (within the classroom or provided by trained personal as supplementary to classroom instruction) could bring to grade level the majority of those students who continued to struggle with literacy difficulties (Blachman et al., 2004; Vellutino, Scanlon, Small, & Fanuèle, 2006; Pinell, Lyons, DeFord, Bryk, & Seltzer, 1994). This research contributed to our understanding of the instructional context within which student literacy abilities are developed, supported, and assessed. In fact, Vellutino, Scanlon, Zhang, and Schatschneider (2008) emphasized that many students' reading disabilities have an instructional etiology that may or may not be in addition to a biological one.

At about the same time, the U.S. Department of Education, Office of Special Education Programs convened a panel of leading LD researchers to discuss LD identification procedures (Bradley, Danielson, & Hallahan, 2002). The panel agreed that the IQ-achievement discrepancy is not a valid way to identify students with LD, and concluded that the most promising method of identifying individuals with LD is to do away with IQ testing and instead to implement RTI.

Also at this time, the National Academy of Sciences convened a panel of experts to summarize the research related to the disproportionate representation of culturally and linguistically diverse students in special education. This group also concluded that IQ is ineffective in the identification of LD, and noted that the use of IQ tests with culturally and linguistically diverse students is fraught with problems (Donovan & Cross, 2002). They recommended alternatives that do not involve the use of IQ tests, and promoted the widespread use of early screening, early intervening services, and RTI. They asserted that if early reading programs help culturally and linguistically diverse students receive more appropriate instruction, then the number of students who continue to struggle will decrease and will more likely be those who require a special education.

**Response to Intervention** With these initiatives and new research findings serving as a backdrop, Congress passed the Individuals with Disabilities Education Improvement Act (IDEA), in 2004. IDEA (2004) promotes RTI as way to identify students with LD and gives states the option of discontinuing the use of IQ-Achievement discrepancy formulas. IDEA also stipulates that schools should screen students and provide Early Intervening Services to all students who are not making progress, as part of general education, before a student is considered for possible special education placement.

Yet some of the assumptions underlying RTI are flawed, especially when applied with culturally and linguistically diverse students. For example, the idea of “research-based” tends to be applied with a one size fits all mentality, without considering issues of population and ecological validity (Klingner & Edwards, 2006). When students do not respond to interventions, educators seem too quick to assume the problem resides in the child, and not with the quality and appropriateness of instruction (Orosco & Klingner, 2010). It appears that there is not enough of a focus on the context for learning and improving instruction and educational opportunities for students. Also, educators seem to be relying on just one or two progress monitoring measures to assess students’ response to instruction, rather than conducting comprehensive evaluations of students’ strengths and areas of need. The most common progress monitoring tools (e.g., Dynamic Indicators of Basic Early Literacy Skills) may hold limited internal or external validity for the literacy skills being targeted by teachers or state assessment systems (Hoffman, Jenkins, & Dunlap, 2009; Shelton, Altweger, & Jordan, 2009; Nelson, 2008).

Although we have questions and concerns about RTI, particularly with culturally and linguistically diverse students, we believe that it provides an opportunity to move in new directions and address problematic past practices. In any case, the IQ-achievement discrepancy way of identifying LD is obsolescent.

**Assessment in RTI.** RTI is an assessment-based system. Yet creating the appropriate system is a challenge in culturally and linguistically diverse schools (Klingner, Soltero-González, & Lesaux, 2010). When considering how best to assess student learning, it is important to consider school contexts and to reflect about school-level models of appropriate assessment and instruction. Rather than simply focusing on individual students and emphasizing students’
differences in relation to a majority culture norm, the goal should be to design assessment systems and effective learning environments tailored to the needs of diverse school populations.

All assessment data should be used expressly to support student learning, whether to inform constructive discussions about classroom-level instruction or student-level supplemental supports. It is important to consider the function of each assessment and not to use assessments for other purposes than those for which they were intended. There is no single best test or assessment strategy. Different assessments—even in the same language or literacy domain—capture somewhat different skills and knowledge. This is because of the assessment format, background knowledge required for the items, and how the task is defined for measurement. All RTI assessment strategies should reflect the multi-dimensional nature of language and reading. Multiple assessments should be used to provide a comprehensive portrait of a given student’s strengths and needs.

Screening batteries should sample from the entire range of language and literacy skills needed for academic success. While many screening batteries focus on print awareness, phonological awareness, and letter-word identification, they typically do not include measures of vocabulary knowledge, oral language proficiency, and/or listening and reading comprehension. This omission is particularly detrimental for linguistically diverse students, many of whom develop age-appropriate word reading skills but need support to further develop their language and comprehension skills.

