State methods for measuring achievement of English Language Learners, and what research suggests is most effective

Shannon Oldham Sampson

University of Kentucky
Abstract
The population of English Language Learners (ELLs) has grown tremendously over the last decade. Recent No Child Left Behind legislation has mandated the inclusion of ELLs in large-scale math and language arts assessments administered in English, and states are increasingly allowing accommodations, adaptations and alternative assessments for ELLs. This paper reviews the literature to determine what accommodations and modifications states are making in their assessments for ELLs, the validity of the results from assessments of ELLs in language arts and mathematics, the fairness of the accommodations or adaptations, and how these accommodations or adaptations are likely to affect measures of adequate yearly progress (AYP).
Objectives

The recently enacted No Child Left Behind (NCLB) legislation requires that achievement in language arts and mathematics of all children in public schools in states that receive federal funding be assessed every year from grade 3 through grade 8 and one year beyond. To ensure adequate yearly progress (AYP) of all subgroups of students, the legislation also requires assessment results be disaggregated based on different levels of student poverty, disability, ethnicity, and language. The goal is to have all students in all subgroups perform at a proficient level on state assessments by the year 2014; states have the responsibility of defining proficient (NCLB, 2001). One of the challenges states face in their efforts to comply with these requirements relates to the accommodations or adaptations that can be made to the form and structure of the assessments, as well as to assessment procedures, in order to accurately and measure the achievement and performance of students whose native language is other than English.

This paper reviews the literature on state assessment systems and testing accommodations/modifications in an attempt to answer the following questions:

- What accommodations or adaptations are states making in their assessments for students who are English Language Learners (ELLs)?
- How valid are the results from assessments of these particular students in language arts and mathematics?
- How fair are the accommodations or adaptations?
- Are the accommodations or adaptations likely to affect measures of AYP?
The results of the review will benefit practitioners in education who work with ELLs, giving insight into the current research, as well as researchers interested in expanding on the knowledge base of assessing ELLs.

Perspectives

In the past, students with limited English proficiency (LEP) were often excluded from large-scale assessments and accountability systems because educators believed their inclusion was not in the best interest of these students, arguing even though many English language learners have the content knowledge and/or the cognitive ability to perform successfully on assessments, the testing experience would be extremely frustrating and the picture of English language learners’ content-knowledge would not likely be valid (LaCelle-Peterson & Rivera, 1994; O’Malley & Valdez Pierce, 1994; Shepard, et al., 1998; August and Hakuta, 1997). There is still controversy over when ELLs can and should participate in tests and produce results from which users can draw valid or useful inferences about the ELLs, but the current reality in most states is that more and more students are required to participate in large scale assessments.

Data Sources

Questions related to how states are assessing and providing accommodations for students who are English Language Learners were answered through a review of the latest Council of Chief State School Officers Annual Survey (2003), in conjunction with a review of individual state policies on assessment, as posted on their state department of education websites. Questions related to the fairness, validity, and effectiveness of accommodations and modifications were addressed through a review of research through UCLA’s National Center for Research on Evaluation, Standards and Student Testing.

Results and Discussion

Accommodations and Adaptations

In order to address the challenge of ensuring meaningful assessment of LEP students’ academic proficiency, states are increasingly allowing a variety of linguistic and non-linguistic accommodations or adaptations, as well as providing alternative assessments for students who are English Language Learners. According to the Council of Chief State School Officers Annual Survey (2003), in 2001, only 6 states and jurisdictions reported allowing no accommodations for any of their assessments. Of those allowing accommodations on at least one of their assessments, nearly all allow for non-linguistic setting accommodations such as administration in a study carrel, in a separate room, and/or in small groups. In most cases, if the state allows one of these accommodations, it allows them all. Most also allow modifications in timing or scheduling of assessments, such as extended testing time in one day, frequent breaks, and extending testing sessions over multiple days, although less allow extending the testing session over multiple days.

