

...Given the need for inclusion of English language learners in accountability systems designed for school improvement, school reformers face real challenges in successfully measuring students' progress through large-scale assessment systems.

## Standards-Based Assessment for English Language Learners

Amie Beckett

What is the yardstick by which our students' achievement is measured? At one time, achievement of limited English proficient students was invisible or underestimated in large-scale assessment systems, either because their achievements did not fit the system or because local expectations permitted these students to achieve at levels below their true potential. Students who were not yet proficient in English were routinely exempted from large-scale assessment programs, because these assessments were designed for proficient speakers of English and they provided few, if any, accommodations for non-native speakers. This was not a satisfactory long-term solution for students acquiring English, particularly given the increased emphasis on assessment in school reform and resource allocation. However, to be fair to the students and the educational process, the drive toward large-scale assessments needs to be counterbalanced by an emphasis on opportunities to learn, including ongoing assessment and feedback (Darling-Hammond, 1994).

The decision about how and to what extent to include ESL students in large-scale assessment systems is a challenging one. The state of the art is limited in defining factors critical to assessment of students acquiring English. We do not know the level of proficiency at which second language learners can first be tested in English with accurate results. We do not know the exact role that native language testing plays in large-scale assessment systems. We know little about the impact of different accommodations and the conditions that must be present for them to provide an accurate picture of the students' content knowledge (NCBE, 1997). We are only beginning to measure the inputs that students receive in and out of school, and the importance these play in the development of content knowledge. We do not know the extent to which the first language might interfere with the demonstration of content knowledge in English as a second language at different levels of proficiency and under different conditions.

Given this incomplete picture, it is not surprising that many educators adopt a cautious stance and exempt students from testing, even when the students might be capable of participating successfully. In a study of the National Assessment of Educational Progress (NAEP), Stancavage, Allen, and Godlewski (1996) found that more than 75% of excluded LEP students could have successfully participated in the assessment from which they were exempted. Changes in NAEP requirements now reflect less permissive language in relation to

exemptions, however, and many students who would have previously been exempted from the testing will participate in the future. These requirements reflect a trend toward greater inclusion of LEP students in standards-based accountability systems (National Center for Educational Statistics, 1996).

### **Purposes and Domains of Assessment of English Language Learners**

Shepard (1995) defines an assessment research framework, in which she identifies three major purposes and four proficiency domains for assessment of students acquiring English. The first purpose, assessment for instructional planning within the classroom, is most directly linked with teacher decision-making about students' current functioning levels and the effectiveness of classroom instruction. To provide concrete information relating to the standards, Shepard (1996) emphasizes the need for conceptualizing and developing performance continua relating to the standards. These continua take into account the current English proficiency levels of students, and provide a mechanism for determining where the student is currently functioning in relation to the criteria. The continua also take patterns of second language acquisition into account. Rubrics and other types of formative assessment tools are useful for this purpose. The Southern California Assistance Center (1998) developed ESL standards for reading that reflect performance at different points in the acquisition process. The Managing the Assessment Process (MAP) Project, developed by Teachers of English to Speakers of Other Languages (TESOL) in 1998 (Katz et al., 1998), highlights standards for ESL and provides educators tools for aligning curriculum, instruction, and assessment using the standards. States have been using these ESL standards to interpret their existing standards through the lens of second language learners.

A second purpose of assessment in Shepard's framework is system-level monitoring and accountability. Large-scale assessments are often used for this purpose. They may be used to compare student achievement with a norm sample (normative assessment), against a criterion measure (criterion-referenced assessment), or in relation to performance expectations for a given task (performance-based assessment). To ensure fairness, validity and reliability of results, multiple measures should be included. Several states have developed performance assessments in addition to normative or criterion-referenced tests, to provide a means for comparability of results using different criterion measures. These measures are group administered, and yield aggregate scores which enable educators to define trends in achievement. They are not appropriate as the sole determiner of high-stakes decisions about program placement and exit for individual students, and their use in such decisions somewhat undermines their use in trend analysis. Teaching to the test becomes more of a problem when system-level instruments are used for high-stakes individual accountability.

Shepard's third purpose of assessment is program placement and exit. Most states have requirements for limited English proficient students to be identified and provided special interventions they determine to be appropriate for second language learners. In instructional settings emphasizing the fluid nature of language learning and acquisition and the interplay between the first and second language, such as two-way bilingual programs, these types of assessments assume less immediate importance than programs with a more limited timeline, such as transitional bilingual programs. For

---

**Amie Beckett is an Assistant Professor in Bilingual Education at Texas Tech University.**

limited English proficient students, language proficiency tests are often used as primary measures influencing program and placement decisions.