Progress monitoring should include analyses of oral reading (e.g., Running Records, Qualitative Reading Inventory) in addition to measures of oral reading fluency. Writing should be analyzed as well, for example, using rubrics such as the 6+1 Writing Trait Model. Content expertise in literacy as well as language acquisition are required in order to analyze and interpret these assessments, but they can yield high validity and be done with a great deal of rigor.


cultural influences on test performance

Any test performance is the result of a complex interaction between the task characteristics and the constraints inherent in the testing situation. Variance across cultural groups in test performance may be due to different interpretations of the nature of the task, the type of problem being solved, and the process used to find a solution (Goodnow, 1976). Cultural variation in assessment results may be due to differences in the situations in which various cultural groups typically apply their skills more than to differences in the actual skills (Cole & Bruner, 1971). Individual interviews can be helpful to determine why a student interprets a test item in a particular way (Solano-Flores & Nelson-Barber, 2001).

Because of how they are developed, standardized tests typically possess strong psychometric properties for some students, particularly those who speak English as their first language and are from mainstream backgrounds. Yet when used with culturally and linguistically diverse students, these tests lack the same levels of predictive and construct validity they show with English-only students (Abedi, 2002; MacSwan, Rolstad, & Glass, 2002). In other words, they tend to underestimate achievement. One potentially problematic aspect of test development is that test developers draw test items from a universe of information to which they assume test takers have been exposed. Yet children from diverse backgrounds might not have been exposed to this body of information, placing them at a disadvantage (Samuda, 1989). Rather, they have been exposed to different information. The reality is that assessments in the United States reflect the abilities, skills, knowledge, and language valued by U.S. “core culture” (Mercer, 1973, p. 13). Yet, this fact is not adequately considered when interpreting test results for students from diverse backgrounds. The evaluator’s challenge is to figure out what a score on a given test means, and what can and cannot be concluded from assessment results.

Research in the neurosciences has more recently provided additional evidence of the needs to examine learning disabilities as a cultural as well as biological phenomenon and to develop assessment tools that attend to individual literacy development as a culturally situated process. “In general, genetic research comparing abilities and disabilities suggests that what we call ‘learning disability’ is merely the low end of the same genetic and environmental factors responsible for the normal distribution of learning ability. In other words, the abnormal is normal” (Kovas & Plomin,
Issues in the Assessment of Culturally and Linguistically Diverse Students with Special Needs

2008, p. 14). This suggests that assessment tools for both regular education students and those identified with special needs should attend to literacy development as a culturally situated process, understanding that any genetic or biological differences are mediated through the instructional environment.

Similarly, efforts to map the brain during various aspects of reading (e.g., word retrieval, phonological processing, and comprehension) indicate differential patterns for students with phonological processing difficulties. However, longitudinal research and cross cultural variation suggest that it is difficult to extricate the cultural origins of differences from those of the biological when examining the etiology of disability (Katzir & Paré-Blagoev, 2006). Current views of brain development emphasize its plasticity and developmental responsiveness to environmental stimuli far beyond what was initially thought of as critical windows for language and literacy development (Miller & Tallal, 2006).

In fact, functional magnetic resonance imaging (fMRI) of the brains of adults and children with dyslexia reveals differential patterns; significantly, such research also indicates that intervention or instruction can reduce or ameliorate these differences as the brain learns, over time revealing more “normal” patterns of brain activity (Gaab, Gabrieli, Deutsch, Tallal, & Tempe, 2007).

Developing protocols for using assessment data in ways that support student development and highlight the plasticity of the brain can be hindered by views of reading (dis)ability as a static phenomenon as opposed to a condition that may be experientially (or instructionally) constructed (Fischer, 2008). Through longitudinal attention to the instructional context of literacy learning as well as the cognitive profiles of students as evidenced on formal, standardized measures, work by Vellutino et al. (2006, 2008) contributes to our understanding of the ways in which assessments can under-examine the socio-cultural dimension of school instruction and practice. What we assess as “ability” is as much a function of the instructional and environmental context and access and opportunity as it is mental constraints. Psychometric assessment is itself, therefore, a cultural activity, although it has tended to portray itself as revealing dimensions of student ability distinct from cultural experience and practice.

**Linguistic Influences on Test Performance** Numerous scholars have written about the potential for linguistic bias in assessments (Rogoff, 2003; Solano-Flores, 2006; Valdés & Figueroa, 1994). ELLs’ test performance is affected by issues of language dominance and language proficiency. They may have only a surface level understanding of vocabulary on an exam and be confused by words with multiple meanings, referents, and pronouns. Assessments rely on particular linguistic forms that are familiar to some children but may be unfamiliar to others (Rogoff, 2003).

Children who are familiar with the language and structure of a test are more likely to respond in ways expected by the examiner.

