# Alternative Assessment for Transitional Readers

Carol Beaumont
Palo Alto Unified School District

Julia Scherba de Valenzuela
University of New Mexico

Elise Trumbull
WestEd Educational Laboratory

## Abstract

This paper presents an analysis of one district's performance based assessment of reading comprehension specifically designed for 5th grade transitional bilingual students in their first year of English instruction. In contrast to the common practice of using one assessment measure for all students and modifying assessment conditions for students from diverse backgrounds, they created a new instrument which was administered instead of the district-wide performance based assessment. Findings from the study indicate that: (a) characteristics of second language learning in the areas of phonology, syntax, and semantics were present to varying degrees in 100% of the students' responses; (b) specific knowledge of the linguistic characteristics of transitional students was necessary for effective interpretation and accurate scoring of student responses; (c) the transitional students demonstrated the ability to comprehend and interpret English texts at inferential levels when assessed on the transitional assessment measure; and (d) the transitional performance-based assessment results suggested to teachers many possible changes they could make in their instruction to meet the literacy needs of transitional students.

Equitable assessment for English language learners remains one of the greatest challenges of educational reform. The complex interaction of language, culture, educational background, and opportunities to learn make equitable, large-scale assessments particularly difficult to achieve (Figueroa, 1990; García & Pearson, 1994; Solano-Flores & Nelson-Barber, 2001; Ulibarri, 1985). Often, English language learners (ELLs) are assessed through the same mechanisms as their English-only counterparts. This practice compromises the accuracy

of results and may lead to inappropriate interpretations about English language learners' proficiency in reading and academic content areas (García, 1991; Miramontes, Nadeau & Commins 1997; Valdés & Figueroa, 1994).

The purpose of this paper is to describe student performance on a specially designed assessment of reading comprehension created for native Spanish-speaking students in their first year of English instruction. This study was guided by the following questions: (a) How do scoring and student evaluation practices in this assessment account for linguistic characteristics of first-year English language learners? (b) What evidence exists that reflects first-year English language learners' ability to process English texts at an inferential level? (c) How can this transition assessment inform curriculum and instruction for these students?

## Assessment Reform and English Language Learners

In the decade-long debate on school reform, some educators favor new methods of assessment to remedy persistent problems in the American schooling. They have suggested that complex, cognitively demanding assessments can drive improvements in curriculum and instruction (e.g., Darling-Hammond, 1994; Mitchell, 1992; O'Day & Smith, 1993). They have also pointed out the overrepresentation of ELLs and other students from non-dominant cultures among poor performers on standardized tests (Figueroa, 1990; Valdés & Figueroa, 1994; Williams, 1996) and the consequent reinforcement and extension of social and educational inequalities (Darling-Hammond, 1994, Durán, 1989; Oakes, 1985).

ELLs' lack of full English proficiency is an obvious reason for their poor performance on standardized tests (LaCelle-Peterson & Rivera, 1994; Cummins, 1989; Genesee & Hamayan, 1994), but it reflects a limited understanding of the factors contributing to a student's performance on an assessment or standardized test. To begin to grasp the relationship between English proficiency and test performance requires a deep understanding of the developmental nature of second language acquisition, the difference between surface fluency and the ability to participate in content area instruction in English, the time required to achieve academic fluency, the kinds of language conventions used on typical assessments and tests, and the language skills necessary for success on standardized tests. Because the length of time required for full English proficiency is often underestimated (Collier; 1988; Cummins, 1989; Valdés, 1998), and students may appear to be more proficient with English than they are, they may be expected to take tests in English long before they are fully proficient with the kind of academic language needed to perform well. All of these factors should be considered in designing appropriate assessment and instruction.

Standardization itself rules out any contextualization of assessment, meaning that linguistic differences among students cannot be accounted for adequately. The practice of assessing English language learners with the same mechanisms as their English-only counterparts may seriously compromise the validity of results and lead to misleading interpretations and unfair decisions affecting their futures (August & Hakuta, 1997; García, 1991; LaCelle-Peterson & Rivera, 1994; Miramontes, Nadeau & Commins, 1997; Valdés & Figueroa, 1994).

English-language assessment prompts that make extensive use of complex or idiomatic language penalize English language learners who may access important concepts in their first language, but not yet in English, or they may access them more slowly in English (Abedi, 2001; Abedi, Leon, & Mirocha, 2001, Figueroa & García, 1994, García, 1991, Heubert & Hauser, 1999). Misinterpretations of the directions or text of an assessment task can lead to flawed conceptualization of the problem to be solved and consequent failure to devise a correct solution (Durán, 1989). In such cases, teachers or others who grade or score an English language learner's performance may falsely underestimate that student's level of understanding or skill.

A move to standards-based performance assessments and other forms of "authentic" assessment (Valencia, Hiebert, & Afflerbach, 1994; Wiggins, 1993) does not ensure assessment validity for ELLs. These kinds of assessments are subject to the same sources of error, particularly given their increased language demand in comparison to multiple choice or short answer tests (August & Hakuta, 1997; Shepard, 1993). Farr and Trumbull (1997) point out that, "good instruction and assessment should look different in different environments, depending on the students served" (p. 2). Assessments designed for native English speakers will simply not meet standards of validity for English language learners. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999):

> For all test takers, any test that employs language is, in part, a measure of their language skills. . . . Language dominance is not necessarily an indicator of language competence in taking a test, and some accommodation may be necessary even when administering the test in the more familiar language. Therefore, it is important to consider language background in developing, selecting, and administering tests and in interpreting test performance. (p. 91)

Specific research studies on the validity of performance-based assessments for ELLs have focused primarily on accommodations (e.g., increasing the time allotted for completion, allowing the use of dictionaries, or modifying the language of test prompts) rather than on major changes in

the instruments themselves (Abedi, 1997; Abedi, Lord, Hofstetter, & Baker, 2000; Kiplinger, Haug, & Abedi, 2000; Kopriva, 1997; Olson & Goldstein, 1997; Sweet, 1997). Some have made efforts to make fairer the scoring of ELLs' responses (Wong Fillmore & Lara, 1996). However, little research literature exists on performance assessments for English language learners.