Corresponding to each of the purposes of assessment are four proficiency domains: Subject matter knowledge, native language and literacy, English language and literacy, and cognitive abilities. All purposes and domains should be included as part of the assessment system. For example, in the proficiency domain English language and literacy, for the purpose of instruction, teachers might use portfolios, rubrics, running records, or measures of writing performance that incorporate knowledge about second language acquisition. For the purpose of system-level monitoring and accountability, a state-adopted, criterion-referenced test of English-language reading might be used. In addition, a state-adopted performance-based writing test might be scored using a standardized rubric incorporating different ESL levels. For the purpose of program placement and exit, an ESL proficiency test could be used.

Assessment tools, whether formal (e.g. standardized tests) or informal (e.g. Informal Reading Inventories), sample the universe of knowledge, skills, and approaches that students need to meet the standards. Effective assessments for students acquiring English tap the most significant and representative elements of the standards-based curriculum, and reflect as authentically as possible the standards on which the assessments are based. They also afford special attention to second language processes and the students' anticipated responses to the tasks they are provided.

### **Standards-Based Accountability Systems**

Increased participation of formerly marginalized groups is characteristic of standards-based school reform. As school districts implement reforms, assessment results, particularly those from large-scale testing, exert greater influence on district needs assessments, resource allocation, district planning and even accreditation in some states. Exemption of ESL students from testing means their needs may not be considered when major decisions are made. From a psychometric perspective, even if only 10%-12% of the students in a district are exempted from district-wide assessments, this presents an inaccurate picture of achievement. At the state or regional level, this distorted picture can affect the validity of state or district comparisons of achievement. Since school reform efforts are often guided by assessment results and intergroup comparisons, exclusion of ESL students from key measures in the assessment system eliminates data necessary to guide the restructuring of schools toward effective instruction inclusive of these students (Shepard et al., 1998). Finally, ESL student exemptions from large-scale assessments remove an important source of data for comparison with classroom achievement. This comparison clarifies the level of English-language proficiency that is really necessary for participation in assessments with native speakers. If large-scale assessment and accountability systems do not penalize schools for the results of ESL students, but disaggregate the scores for analysis, this provides a further incentive to include them in assessments.

Given the need for inclusion of English language learners in accountability systems designed for school improvement, school reformers face real challenges in successfully measuring the students' progress through large-scale assessment systems. Achievement measures designed for fluent English speakers are not likely to yield accurate

data about ESL student achievement in the content areas, unless adjustments are made for the level of English proficiency through accommodations, or alternative measures are used (August and Hakuta, Eds, 1997). For example, Abedi, Lord, and Plummer (1997) found in an experimental study that mathematics assessments relying heavily on complex English language structures and low-frequency vocabulary were significantly less likely than simplified word problems in English to accurately measure ESL students' mathematical understanding. Accommodations enabled the students to demonstrate their content knowledge despite the fact that they did not comprehend complex linguistic structures in English. In general, a good accommodation should eliminate or significantly reduce the language barrier, to enable students to respond more directly to the content, using the most efficient means for the student.

The type of standard to be measured is also important to consider when adjusting assessments for limited English proficient students. School reform efforts encompass several categories of expectations for student learning, including content standards, performance standards, and benchmarks. Content standards refer to what students are expected to learn, and what schools are expected to teach. Performance standards define how students demonstrate their progress toward proficiency in the standard. Benchmarks define the expected proficiencies at a certain point in time (e.g. in Grades 3, 7, and 10). To address these different categories of standards, multiple measures are necessary.

Standards-based reform efforts drive accountability systems which are focused on continuous instructional improvement. Although standards-based accountability can be a powerful tool for instructional improvement for students acquiring English, it can also be highly problematic. Standards developed using only a monolingual English-language framework tend to ignore strengths and needs of English language learners. For example, literacy development in the first language is a better predictor of reading in English as a second language than oral language development in the two languages (Lanauze and Snow, 1989), yet students' first language reading is seldom mentioned in state standards. If it does not appear in the standards, it is not likely to be reflected in the assessment system. A 1996 study by the Council of Chief State School Officers (Lara, 1996) indicated that only six states provided native language assessments aligned with state standards, including reading, and three additional states were developing such measures.

If students acquiring English are to be successful in standards-based instruction, attention to linguistic processes, including second language learning and acquisition is necessary at the level of standards development and interpretation, curriculum design and implementation, instruction, and assessment. These levels are interactive, interrelated, and interdependent, and begin with the development of the standards. If standards are not developed first and assessments drive the standards or curriculum and instruction, ineffective policy decisions or poor long-term achievement are too often the result (Raimi, R.A., and Braden, L.S., 1998). Second language processes and their interaction with the native language also need to be a part of standards development, or an incomplete picture of student achievement is likely to result (August, 1994). Even if standards are based on a monolingual framework only, however, an educator well trained in second language pedagogy can interpret and apply those standards in a way that facilitates the development and

extension of dual language capabilities, by utilizing networks of support outside the classroom and providing materials for instruction in the home language as well as English.