States vary more on the linguistic accommodations allowed. Many allow oral reading of the test in English, excluding the reading tests. Twenty-eight states allow
translation of directions, and 18 allow oral reading of the test in the students’ native language. Of the 13 states that allow tests to be translated into the students’ native language, 8 states restrict that language to Spanish, while other states such as Minnesota and New York also include Hmong, Vietnamese, Somali, Chinese, Haitian, Creole and Korean. Oklahoma and Utah report allowing for translation into whatever language is necessary provided a translator is available. Arizona, Massachusetts, Minnesota, Oregon, and South Dakota report allowing a bilingual version of the test; the language used is not specified. As for response options, 10 states allow students to respond in their native language and 9 allow students to respond in their native language and English. Twenty-four states allow other accommodations such as word lists, dictionaries, and bilingual glossaries.

Only 19 states allow alternative assessments for one or more components of their tests. These range from alternate portfolios in literacy and math in Arkansas, a test of emerging academic English in Minnesota for students who have been in the U.S. three years or less, alternate assessments which could include native language achievement tests, performance assessment and/or writing samples and other classroom work from the student in New Jersey, and alternate assessments which are standards-based and locally developed in Wisconsin. The Council of Chief State School Officers (CCSSO) predicts that more states will report allowing alternate assessments as a result of the recent No Child Left Behind legislation, but notes that states offering an alternative assessment face the challenge of combining the results from that assessment with the results from the regular assessment.
Validity of Results

Researchers disagree on the validity of the results of LEP students’ language arts and mathematics assessments. In the past, ELLs were often excluded from large-scale assessments and accountability systems because educators believed it was not in the best interest of students to take the tests, arguing that even though many ELLs have the content knowledge and/or the cognitive ability to perform successfully on assessment tasks, the testing experience would be extremely frustrating and the picture of ELLs content-knowledge would not likely be valid (LaCelle-Peterson & Rivera, 1994; O’Malley & Valdez Pierce, 1994; Shepard, et al., 1998; August and Hakuta, 1997). Even though a test may be designed to measure math skills, to a limited English proficient student it becomes a test of English language in addition to, or possibly instead of, math. The Standards for Educational and Psychological Testing (1999) recognizes this in the background section of Standard 9: Testing Individuals of Diverse Linguistic Backgrounds, stating, “For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. Test use with individuals who have not sufficiently acquired the language of the test may introduce construct-irrelevant components to the testing process” (p. 91). Without accommodations, it seems that test users cannot draw valid or useful inferences about the English language learners.

Fairness of Accommodations/Adaptations

One could argue it is not fair for a student to have to take an academic test in a language in which he/she has limited proficiency. In looking at fairness, one must consider if accommodations or adaptations can truly level the playing field by reducing
the linguistic barriers and assessing only the intended content. According to Cummins (2000), “There is little evidence that… accommodations result in equitable assessment” (p.141). Common accommodations such as setting and presentation changes such as small group administrations or short segment test booklets may not be as effective for LEP students whose real hurdle in performing to the best of their abilities is the language barrier (Rivera & Vincent, 1997). Non-linguistic accommodations such as modified time and setting make sense only if the ELLs have some degree of English proficiency (Shepard, et al., 1998).

While non-linguistic modifications may make limited difference for students with higher levels of English proficiency, linguistic accommodations appear to be most effective. Abedi, et al., (1997) found that ELL students who received modified English versions of a math test performed significantly better than those receiving the original items. In 2000, Abedi, Lord, Hofstetter and Baker conducted a study in which eighth graders took a math test under standard NAEP conditions and four different types of accommodations: linguistically modified English version of the test, standard NAEP items with glossary only, extra time only, and glossary plus extra time. In their study, use of a glossary plus extra time was the most effective form of accommodation for both LEP and non-LEP students. The only type of accommodation that narrowed the score difference between LEP and non-LEP students was modified English.