## Second Language Learning and Reading Comprehension

The theoretical foundation for much of the literature on second language reading is the construct of linguistic interdependence based largely on Cummins' (1981) construct of a Common Underlying Proficiency (CUP), whereby skills transfer from one language to another. The theory of linguistic interdependence indicates that well-established skills in the first language transfer to the second language. The theory of linguistic interdependence incorporates both content and processes in reading (e.g., subject matter knowledge, conceptual knowledge, higher-order thinking skills and reading strategies). Support for the transfer phenomenon in studies investigating reading behavior has been widespread (Fitzgerald, 1995; Krashen, 1996). Evidence cited for the transfer hypothesis centers around (a) use of similar strategies to read in both languages (e.g., Pritchard, 1990); (b) predictability of level of reading proficiency in the second language on the basis of reading proficiency in the first (e.g., Calero-Breckheimer & Goetz, 1993; Carrell, 1991; Droop & Verhoeven, 1998; García, 1998; Geva, Shany, & Himel, 1992; Saville-Troike, 1984); and (c) an apparently closer correlation between reading proficiencies in the two languages than between reading and oral proficiency in the second language (Geva, Wade-Woolley, & Shany, 1993; Saville-Troike, 1984).

Some research has focused on specific language skills and the degree to which they may transfer from reading in one language to reading in another. Durgunoglu, Nagy, and Hancin-Bhatt (1993) showed that young children's phonological awareness and word recognition skills in Spanish were predictive of phonological awareness and word recognition in English. However, Heubert and Hauser (1999) conclude that transfer is not automatic: "It occurs only when conditions for the emergence of the analogous second-language skills exist, and it can be aided by explicit support for the process of transfer" (p. 224). Others caution that when cognitive and linguistic factors are taken into account to the exclusion of macro social factors (i.e., status of languages and relations among groups of people), evidence of variance in literacy success is overlooked (Gee, 1999).

Information on text processing among bilinguals suggests that several factors may create difficulties for students reading in their second language: (a) the degree to which reading comprehension strategies are well developed in the first language (Jimenez, García, & Pearson, 1996; Langer, Bartolome, Vasquez, & Lucas, 1990; Thonis, 1983; Westby, 1989; Laberge & Samuels,

1974; Jimenez, 1997; Bialystok, 1991); (b) characteristic features of the English language and the specifics of English text structures (Kucer & Silva, 1995); (c) inaccurate or incomplete background knowledge (Jimenez, García, & Pearson, 1996; Kucer & Silva, 1995; Flood & Menyuk, 1983; Vellutino, Scanlon, & Tanzman, 1990; García, 1991; Westby, 1989; Bartolome, 1994; Edelsky, 1983); (d) vocabulary limitations (Wong Fillmore, 1989; Gibbons, 1991; Nagy, García, Durgunogulu, & Hacin-Bhatt, 1993; García, 1991; Baker, 2000); and (e) speed of processing (Fitzgerald, 1995; Jimenez, García, & Pearson, 1996).

When reading comprehension is assessed through writing, these difficulties are compounded (Gutiérrez, Baquedano-López, Alvarez, & Chiu, 1999; Reyes, 1992). As writing development may tend to lag somewhat behind reading development in English language learners (DeAvila, 1997), assessment of reading through writing could underestimate reading proficiency. Also, time limits fail to take into account the slower rate at which English learners process texts (Fichtner, Peitzman, & Sasser, 1994; Peitzman, personal communication, April 7, 2000; Carrell, 1991; August & Hakuta, 1997; García & Pearson, 1991; Mercado & Romero, 1993; Pérez & Torres-Guzmán, 1996). Finally, when second language readers have the opportunity to use their dominant language to discuss texts written in their second language, they reveal deeper understandings than they were able to demonstrate when required to talk about them in English (Díaz, Moll, & Mehan, 1986).

It seems reasonable to assume that these factors may contribute to English language learners' poor performance on alternative assessments, unless they have been developed with English language learners in mind. These issues in the assessment of bilingual readers and particularly transitional students, point to the need for continued exploration of effective assessment instruments that take into account the unique profile of bilingual readers.

## Background to the Study

This study took place in an ethnically diverse urban school district in the West San Francisco Bay. Of the 8,500 students in the district, 49% were minority and 95% of those were Mexican-American. The district's bilingual program had been in place for many years and followed the multi-year program design based on California's "State Program for Students of Limited English Proficiency," where students receive primary language instruction in content areas for four years, in addition to daily English as a second language instruction. Students begin English content area instruction in science and/or math in fourth grade and language arts instruction in English in fifth grade. Students who enter after kindergarten are placed in primary language or English content area instruction based on their school background and language proficiency test scores.

During the study the district was in the fourth year of a major standards-based reform effort in language arts, mathematics, and social studies, K-8. The district had changed from a standardized, norm-referenced test to a Performance-Based Assessment (PBA) given at second, fifth, and seventh grades. In general, English learners had scored poorly in the first few years of the PBA. Many of the district's bilingual teachers had been involved in scoring the district assessment and believed the English learners' poor performance was indicative of an inappropriate instrument rather than a reflection of the students' abilities. They were particularly concerned that the reading level, vocabulary, content, and thematic focus of the reading selections were inappropriate for these students. After trying several alternative approaches and accommodations, teachers in the bilingual department proposed the development of an alternative assessment specifically designed for the transitional bilingual students. This paper discusses the results of the first administration of this teacher-developed alternative assessment.

## Method

### Participants

The primary participants in the study were 89 fifth-grade students in transition, completing their first year of English language arts instruction. All were from lower socio-economic status (SES), Mexican-American homes where Spanish was the primary language. While the criterion for transitioning to English instruction was grade level reading in Spanish, in practice many students were transitioned before that level of proficiency was reached. The test development workgroup consisted of 10 teachers, the director of bilingual education, and two consultants from the WestEd Regional Educational Laboratory. The teachers were either state certified bilingual teachers or Specially Designed Academic Instruction in English (SDAIE) teachers who had participated in specialized training in effective techniques for teaching transitional students. The director of bilingual education and the regional laboratory consultants all had extensive instructional experience with English language learners and alternative assessment and advanced degrees in linguistics and education.

### Data Sources

The data consisted of 89 performance-based assessments completed by the transitional students and field notes and audiotapes from test development meetings, scoring sessions, and interviews with the workgroup members. The summary and interpretation of events were triangulated with interviews with the project participants upon completion of the analysis. Project participants and one of the consultants read a draft of this manuscript and provided clarifications to the background and test development sections of this paper and feedback on the analysis of the tests.

*The Transitional Performance-Based Assessment (TPBA)*

The assessment was developed over an 18-month period during which workgroup members met to discuss relevant language issues; look for appropriate reading materials; create, field test, and evaluate test items; complete a final assessment measure; and test, score, and interpret student performance. The transitional assessment instrument, which was framed in second language acquisition theory and Vygotsky's construct of assisted performance in assessments (1978), paralleled the district-wide Performance-Based Assessment (PBA), (as required by the district assessment department): (a) students were required to read and respond to two English texts on related topics, one fiction and one nonfiction; (b) test items required constructed written responses targeting the fifth-grade language arts standards; and (c) the texts were given in English to evaluate a transitional program for students participating in English-only classroom instruction.