### **Policy Directions for Inclusion of Limited English Proficient Students in Assessments**

The national call for higher standards for all students, prompted by the reauthorization of the Elementary and Secondary Schools Act in 1994, moved practice toward inclusion of limited English proficient students in assessment systems and in the direction of productive use of data to improve instruction. Law following that reauthorization required Title I programs to implement standards and accountability procedures at the state level, and also required that limited English proficient students be included in programming and assessments to a much greater degree than before. Limited English proficient populations in Title I since that time have remained relatively stable, at around 17% of the total Title I population (Sinclair, personal communication; Sinclair and Guttman, 1994).

Title I law stipulates that required annual assessments for program evaluation “provide for the inclusion of limited English proficient students who shall be assessed, to the extent practicable in the language and form most likely to yield accurate and reliable information on what students can know and can do, to determine such students’ mastery of skills in subjects other than English (Improving America’s Schools Act, sec. 1111, 1994).” It also provides that testing be designed to measure group achievement for program accountability, rather than a high-stakes test for individual program entry or exit.

In addition to Title I, Title VII bilingual statutes require that grantees evaluate the extent to which achievement gaps are narrowed or eliminated between limited English proficient students and their native English speaking peers. The National Assessment of Educational Progress (NAEP) in 1995 strengthened its inclusion criteria for limited English-proficient students to include students who have received academic instruction primarily in English for three years; or who are sufficiently proficient to take the English-language test; or who would be capable of taking the test in Spanish, if the test is available. In addition, various accommodations were permissible on the math section of the test (NCES, 1996).

In addition to the federal initiatives, states have responded with a variety of mandates for inclusion of limited English proficient students in standards-based accountability systems, and the use of these data for instructional improvement (Raimi and Braden, 1998; Lara, 1996). In most states, however, students are exempted for several years before they are required to take the test.

### **Accommodations in Assessment Systems**

Once the standards are developed and interpreted inclusive of the needs of students acquiring English, instructional accommodations can enable these students to achieve at the same high levels as their English proficient peers, or the gap in achievement between the two groups can be narrowed significantly. Along with these instructional accommodations, it is also often necessary to make adjustments in assessment procedures, to ensure that content, and not just English language proficiency, can be measured.

Decisions about appropriate accommodations themselves require an assessment. Information about the student’s functioning in the first language, knowledge of the content to be tested, behavior under conditions of testing, and English proficiency level all have a bearing

on the way a student may respond. This information is most accessible through the teacher’s ongoing classroom assessment. Teachers are not able to assess what they cannot see (Igoa, 1995). They need the tools for assessing second language learners, through support and training activities.

Accommodations fall into two major categories: Modifications of the test and modifications of the test procedure. Butler and Stevens (1997) listed a number of possible accommodations for limited English proficient students under each category. Modifications of the test might include assessment in the native language, changes in vocabulary to reflect more high-frequency terms, modification of grammatical complexity, addition of visual supports, such as pictures or objects, use of glossaries in the native language, use of glossaries in English, linguistic modification of test directions, and additional example items. Modifications in procedures might include additional assessment time, breaks during testing, administration in several sessions, oral directions in the native language, small-group administration, separate room administration, use of dictionaries, reading the questions aloud, answers written directly in test booklets, or directions read aloud or explained.

Shepard et al. (1998) found in their study of mathematics performance assessments that many educators have received little guidance in the appropriate use of accommodations. They tend to modify the test administration rather than the test itself. Accommodations are either used with all of the students, or no accommodations are used. Individual needs of students are seldom considered when accommodations are selected. Educators also need to know when an accommodation is effective only for LEP students. If an accommodation results in higher levels of achievement for proficient speakers of English as well as English language learners, then it is not effective in removing a barrier for English language learners alone. More training and discussion about appropriate accommodations are needed, to refine testing practices for limited English proficient students. Discussion to follow highlights selected recommendations.

### **Capacity Building for Meeting the Standards: Using Assessment Results**

1. Large-scale assessments occur only a few times during the year. Despite their influence on policy, they occur too infrequently to provide sufficient information for instructional improvement. Therefore, it is essential to compare results of large-scale assessments to campus-level and classroom assessments. What trends can be noted for instructional improvement toward the attainment of the standards from the various measures? Are certain standards not being met? For example, if students acquiring English consistently demonstrate difficulty writing a topic sentence on classroom as well as large-scale measures, what can be done in writing process instruction and assessment to address this? The students’ native language may employ a different text structure that involves setting the context before stating the topic. It is important to know about the students’ languages and prior experiences, to bridge their knowledge with the goals of the school. Then, it is important to discuss the similarities and differences in text structure directly with the students, providing examples and bridging the two languages. Encourage the students to compare and contrast, view models of the English text structures while appreciating the structures of their native language, observe

storytelling using English text structures, or work with an English-proficient partner on writing topic sentences in English. The modeling and partnerships plus the direct feedback can be very productive, if the student is beginning to be aware of the differences.