Test developers typically do not take into consideration all the ways language, dialect, and register affect students’ understanding of test items. Addressing the language adequacy of test materials requires the use of sophisticated approaches to test development (Solano-Flores, Trumbull, & Nelson-Barber, 2002; Tanzer, 2005). Solano-Flores (2008) contends that “current ELL testing practices are limited in their effectiveness to produce valid measures of academic achievement because they are based on categorical, deterministic views of language and erroneous assumptions about the capacity of assessment systems to effectively communicate with ELL students” (p. 189). Problematic procedures include classifying ELLs into a few overly simplistic categories of language proficiency and treating ELLs as if they are more linguistically homogeneous than they are.

**Problematic Assessment Procedures** In numerous studies of ELL testing practices over the years, researchers have found that psychologists tend to over-rely on the results of English-language testing, exclude native language test results, and give inadequate attention to language acquisition issues or classroom context as possible explanations for students’ struggles (Figueroa & Newsome, 2006; Harry & Klingner, 2006; Klingner & Harry, 2006; Maldonado-Colon, 1986; Ochoa, Rivera, & Powell, 1997).

For example, Harry and Klingner (2006) found that the psychologists in their investigation of disproportionate representation seemed to be overly influenced by teachers’ informal diagnoses of children’s problems, school personnel’s impressions of the family, and external pressures for identification and placement in special education. They sometimes disregarded established identification criteria and placed students in special education programs for which they did not truly qualify, to “save” a child or to “protect” him. They apparently lacked understanding of cultural and linguistic influences on test performance. To address these problematic assessment procedures, we recommend shifting from a focus on finding out what is wrong with students to figuring out how best to help them. In other words, assessments in special education or as part of RTI should be used in more formative ways than in the past, or, in other words, to identify learning needs to improve teaching by making instructional adjustments (Shepard, Hammerness, Darling-Hammond, & Rust, 2005).

**Culturally Responsive Assessment** All learning is cultural (Cole, 1996; Rogoff, 2003). Therefore, as Moje and Hinchman (2004) observed, “All practice needs to be culturally responsive in order to be best practice” (italics added, p. 321). Culture is not a static set of characteristics located within individuals but is fluid and complex, varying across contexts (Gutierrez & Rogoff, 2003). Children are socialized to learn in particular ways in their homes and communities. In order to provide valid measurements of what students know and can do, assessment practices must access student knowledge in ways that are consonant with how they...
organize and understand subject matter, and with how they express their understandings.

Culturally responsive literacy assessments must go beyond examining the degree to which students have acquired basic components or subprocesses in reading and writing and also examine the ways that best support meaning making for particular students. We draw here upon the accommodation and incorporation principles of Wiley’s (1996) framework as possible ways of developing and interpreting assessment practices in culturally diverse contexts (Klingner & Edwards, 2006). Embedded in each of these perspectives is an assumption that assessment is inextricably linked with instructional practice. That is, what and how we assess student learning is reciprocally related to the strategies and goals of instruction.

Accommodation, according to Wiley (1996), requires educators to deepen understandings of the communicative styles and literacy practices among their students and to account for these in their instructional and assessment practices. Several qualitative studies have shown that, even in impoverished conditions, homes can be rich in print and family members can engage in literacy activities of many kinds on a daily basis (Anderson & Stokes, 1984; Heath, 1983; Purcell-Gates, 1996; Taylor & Dorsey-Gaines, 1988). Assessment practices that accommodate the rich linguistic and cultural resources that students bring to literacy learning require tools and protocols that address the resources that students develop in home and community contexts to support goal-related literacy activities (Heath, in press; Moje & Hinchman, 2007). For example, Moje and colleagues have developed survey tools for examining adolescent’s interests, skills, and experiences with literacy (Moje et al., 2008).

Incorporation (Wiley, 1996) involves studying community practices that have not been valued previously by schools and incorporating them in the curriculum. It also means surrendering a privileged position and acknowledging that much can be learned from diverse ethnic groups. Assessment tools explicitly developed to address how students define, value, and use literacy as part of their practices in and out of school “offer the potential for schooling to be adjusted to meet the needs of families” (Cairney, 1997, p. 70). These kinds of assessment tools provide the potential for the reflexive analysis that reveals not only what students know and can do but the conditions under which such behaviors are most likely to occur. These tools move beyond descriptions of individual capacities to broader examinations of the instructional context. For instance, Moll (2009) described how teachers learn to engage in new forms of assessment and instructional practices through involvement in their communities.

Conclusion

In conclusion, assessment practices should not be isolated from the goals of the literacy curriculum and associated assumptions about the meanings of literate practice in diverse communities (Klingner & Edwards, 2006). Assessments do not provide instructionally valid data if they are decontextualized from the practices within which literate knowledge and skills are constructed. Whether as part of RTI or another purpose, assessment results must be considered in relation to the sociocultural contexts in which instruction and assessment take place (Artiles, 2002; Gee, 2001). In this chapter, we have tried to provide a framework for developing and using assessment data that moves closer to leveling the educational playing field for African American, Hispanic, and other culturally and linguistically diverse students in the United States.

References