The workgroup considered the following most important: (a) adjust the reading level to make the texts accessible to transitional readers; (b) test more than one standard per item, enabling students to build on their knowledge across items; (c) design holistic scoring procedures to complement the integrated structure of the assessment; and (d) use of bilingual teachers for scoring.

In adjusting the reading level of the texts, the workgroup members considered three factors. The workgroup chose shorter texts than those used for the district PBA to allow bilinguals enough processing time. The workgroup also believed that the students could not be assessed fairly using fifth grade reading material since they had had only one year of English reading instruction. They selected texts they believed were age-appropriate in content but were less dense and vocabulary-dependent and culturally more familiar than the texts used in the district PBA. A classroom teacher administered the test in two periods: one for pre-reading activities to activate background knowledge and set the purpose for reading, and the other for administering the test itself. All classroom teachers used the same pre-reading activities, as did the English only teachers in administering the PBA.

Since many items of the TPBA measured multiple standards, the group chose holistic scoring to evaluate responses. In addition to changes in the content and structure of the test, which provided a kind of built-in scaffolding, other features of the test such as graphics, formatting, and pre-assessment activities provided for assisted performance consistent with fair assessment practice for English language learners. Strategies included development of necessary vocabulary, activation of background knowledge, and providing the opportunity to formulate ideas orally before writing.

District bilingual teachers scored the assessment during a two-day period approximately two weeks after it was given. Consultants who had read all of the tests developed a rubric and chosen anchor papers from student work,

facilitated scorer calibration. During the calibration process teachers read the anchor papers, assigned a score based on the rubric, and recorded evidence for their decisions, noting issues such as language or test formatting for later debriefing. During the subsequent discussion, teachers explained their rationale for assigning certain scores, and disagreements were resolved by reference to the rubric and advisement with the consultants.

After calibration, pairs of teachers scored each test independently, checked their scores with each other, discussed discrepancies, and reconciled their scores. Differences that could not be reconciled (25% of the cases) were refereed by one of the consultants.

### Data Analysis

All 89 tests were read several times, and 5 of the 11 test items were chosen for in-depth analysis because they required inferential thinking beyond basic identification and summarization tasks, and allowed for more detailed analysis of reading comprehension.

Each student's responses to the five items were then re-read multiple times without reference to pre-determined analytic categories. As patterns began to emerge from the readings, tentative analytic categories were identified. The tests were then read again to see how successfully each category captured aspects of student performance. They were then coded according to the evaluation scheme that developed. As examples of various response patterns were logged and preliminary interpretations were made, the categories were refined and examples re-examined for consistency.

Seventy-four percent of the students received holistic scores in the 2–3 range. According to the rubric, this score range reflects a processing of English texts primarily at the literal level. Levels 2 and 3 are described as "limited" and "adequate," respectively. However, because the tests were scored holistically, many individual examples of more abstract or inferential processing that might provide more specific information about the nature of transitional reading comprehension were not reflected in the score distribution. To further investigate the validity of this measure in assessing transitional students' reading comprehension skills, the holistic scoring was supplemented by more in-depth item analysis by this paper's authors. This analysis was undertaken to gain more specific information about the conditions under which higher level, inferential responses were elicited.

The analytic categories included phonological, syntactic, and semantic aspects of language as well as deeper levels of text processing regarding character and author motivation, text themes, and ability to assume other points of view (Westby, 1989). Although, originally, surface features or "conventions" of the written responses were not scored, for the purposes of this study they are part of the analysis here to amplify scorer response.

## Results and Discussion

Analysis of the students' performance revealed four significant findings: (a) Characteristics of second language learning in the areas of phonology, syntax, and semantics were present to varying degrees in 100% of the students' responses; (b) specific knowledge of the linguistic characteristics of transitional students was necessary for effective interpretation and accurate scoring of student responses; (c) the transitional students demonstrated the ability to comprehend and interpret English texts at inferential levels when assessed on the transitional assessment measure; and (d) the transitional performance-based assessment results suggested to teachers many possible changes they could make in their instruction to meet the literacy needs of transitional students.

### Finding 1: Second Language Learning Characteristics

In the first level of analysis, phonological, semantic, and syntactic features of the students' responses were analyzed to determine the role they played in the students' ability to express meaning and the scorer's ability to interpret meaning. Characteristic transitional language features such as phonological and syntactical evidence of Spanish language influence, semantic confusion, limited vocabulary, and limited variety in sentence structure appeared as expected. Two of these features, phonology and semantics, are discussed below.

#### *Phonological features*

Phonological indicators of the influence of Spanish appeared frequently. Spelling patterns such as writing "starbing" for starving, "kip" for keep, "slipping" for sleeping, "mit" for meet, "veri" for very, "way" for why, "many" for money, "nating" for nothing, and "so posto" for supposed to, reflect the orthographic influence of Spanish, particularly in the vowel substitutions. Other spelling patterns such as promus/promise, humpbell/humble, and concidaret/considerate reflect the students' attempts to sound out words in English using a one-to-one match between sound and symbol, a strategy that works well in Spanish but causes problems when applied to English, given the inconsistency of English orthography. Not only the inconsistency of English orthography but also the variations in how similar sounds are represented between the two languages pose problems for Spanish readers. For example, the /h/ sound in English can be represented in Spanish by the letters "g," when followed by "e" and "i," and "j."

The following examples, taken from responses to several items, illustrate these phonological patterns as they occurred in connected text: "He geib ibriting hi had hi dident kip nating" [He gave everything he had. He didn't keep nothing]; "I help some homeless people the was slipping in the

estrits" [I help some homeless people they was sleeping in the streets.]; "Yo yahto knoe way I dont gib many bicas wen I diden hab many they !deden Helpmy!" [You have to know why I don't give money because when I didn't have money they didn't help me!]; "!Stil you so posto gib many the piople in this vilig ar yor famali!" [Still, you supposed to give money the people in this village are your family!]. When numerous errors such as these occurred within connected text, interpretation was often initially difficult for the scorers. However, when scorers read the papers aloud and could hear the Spanish language influence as reflected in the students' approximations of English words, the effect of phonological variations on meaning was reduced. This supports the earlier finding that having knowledgeable, bilingual scorers was essential to equitable assessment of students' reading/writing abilities.

### *Semantics*

At the semantic level, vocabulary limitations and confusions appeared in three commonly seen patterns: (a) the use of phrases in lieu of a single adjective; (b) the frequent use of nonspecific vocabulary such as "nice," "good," and "bad"; and (c) incorrect use of word forms.