2. Classroom, campus-level, and large-scale assessments can also provide the basis for an assessment of additional training and resource needs. After identifying the trends in the data, develop a cohesive plan that targets a few areas at a time. This will in most cases involve prioritizing and identifying short- and long-term goals.

3. Check the accommodations that are permitted and disallowed on the large-scale assessments in your district. Tailor the accommodations to afford each ESL student the best opportunity to demonstrate knowledge in relation to the standards, without providing unfair advantages. Classroom assessments provide a good indication of whether or not the accommodation provided an accurate measure of the students' progress toward meeting the standards.

Finally, McTighe (1996) identified 7 principles of instruction with attention to standards, to use between large-scale assessments.

1. Establish clear performance targets.
2. Strive for authenticity in products and performances.
3. Make criteria and performance standards known.
4. Model excellent performances and products.
5. Teach strategies explicitly.
6. Use ongoing assessments and provide continuous feedback.

When students, educators, and families are focused on standards and can identify goals for attaining them, assessment becomes a reflection of their direction, guiding them to progress. It is important that they be actively involved in self-assessment and peer assessment as well as classroom and large-scale assessments, for optimal growth to occur.

### **Bibliography**

Abedi, J., Lord, C., and Plummer, J.R. (1997). *Final report of language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No. 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

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

August, D., Ed. (1994, Fall). *For all students: Limited English proficient students and Goals 2000*. [Online] Available: <<http://www.ncbe.gwu.edu/ncbepubs/focus/focus10.htm>>.

Butler, F.A., and Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other characteristics* (CSE Tech. Rep. No. 448). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Igoa, C. (1995). *The inner world of the immigrant child*. New York: St. Martin's Press.

Katz, A., Malone, M., and Gottlieb, M. (1998). *Standard bearer: Next steps in assessment from the ESL standards and assessment project*. [Online] Available at <<http://www.tesol.edu/assoc/kl2standards/articles/eslart9808-01.html>>.

Lanauze, M. and Snow, C. (1989). The relation between first- and second-language skills: Evidence from Puerto Rican elementary children in bilingual programs. *Linguistics and Education*, (1), 323-340.

Lara, J. (1996) and August, D. (1996, February). *Systemic reform and limited English proficient students*. [Online] In Council of Chief State School Officers. Available: <[http://data.ccsso.org/pubs.taf1\\_function=detail&publications\\_uidl\\_155](http://data.ccsso.org/pubs.taf1_function=detail&publications_uidl_155)>.

McTighe, J. (1996, December). *What happens between assessments?* [Online] Available at <<http://www.ascd.org/pubs/el/dec96/mctighe.html>>.

National Center for Educational Statistics (1996, July). *Increasing the inclusion of students with disabilities and limited English proficient students in NAEP*. [Online] Available: <<http://nces.ed.gov/pubs/96894.html>>.

National Clearinghouse for Bilingual Education (1997) *High stakes assessment: A research agenda for English language learners*. Symposium Summary. [Online] Available: <<http://www.ncbe~gwu.edu/ncbepubs/repons/highstakes.htm>>.

Raimi, R.A., and Braden, L.S. (1998). *State mathematics standards: An appraisal of science standards in 46 states, the District of Columbia, and Japan*. Washington, DC: Thomas B. Fordham Foundation.

Sinclair, B. and Guttman, B. (1994). *State Chapter I participation and achievement information-1992-93*. Westat, Inc.: Rockville, MD.

Shepard, L. (1995). Assessment of language-minority students. In E.L. Baker, Principal Investigator, *Institutional grant proposal for OERI Center on Improving Student Assessment and Educational Accountability. Integrated assessment systems for policy and practice: Validity, fairness, credibility, and utility*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Shepard, L., Taylor, G., and Betebenner, D. (1998, September). *Inclusion of limited English-proficient students in Rhode Island's grade 4 mathematics performance assessment* (CSE Technical Report No. 486). Los Angeles: University of California. National Center for Research on Evaluation, Standards, and Student Testing.

Stancavage, F., Allen, J., & Godlewski, C. (1996). Study of exclusion and accessibility of students with limited English proficiency in the 1994 Trial State Assessment of the National Assessment of Educational Progress. In *Quality and utility: The 1994 trial state assessment in reading*. Stanford, CA: National Academy of Education.