In a study of eighth grade students taking the NAEP science tests (Abedi, Lord, Kim, & Miyoshi, 2000), students were distributed tests in three formats: no accommodation, one booklet with an English glossary and Spanish translations in the margins, and one booklet with a customized English dictionary at the end of the booklet.
ELLs scored highest on the customized dictionary accommodation, and there was no significant difference between the scores of the English proficient students in the three test formats; thus, the accommodation strategies likely did not affect the construct. A study by Abedi, Lord, and Plummer (1997) indicated that when the National Assessment of Educational Progress (NAEP) test items were grouped into long and short items, English language learners performed significantly lower on the longer test items, regardless of the level of content difficulty of the items. Modifying and simplifying the language of the test items has consistently resulted in ELL performance improvement. The Center for Research on Evaluation, Standards, and Student Testing (CRESST) has been able to repeatedly improve ELL performance by 10-20% on many tests by modifying and simplifying language and allowing students to use dictionaries or glossaries. Language simplification thus appears to be a useful accommodation.

In using accommodations, one has to be careful that the accommodations do not overcompensate and compromise the validity of the results. Many would argue that any alteration to the standard administration necessarily alters the validity of the test results and the construct under measurement. Research suggests that assessment accommodations may alter the equivalence of the assessments and give an unfair advantage to recipients (Koretz, 1997; Willingham, et al., 1988). Ideally, accommodations will have no effect on native English speaking students while reducing the language barrier for ELLs (Shepard, Taylor & Betebenner, 1998; Rivera & Stansfield, 1998). Some commonly used accommodations such as extra time may lead to higher scores for many ELLs, but this accommodation also aids students already proficient in English (Abedi, Lord & Hofstetter, 1998; Abedi, Lord, Kim & Miyoshi, 2000).
Research suggests that native-language assessments are a useful accommodation, but only when students have learned content and concepts in that language. Abedi, et al., (1998) found that translating math tests into the native language of LEP students was not helpful if math instruction took place in English. They found that eighth grade Hispanic ELL math students taught in English or sheltered English scored higher on NAEP items in English than their peers who received the same math items in Spanish. Students who were instructed in Spanish performed higher on the Spanish-language math items than those in either standard or linguistically modified English.

In the case of students receiving native language instruction and testing, validity could be compromised because of linguistic differences between the variations of languages. According to Olmedo (1981), translated items may exhibit psychometric properties substantially different from those of the original English items. For languages spoken in numerous parts of the world, Spanish for example, which variation of the language would be the standard for translation? Use of one variation of Spanish when a student has been taught in another could result in inaccurate reflection of the students’ content knowledge. The Standards for Educational and Psychological Testing cautions that translations do not necessarily produce a version of the test that is equivalent in content, difficulty level, reliability and validity to the original version. Further, the test of content of the translated version may not be equivalent to that of the original version, and there are many language and dialectic variations that would need to be included in translation considerations (pp. 92-93.)

Math and science are easier to modify linguistically than assessments of reading ability, because they are generally less dependent on language proficiency. In a National
Center on Educational Outcomes (NCEO) study to assess English reading with native-language support, Anderson, Liu, Swierzbin, Thurlow & Bielinski (2000) provided bilingual accommodations for Limited English Proficient students in reading tests using the Minnesota Basic Skills Test. They found that students tended to use the Spanish as a reference when they did not understand a word in English. They suggested it might be more cost effective and appropriate to instead provide students with English/native-language dictionaries. They concluded that linguistic accommodations may help to a certain extent, but they cannot make a difference if the students have not had the opportunity to learn the content tested, either in their first language or in English. The tests for ELLs should align with the content being taught, and students should be ensured the opportunity to learn the content being tested.