Phrase substitutions appeared primarily in responses to item 1 of one reading passage and item 2 of another, which asked for descriptions of character traits. In describing the main characters, for example, phrases such as "do not do good to beggars" and "not nice to poor people" appeared in place of single adjectives such as "mean" or "stingy" and "gives things to homeless" or "helps them live" instead of adjectives such as "generous" or "helpful." Interestingly, these phrases often suggested a higher level of text comprehension than did single adjectives because they indicated the students had not simply searched the text for words to fill in the blank but rather understood what the characters' actions conveyed.

## Finding 2: The Role of Bilingual Scorers

As mentioned above, when the transition students took the district PBA, the English-only scorers had difficulty scoring the tests because of their unfamiliarity with transition students' writing. In this case the bilingual scorers were able to interpret unconventional responses based on their knowledge of the second language acquisition process. In their evaluation of test responses, they were able to consider how Spanish influenced a student's understanding and use of English orthography, syntax, and vocabulary. Because the test assessed reading comprehension and not writing, it was critical that scorers were able to see beyond problems in form to the underlying content students were communicating. Although we did not have a control group of English-only scorers, we believe the examples below show the importance of bilingual scorers for accurate test interpretation. Certainly, teachers who do not know the phonological and orthographic rules

of Spanish cannot be expected to infer children's intentions in choosing certain spelling patterns, and hence, the target words they may be aiming for (Fashola, Drum, Mayer, & Kang, 1996). Nor can they be expected to disentangle syntax that makes sense from the perspective of another language with which they are unfamiliar.

As demonstrated above, the influence of Spanish was evident at several levels of language form and content. It was at the semantic level, however, that translations from Spanish to English could have caused potential misinterpretation of student responses. For example, one student wrote, "I like wasting money on poor people." As several bilingual scorers recognized, this vocabulary usage suggests that the student translated his/her response from the Spanish *"gastar,"* which can mean either to spend or to waste. The student almost surely meant to say, "I like spending money on poor people," which would be more in keeping with the intended response to the question.

Another example of the semantic influence of Spanish in this item occurred when a student began a dialogue by having one character try to persuade another to give some money to the poor. After getting nowhere trying to convince her to be more generous, the student writes that the first character says, "I think no one can change you. Of your form your act." Again, an English-only teacher scoring the test might wonder what "of your form your act" really means. A bilingual scorer, however, might realize the student is most likely translating the notion of *"manera de ser"* or "form of being," which refers to character traits, not transitory behavior. From this perspective, it is clear that the student was able to take two points of view and sustain them through a dialogue, demonstrating a well-developed sense of character.

On another item that asked students to describe character traits, one student argued that the poor beggar woman "should be shy for that." This word usage suggests that the student was translating from the Spanish word *vergüenza*, which means to be embarrassed or ashamed as well as to be shy. The student most likely wanted to say "She should be ashamed of herself for begging." These answers and many others like them could have been misinterpreted if the scorers had not been aware of the Spanish equivalents that were implied by the students' answers.

There was, of course, a relationship between form and meaning and many of the discussions among scorers revolved around how to score consistently and fairly for meaning rather than form. The scorers recognized that students who had more control over form were able to express their meaning more clearly, especially in items that required extended responses. In looking at text processing patterns suggested by the nature of the students' answers, the researchers' item analysis built on the effort to interpret meaning that the scorers demonstrated.

## Finding 3: Comprehension of English Texts

Looking beyond the surface features of the students' written responses, we next examined text processing patterns suggested by the nature of their answers. The transitional assessment addressed a number of reading comprehension skills including summarizing, identifying main idea, and locating information in text. The items chosen for in-depth analysis, however, were those that required more inferential responses, such as identifying character traits, noting character development, determining the author's purpose, and contrasting points of view. As mentioned earlier, the distribution of holistic scores after the first test administration reflected the transitional students' processing of English text primarily at the literal level, according to the district rubric. One might expect this for first year transition students, given their still-developing skills in decoding English. However, item analysis gave more specific information about what students could do, what they had difficulty doing, and the conditions under which students were able to go beyond basic, literal answers to expand on text and make connections, generalizations, and extensions.

### *Character traits*

The early items in the assessment required students to identify traits of the main characters and to provide text-based justifications for their choice of adjectives. By requiring textual support, the item gave an indication of whether the adjectives used were meaningful to the students or were simply copied from the text in a "match to sample" strategy (this strategy of lifting words from the text that match those in the comprehension questions has been identified as a common approach to comprehension test questions used by second language readers (Anderson & Roit, 1996; Cole & Griffin, 1983). Three frequently occurring response patterns to this item were identified. In the first type of response, difficulty with text processing was evident from the mismatch of adjectives and textual support. In these examples, students may have been able to identify adjectives by lifting key words from the text but, when asked to justify their answers, were unable to do so. For example, the students might identify "stingy" as an appropriate description of a rich character in the story, and then fail to use textual excerpts that demonstrate that they understood the meaning of stingy and what actions proved the character deserved that label.

The second type of response involved the use of nonspecific adjectives coupled with appropriate text support that would earn a 3 (adequate understanding) on the rubric. This could be the result of difficulty with text in-depth processing, suggesting that students got only a general idea of the character's traits. However, as suggested above, the use of vague descriptors such as "nice," "bad," or "good" may also be a reflection of limited vocabulary

rather than limited text processing. In describing one character, students used the vague descriptor "nice" in 20% of the responses. However, in 90% of those cases the students provided appropriate text support that suggests that they did understand the basic description of the character and what evidence was available to support that description. These three responses are illustrative examples of this type of response:

> "Nice": (a) he was nice because he made a home fur them; (b) because he helped the homeless by giving them food; (c) when the señora gave her [referring to the beggar] the onion skins she humpbull [humbly] said *muchas gracias*.

For the first two definitions, more specific adjectives that would have been appropriate include "helpful," "generous," or "kind," but "nice" is not inappropriate, and the evidence given is accurate. In the third definition, "nice" is used in the sense of "polite" and, again, the evidence is appropriate.

The third definition gave the strongest suggestion of accurate text processing by indicating that students had understood the meaning of the adjectives they used and chose these descriptors based on the text. The examples below demonstrate a clear match between specific, rather than generic adjectives and textual support, and would have received a 4 or 5 on the rubric—"clear" or "thorough" understanding of the text. Sixty-four percent of the students responded with this type of response:

> Greedy: Quickly the rich señora threw a cloth over her plate to hide the food piled.
>
> Selfish: There were plenty of meat lots of corn and tortiallas. But the rich señora did not want to waste such fine food on a beggar.
>
> Stingy: Every day poor people came to her house to beg for food and every day she sent them away with nothing.

Some differences were noted in the responses to the character trait items for two characters. In one question students were asked to copy passages from the text that supported their adjectives, but for the other they were asked to use their own words. The match-to-sample strategy for finding textual support would not be as effective in the latter case. Perhaps as a result of being freed from copying exact information from the text, students used 18 different adjectives to describe the character in the second question, including sophisticated adjectives such as convincing, successful, powerful, intelligent, active, and considerate, but only six different adjectives to describe the character in the first question (stingy, bad, mean, selfish, greedy, rude). This stronger use of vocabulary in the second passage may also be a reflection of the text,

which provided much more detail about the character than the fictional piece did about its two main characters. This level of analysis is important in the assessment of language learners where it is often difficult to ascertain which factors plays a role in the students' responses.

*Character development*

Subsequent items required the students to expand on this basic sense of character to make more complex inferences. Asking students to write what the character is thinking about the things she did in her life assessed understanding of character development. Asking students to imagine a character's final thoughts as she drops into a sea of darkness required them to attribute feelings, plans, motivations, and thoughts to the character. Two general types of responses were identified. In the first type, students inferred that the character regretted her actions because of their impact on her. In the second type, students extended their understanding a bit by suggesting that the character felt remorse for hurting other people. Both types of responses demonstrate inference and would have received scores of 3–5 on the rubric.

Type 1 examples (regret):

1. I tink tat she was tinking about tat she didnot give food to the poor people and that she slamd the door to the old laydy, and god puneshd her becues she didnot give food to the poor people and she was stingy
2. If I have gave that old woman more that old onion, now I should be in heaven.

In these examples the students use details from the story to frame their answers, a response that would be considered a type of restatement of the text and thus earn a score of 3. Although this type of response does reflect inference, it is a very basic level given the obvious result of her actions.

Students who went beyond this level ascribed more complex motivations to the Rich Señora and by so doing, indicated their ability to go beyond literal thinking to, "understand connections among story elements and overall meaning," as level 4 of the rubric described, or "make plausible interpretations or generalizations," as characterized by the following examples.

Type 2 examples (remorse):

1. I had been nicer with the beggar womans. I had think of it before. But why was I mean. Poor womans I hope somebody saves me. So I can help beggar womas. Specialy the old one. Now am going to help peeple. But first some one help me.
2. She was thinking that she was greedy to the poor people and mean and she was going to help the poor pleople and she was sorry.

3. I was very selfish I had never given those dry skins I should given meet, corn and tortillas to the old women, I should given food the poor peple ho came for food. I'm sorry, I'm sorry, sorry, sorry, sory, I'm sorry!
4. Oh I should have never rejected beggars I-I-I wish I could live again and fixed my errors. And never give poor people only a slice of old onion. I wish I could just stay poor put have given poor people food before getting poor.

*Author's point of view*

In another item, students are asked to step out of the story and imagine the author's point of view and purpose in response to the question, "Write why you think the author made this happen to the Rich Señora." The analysis of responses to this item revealed that 93% of the students had a sense of the author's purpose and were able to imagine the author's point of view, 71% at primarily a literal level, and 22% at a more inferential level. Two general types of responses were identified. In the first type of response students demonstrated a "basic understanding of story elements" (score of 2–3 on the rubric) and assumed that the author was telling a tale of retribution for misdeeds as shown in the following examples:

1. Because when the beggar woman came to the rich woman's house the rich woman threw a cloth over her plate and she didn't wanted to give food to the people because she didn't want to wasted food for a beggar.
2. I think the author made this happend to her because she was berry stingy and mean and she was berry rich she had enouhgf food to give to poor and for her.

These examples illustrate some of the differences in students' responses and suggest different levels of text processing. Although both are essentially restatements of obvious information in the text—literal interpretations—in the first example, the student used details from the story to describe the character's behavior. In the second example the student generalized from the behavior to character traits—"she was very stingy and mean." Both are examples of building on earlier test items where students are asked to describe the character and give examples of actions that prove the character was as the student described.

The next level of response illustrates the students' ability to go beyond a simple text theme of retribution and restatement of the text to imagine the author's interaction with the reader, suggesting that the author wanted to convey a moral or lesson.

Type 2 examples:

1. I think the author made this happen to the rich señora because it was a lesson to everybody to know that there are people less fortunate then you so don't be so gretty because you never know what will happen next.
2. I think that the author made this happen is for the people to see that if you want to recibe something you also have to share. I also think that happen for people to see that been selffish is not good.

*Contrasting characters' points of view*

In the comparison portion of the assessment, the students are asked to construct a dialogue between the main characters of the fiction and nonfiction pieces. The task requires students to take the point of view of both characters alternately and sustain a dialogue that reflects each of their imagined points of view. There were three general types of responses to this item. The first type of response demonstrated minimal text comprehension or inappropriate departure from the text. The second and third types of responses were of an inferential nature, giving "evidence that the reader is beginning to interpret text by making connections between details and larger meanings" (score 4) or "make plausible interpretations or generalizations" (score 5). Eighty-four percent of the responses to this item were at an inferential level. The examples below illustrate these two higher levels of text processing.

Type 1 and 2 examples:

*Example 1*

Trevor: Do you know how much te people need that food? You are rich, yo'u should at least give a little meat the people are staving. Please give them something.

Rich Señora: Yeah Yeah I'm not going to waste my food let them work so they could buy what they need and I'm not giving nothing to anybody.

Trevor: You have plenty of food they dont have nothing dont you have any sipithy for them.

Rich Señora: Ok, ok, I'll just give some little of meat thats all.

*Example 2*

Trevor: Wy did you youst gave that poor woman a skines onion. Why are you so neine to people. be more nise to them.

Rich Señora: but I just hate when thet women comes beggin for things. the should feall she should be shy for that.

Trevor: but dirint you now if that lady had a famaly thats why you should help.

Rich Señora: No I can't I keape my richnes and my good to myself ok.

In both examples the students used textual information to create a dialogue that was plausible for the characters. They went beyond restating the text to extend the characters in appropriate ways. For example, Trevor's question, "Why?" asks the Rich Señora to reflect on her actions. She responds that she thinks the beggar should be ashamed of herself for begging and Trevor suggests she may have a family and needs help.

There are a number of possible explanations why this item drew high level responses from so many students. According to the TPBA work group, the alternative assessment was constructed so that the items built on each other throughout the text, providing a kind of internal scaffolding. From the workgroup's point of view, the students actually learned from, as well as performed, on the test and as they went through the items. This scaffolding allowed them to see connections between plot and character that were then reflected in this item. The task was also an engaging one in which students had more interpretive freedom. According to the teachers, the task also resembled those used in instruction as students worked toward the standards in their Language Arts classes. They were familiar with projecting character traits through dialogue and using "thought bubbles" to indicate the character's exact words. Although the item came at the end of the assessment when performance could have been affected by fatigue, the opposite appeared to occur. The text processing patterns suggested by the students' answers indicate that students can make inferences and apply them to elaborate on text even though their ability to express themselves clearly is still developing.