Affect on Measures of Adequate Yearly Progress

To determine if accommodations and adaptations are likely to affect measures of Adequate Yearly Progress (AYP) it is important to look at the group under review. While accommodations may make a difference in reducing the language barrier, especially in math and science, the instability of the subgroup creates a barrier affecting AYP. As noted by Abedi and Dietel (2004), one of the goals under No Child Left Behind is that of redesignating high-performing ELL students as language-proficient students. Exiting the high achieving English learners from the subgroup of ELLs, or Limited English Proficient (LEP) students in this context, results in downward pressure on ELL test scores worsened by the constant addition of new ELL students, who are typically at the beginning stages of English proficiency, and thus low achievers on the tests. A look at the Massachusetts state testing results illustrates this point. Abedi, et al. (2004) report
that all ELLs are required to participate in Massachusetts state testing. The target for the LEP subgroup was 48% proficient or above in English language arts for 2003-2004, 56% for 2004-05, and 64% in 2006-2007. However, the ELL statewide performance did not increase 8% points in English language arts during an entire 5-year period (1998-2003), moving unsystematically between 6% and 12% proficient. These results are not encouraging in the light of expectations for Adequate Yearly Progress toward 100% proficient by 2014. Abedi et al. (2004) suggest that the removal of proficient English learners from the LEP subgroup and the addition of non-English proficient learners is likely the central cause of the flat LEP test scores. How realistic is it for the non-English proficient students who enter U.S. schools in 2013 to be proficient in 2014? A key NCLB goal is for all subgroups, including LEP students, to reach 100% proficiency in English language arts and math; however, by definition, LEP students are not proficient in the state’s assessments. Sec. 9101(25) of No Child Left Behind defines an LEP student as one “whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual the ability to meet the State’s proficient level of achievement on State assessments described in section 1111(b)(3).” If ELL students were proficient in English language arts, they would not be LEP students in the first place; thus, if NCLB goals were attained, the LEP subgroup would cease to exist. This is not only unrealistic; it seems to attach a negative connotation to being a member of that group.

Implications and Recommendations
Abedi, et al. (2004) recommend that all state tests undergo rigorous review for language difficulty and that test questions be modified to reduce the level of unnecessary linguistic complexity and cultural bias. It is necessary to review tests to reduce unnecessary language complexity and improve language accessibility. Furthermore, there is a need to expand scoring and training manuals to include sample LEP responses that reflect benchmark level answers to content questions. Scorers should be trained to be sensitive to native language influences such as code-switching or mixing of languages, transposition of words, phonetic writing, spelling conventions from the native language, merging of words, interchanging of periods and commas. The State Collaborative on Assessment and Student Standards (SCASS) for Limited English Proficiency has developed procedures and materials for more appropriate assessment of English Language Learners called *Guide to Scoring LEP Student Responses to Open-Ended Mathematics Items* (Kopriva & Saez, 1997). This guide would be helpful in informing this work.

It is also important that teachers of ELLs incorporate test-taking skills as part of their curriculum. The SCASS LEP Mathematics assessment training indicators manual (1999) suggests students be given ongoing opportunities to verbalize and write about their understanding in the language of the test so that the items better measure skills and knowledge and not coping skills. As students participate in assessment situations similar to those that they will encounter in state testing, they have the opportunity to learn the culture of U.S. assessment.

Finally, states should consider including performance assessment in the accountability measures. Cummins (2000) recommends that the traditional standardized
norm- or criterion-referenced tests be complemented by alternative assessment procedures that are sensitive to the academic growth and accomplishments of all students (p. 142). The State Collaborative on Assessment and Student Standards (1999) encourages the use of performance activities because they give substantial access to those with kinesthetic, tactile, spatial and related strengths. Also, these generally provide students with background and place vocabulary into a meaningful context, making the language more accessible for English learners. In order to proceed with incorporating alternative assessments, it would be recommended to contact testing coordinators in the states currently offering alternative assessments to obtain information about their experiences, so that those experiences can be build upon and modified to create a model of success.

Conclusion

Each state has a unique way of assessing ELLs. This paper reviewed the broad picture of how states are addressing issues of assessment related to ELLs, then turned to the research that suggests which accommodations are most effective. As states continue to address the educational needs of the nationally growing group of ELLs, it is wise to review the research on accommodations and modifications that appear to be most effective. This review is relevant to teachers who work with ELLs, providing insight into adaptations they can make in classroom assessments to better measure the constructs they intend to measure. The research opportunities in this field are wide open as there is much to be done to create measures of student academic progress that provide valid information to teachers, administrators, parents and community members, and policy-makers.
References


National Center for Research on Evaluation, Standards and Student Testing,

http://education.umn.edu/NCEO/OnlinePubs/MnReport31.html