## Finding 4: Implications for Instruction

Unlike the district PBA where transitional students' performances revealed global difficulties, the transitional performance-based assessment provided more specific information about the students' strengths and difficulties. The phonological, syntactic and semantic patterns demonstrated in the students' work reflected the influence of their first language, Spanish. When the papers were read out loud, it became clear that students were using Spanish sound/symbol relationships to spell English words. While sounding out words is often useful in Spanish, it is a much less reliable strategy in English. Students would benefit from explicit instruction in the similarities and differences between Spanish and English sound systems and their orthographic representations. In addition, more emphasis should be placed on the conceptual basis for English spelling patterns, as well as memorization of high frequency sight words. Attention to how morphemes

are combined to make different grammatical categories in English is clearly needed (e.g., root word plus *ing*). Understanding the conceptual basis of words (e.g., root words, word families) would be more appropriate than memorizing lists of words. The semantic content of the students' answers indicate that students need instruction to increase lexical variety, vocabulary specificity, and understanding of word forms. A study of Spanish/English cognates would be one useful strategy, as would study of synonyms and the use of a thesaurus.

Analysis of students' text comprehension patterns indicated that students were able to read at an inferential level with an appropriately leveled text, graphic support, and scaffolded test structure. Students were able to use text information to take a variety of viewpoints and make plausible extensions of text information and connect it to personal experience. The use of open-ended questions gave specific information on what the students were getting from the text. In addition to revealing the students' strengths, the test answers also indicated that students need more work on text analysis skills, such as identifying text passages that support ideas, building from recall of detail to inferences, and developing metacognitive skills for processing texts (e.g., using graphic organizers to track multiple characters, and comprehension monitoring).

While the influence of Spanish was an obvious and expected feature of transition students' language, it is important to look beyond this developmental process in language acquisition to the instruction students receive in both languages. Anderson and Roit (1996) point out that instructional practices can exacerbate the difficulties second language readers experience in processing text. They suggest that typical instruction for language minority students emphasizes high frequency nouns, verbs, and adjectives but ignores vocabulary that "carries much of the logic of the language, such as negatives, conjunctions, prepositions and other abstract words" (p. 298) that would enable students to process texts more accurately. Another difficulty bilingual students may experience once they are able to decode English text is the task of distinguishing important from unimportant text segments. Students struggling with text comprehension may focus on noticeable or dramatic portions of a text while failing to understand the relevance of particular text portions to the whole. This difficulty may also be instructionally induced when, in the desire to encourage discussion, teachers allow students to engage in lengthy discussions about trivial aspects of a text.

## Discussion

The development of the Transitional Performance Based Assessment was an effort on the part of one school district to adjust its assessment practices to the local context. While the project affected only 89 students

out of the nearly 600 who take the PBA at fifth grade, the issues raised are pertinent to the assessment of all students, particularly those from linguistic and cultural backgrounds.

The creation of a specific test for transition students provided a great deal of useful information. Findings from the research project concluded that the TPBA showed that students who traditionally performed poorly on district-wide assessments could successfully engage with English texts and demonstrate progress toward the district's language arts standards when the assessment took into account their unique language and literacy profiles and planned for their language development needs. The transition assessment, with its emphasis on assisted performance, enabled the students to demonstrate their ability to comprehend English text. Even those students whose overall performance was low in the holistic scoring demonstrated ability to go beyond concrete responses on certain items.

This detailed analysis of student performance showed that while difficulties with language processing clearly affected performance, the specific difficulties with vocabulary, syntax, spelling, and text comprehension did not preclude engagement with text or larger ideas. In her reflection on the results of the assessment, one of the consultants noted:

> It's so seductive to say "you better get the basics, the parts, before you get the whole" rather than recognizing that it's kind of going back and forth between the two. That kids can deal with big ideas long before they have all the parts. A kid may have a lot of insight into character and miss some kind of a sequential thing. . . .The teachers got so much from the kids and I think it's because of the format [of the alternative assessment], its nature, the kinds of things we ask kids to do. We asked them to do fairly interesting and demanding things but [with] a lot of latitude for personal interpretation.

The presence of knowledgeable scorers capable of interpreting students' responses was a key factor in revealing the students' strengths and weaknesses. Whereas similar performances of transitional students on the district PBA in previous years might have been marked "unscoreable" due to the difficulty with conventions and language expression, these tests were interpretable by readers familiar with the developmental nature of second language acquisition in general and the influence of Spanish in particular.

As districts and states grapple with the challenges of equitable assessments for language minority students, many questions remain unanswered. Specific questions about test construction and administration exist alongside larger questions of the role of language in learning and the responsibility of schools to prepare all students for equal participation in society. In spite of its unanswered questions and unmet challenges, this assessment project can inform other districts facing similar challenges of

equitable assessment. Its results give us a regard for the complexities of making sense of the performance of English language learners on assessments in English.

## Questioning the Rationale of the TPBA

Regardless of the strengths shown in the students' performance, many critics argue that while we may be sympathetic to these features of second language acquisition, the students must eventually perform proficiently in English and efforts such as these to document their strengths in the transition phase diverts attention from the pressing need to move these students forward in English proficiency. Goldenberg (1996) for example, questions the appropriateness of program structures that are specifically designed for students making the transition from primary language to English reading. He cautions "perhaps reifying transition—making it into an explicit phase of students' school program—is precisely the wrong thing to do" (p. 357). He contends that the wide variability in the rate and manner of second language acquisition argue against structuring formal transitional programs. Secondly, he suggests that schools that fail to provide a gradual shift in language instruction and instead abruptly move students to English language instruction before they are prepared may in large part socially construct the "transition" period of second language acquisition. He recommends instead a model of gradual transition that would provide primary language support for longer periods of time and more gradual exposure to English, possibly circumventing the clustering of intense language needs during a particular phase of second language reading development. While the district under study here did have such a program, many districts with fewer language minority students may not be able to provide carefully sequenced programs.

Those who oppose the creation of special assessments for transitional students often advocate reporting students' performance on one assessment used for all students but removing the high stakes associated with those scores for students in the transition process. That adjustment does not satisfactorily answer the need for valid information about such students, nor does it address the ethical issues associated with putting students through potentially frustrating and painful exercises from which they gain little, if any, benefit. In this case, however, for a number of reasons, including the nature of the assessment and the scoring procedures, the district PBA left teachers disheartened with the dismal performances of their students without receiving any useful information for instruction. In contrast, the TPBA gave direction for instruction, as discussed in the previous section.

This study was designed to shed light on the performance of transition students on a test of reading comprehension specifically designed with their linguistic profiles in mind. We specifically wanted to know whether an

alternative test would yield more information relevant for instruction than the district-wide test had done. In considering the generalizability of the findings, a number of issues emerged.

While the above approach yielded much information on the form and quality of responses to selected texts, it did not provide comparative information. A useful next step would be to give subsets of students, matched for particular characteristics, similar types of performance assessments using different texts to help determine the influence of the texts themselves on student performance.

It would also be of interest to have a subset of students take both the district PBA and the transitional PBA to more specifically compare performances. Information on the transition students' educational backgrounds, level of Spanish oral fluency and literacy, and oral English fluency levels would enable us to make more inferences about the relation between these factors and English reading comprehension. Ethnographic methodology could be used to gain information on instructional practices that may have contributed to the prevalence of performance at the literal level.

In addition to these specific research design challenges, the issue of assessing reading through writing poses particular dilemmas for transition students and is a critical limitation of the approach of both the PBA and the TPBA. It is likely that additional means of assessing reading (think-alouds, portfolios, exhibitions, oral discussions) would yield more definitive information on students' ability to meet the standards and on the reading process itself (see Fitzgerald, 1995 for a review of various methodologies used to study second language reading).

Another issue raised by these findings is the feasibility of creating assessment instruments specific to particular groups of students. The cost of test development and the difficulty in finding knowledgeable scorers would present significant problems to many districts. However, as discussed above, the staff development model adopted in this district demonstrated an economical way of accomplishing the test development at the same time that teacher expertise was well used and further developed. The use of knowledgeable district employees kept the cost manageable and provided professional development opportunities for the teachers whose expertise had previously been underutilized. For less equipped districts, pairing bilingual and English-only teachers for scoring purposes may be one solution to this dilemma (possibly even using inter-district teams). In these cases significant attention would need to be paid to the establishment of inter-rater reliability.

The work group realized that one assessment measure could not definitively answer complex questions of why particular response patterns occurred, especially when the assessment was scored holistically. The clustering of scores at the 2 or 3 level may have been due to processing

factors, but could also have been influenced by instructional factors, writing ability, the particular texts chosen, or a combination of factors. Without more information about the students' previous Spanish instruction, more information about results on multiple measures of reading comprehension, and/or a comparison group using other texts, it is still difficult to answer these questions.

Perhaps the most important result of having a specially designed assessment for students in transition was the opportunity to break the cycle of failure and the culture of discouragement over the school performance of language minority students. The Transitional Performance Based Assessment allowed teachers to see what their students were able to do and gave them information on how to build on their strengths and, through appropriate instruction, guide students to more proficient use of English and fuller access to educational opportunities.

## References


Abedi, J. (1997, March). *The impact of linguistic features of the NAEP test items on students' performance on NAEP assessments*. Paper presented at the meeting of the American Educational Research Association, Chicago, Illinois.

Abedi, J. (2001). *Assessment and accommodations for English language learners: Issues and recommendations*. Policy Brief 4. Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, *19* (3), 16–26.

Abedi, J., Leon, S., & Mirocha, J. (2001). *Impact of students' language background on content-based performance: Analyses of extant data. Executive Summary, CSE Technical Report*. Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, V., & Roit, M. (1996). Linking reading comprehension instruction to language development for language-minority students. *Elementary School Journal, 96* (3), 295–309.

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

Baker, E. (2000, March). *Assessment and the English language learner*. Paper presented at the Annual Meeting of the California Association for Bilingual Education, San Francisco.

Bartolome, L. (1994). Beyond the methods fetish: Towards a humanizing pedagogy. *Harvard Educational Review, 64* (2),173–194.

Bialystok, E. (1991). Metalinguistic dimensions of bilingual language proficiency. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 113–140). Cambridge, England: Cambridge University Press.

Calero-Breckheimer, A., & Goetz, E. T. (1993). Reading strategies of biliterate children for English and Spanish texts. *Reading Psychology*, *14* (3), 177–204.

Carrell, P. (1991). Second language reading: Reading ability of language proficiency? *Applied Linguistics, 12* (2), 159–179.

Carrell, P. (1992). Awareness of text structure: Effects on recall. *Language Learning*, *42* (1), 1–20.

Cole, M., & Griffin, P. (1983). A socio-historical approach to remediation. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 5* (4), 69–74.

Collier, V. (1988). The effect of age on acquisition of a second language for school. *New Focus*, (2), 1–8.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In Bilingual Education Office (Ed.), *Schooling and language-minority students: A theoretical framework* (pp. 3–47). Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Cummins, J. (1989). *Empowering minority students*. Sacramento, CA: California Association for Bilingual Education.

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

DeAvila, E. (1997). Setting expected gains for non and limited English proficient students. *NCBE Resource Collection Series No. 8*. Washington, DC: National Clearinghouse for Bilingual Education.

Díaz, S., Moll, L. C., & Mehan, H. (1986). Sociocultural resources in instruction: A context-specific approach. In Bilingual Education Office (Ed.), *Beyond language: Social and cultural factors in schooling language minority students* (pp. 187–230). Los Angeles: Evaluation, Dissemination and Assessment Center, California State University.

Droop, M., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research*, *30* (2), 253–271.

Durán, R. P. (1989). Testing of linguistic minorities. In R. Linn, (Ed.). *Educational Measurement* (3rd Edition). New York: American Council of Education, Macmillan Publishing Company.

Durgunoglu, A., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, *85* (3), 453–465.

Edelsky, C. (1983). *Writing in a bilingual program.* Norwood, NJ: Ablex.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms.* Norwood, MA: Christopher Gordon Publishers.

Fashola, O. S., Drum, P. A., Mayer, R. E., & Kang, S. (1996). A cognitive theory of orthographic transitions: Predictable errors in how Spanish speaking children spell English words. *American Educational Research Journal, 33*, 825–844.

Fichtner, D., Peitzman, F., & Sasser, L. (1994). What's fair? Assessing subject matter knowledge of LEP students in sheltered classrooms. In F. Peitzman & G. Gadda (Eds.), *With different eyes* (pp. 143–159). Los Angeles: UCLA Publishing Center for Academic Interinstitutional Programs.

Figueroa, R. A. (1990). Assessment of linguistic minority group children. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement: Vol. 1* (pp. 671–696). New York: Guilford Press.

Figueroa, R. A., & García, E. (1994). Issues in testing students from culturally and linguistically diverse backgrounds. *Multicultural Education, 2* (1), 10–18.

Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive processes: A review of the research in the United States. *Review of Educational Research, 65* (2), 145–190.

Flood, J., & Menyuk, P. (1983). The development of metalinguistic awareness and its relation to reading achievement. *Journal of Applied Developmental Psychology*, *4*, 65–80.

García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26* (4), 371–390.

García, G. E. (1998). Mexican-American bilingual students' metacognitive reading strategies: What's transferred, unique, problematic? *National Reading Conference Yearbook*, *47*, 253–263.

García, G. E., & Pearson, P. D. (1991). The role of assessment in a diverse society. In E. Hiebert (Ed.), *Literacy for a diverse society. Perspectives, practices and policies* (pp. 253–276). New York: Teacher College Press.

García, G. E., & Pearson, P. D. (1994). Assessment and diversity. *Review of Research in Education*, *20*, 337–391.

Gee, J. (1999). Reading versus reading something: A critique of the National Academy of Sciences' report on reading. *Yearbook of the American Reading Forum, 19,* 1–12.

Genesee, F., & Hamayan, E. (1994). Classroom-based assessment. In F. Genesee (Ed.), *Educating second language children: The whole child, the whole curriculum, the whole community* (pp. 212–239). Cambridge, England: Cambridge University Press.

Geva, E., Shany, M., & Himel, N. A. (1992, April). *Parallel processes in the emergence of first and second language literacy skills*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Geva, E., Wade-Woolley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in different orthographies. *Journal of Reading Behavior, 25* (4), 383–340.

Gibbons, P. (1991). *Learning to learn in a second language*. Portsmouth, NH: Heinemann.

Goldenberg, C. (1996). The education of language minority students: Where are we, and where do we need to go? *Elementary School Journal, 96* (3), 353–61.

Gutiérrez, K., Baquedano-López, P., Alvarez, H., & Chiu, M. (1999). A cultural-historical approach to collaboration: Building a culture of collaboration through hybrid language practices. *Theory into Practice, 38* (2), 87–93.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). Committee on Appropriate Test Use, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. *High stakes testing for tracking, promotion, and graduation*. Washington, DC. National Academy Press.

Jimenez, R. T. (1997). The strategic reading abilities and potential of five low-literacy Latina/o readers in middle school. *Reading Research Quarterly, 32* (3), 224–243.

Jimenez, R. T., García, G. E., & Pearson, P. D. (1996). The reading strategies of bilingual Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly, 32* (1), 90–112.

Kiplinger, V. L., Haug, C. A., & Abedi, J. (2000, April). *Measuring math—not reading—on a math assessment: A language accommodations study of English language learners and other special populations.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Kopriva, R. (1997, March). *Piloting the LEP scorers' training indicators manual*. Paper presented at the meeting of the American Educational Reseach Association, Chicago, Illinois.

Krashen, S. (1996). *Under attack: The case against bilingual education*. Culver City: Language Education Associates.

Kucer, S. B., & Silva, C. (1995). Guiding bilingual students "through" the literacy process. *Language Arts, 72* (1), 20–29.

Laberge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6,* 293–323.

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

Langer, J. A., Bartolome, L., Vasquez, O., & Lucas, T. (1990). Meaning construction in school literacy tasks: A study of bilingual students. *American Educational Research Journal, 70*, 350–354.

Mercado, C., & Romero, M. (1993). Assessment of students in bilingual education. In B. Arias & U. Casanova (Eds.), *Bilingual education: Politics, practice, and research* (pp. 144–170). Chicago: University of Chicago Press.

Mitchell, R. (1992). *Testing for learning*. New York: The Free Press.

Miramontes, O., Nadeau, A., & Commins, N. (1997). Restructuring schools for linguistic diversity: Linking decision making to effective programs. New York: Teachers College Press.

Nagy, W. E., García, G. E., Durgunogulu, A. Y., & Hancin-Bhatt, B. (1993). Spanish-English bilingual students' use of cognates in English reading. *Journal of Reading Behavior, 25* (3), 241–258.

O'Day, J. A., & Smith, M. S. (1993). Systemic school reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 250–311).

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Olson, J. F., & Goldstein, A. A. (1997, March). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* Paper presented at the meeting of the American Educational Research Association, Chicago, Illinois.

Pérez, B., & Torres-Guzmán, M. E. (1996). *Learning in two worlds: An integrated Spanish/English biliteracy approach* (2nd ed.). London: Longman.

Pritchard, R. (1990, December). *Reading in Spanish and English: A comparative study of processing strategies.* Paper presented at the annual meeting of the National Reading Conference, Miami, FL.

Reyes, M. de la Luz. (1992). Challenging venerable assumptions: Literacy instruction for linguistically different students. *Harvard Educational Review, 62* (4), 427–446.

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *TESOL Quarterly, 17*, 199–219.

Shepard, L. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405–450.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38(*5*)*, 553–573.

Sweet, D. (1997, March). *OERI-Sponsored research projects to develop inclusive statewide assessment systems.* Paper presented at the meeting of the American Educational Research Association, Chicago, Illinois.

Thonis, E. (1983). *The English-Spanish connection.* Northvale, NJ: Santillana.

Ulibarri, D. M. (1985). *Standardized achievement testing of non-native English-speaking students in elementary and secondary schools.* Rosslyn, VA: National Clearinghouse for Bilingual Education.

Valdés, G. (1998). The world outside and inside schools: Language and immigrant children. *Educational Researcher, 27* (6), 4–18.

Valdés, G., & Figueroa, R. (1994). *Bilingualism and testing: A special case of bias.* Norwood, NJ: Ablex.

Valencia, S. W., Hiebert, E. H., & Afflerbach, P. P. (1994). *Authentic reading assessment: Practices and possibilities.* Newark, DE: International Reading Association.

Vellutino, F., Scanlon, D., & Tanzman, M. (1990). Differential sensitivity to the meaning and structural attributes of printed words in poor and normal readers. *Learning and Individual Differences*, *2*, 19–43.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Westby, C. (1989). Assessing and remediating text comprehension problems. In A. G. Kamhi & H. W. Catts (Eds.), *Reading disabilities: A developmental language perspective* (pp. 199–259). Boston: College Hill.

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing.* San Francisco: Jossey-Bass, Inc.

Williams, B. (Ed.). (1996). *Closing the achievement gap: A vision for changing beliefs and practices*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wong Fillmore, L. (1989). Language learning in social context. The view from research in second language learning. In R. Dietrich & C. F. Graumann (Eds.), *Language processing in social context* (pp. 277–30). North-Holland: Elsevier Science Publisher B.V.

Wong Fillmore, L. & Lara, J. (1996). *Summary of the proposal setting the pace for English learning: Focus on assessment tools and staff development*. Washington, DC: Council of Chief State School Officers.