

**Assessment and Accountability Provisions of No Child Left Behind for English-Language Learners and Related Issues**

Diane August  
Center for Applied Linguistics

## Summary

### ASSESSMENT

**Language Proficiency Assessments** – Issues with many of the language proficiency tests currently available include:

- How we conceptualize language arts and language proficiency, especially in the areas of reading and writing.
- The degree to which off the shelf tests are aligned with state English language proficiency standards
- How well the English proficiency standards define proficiency levels that will help English-language learners acquire the language skills necessary to meet academic content and achievement standards
- The feasibility and practicality of giving an assessment that requires a substantial amount of time to administer
- How well any one assessment can meet the needs of students from diverse language backgrounds
- The level of resources (both financial and technical) available to states for the development and validation of assessments, and for technical assistance in the use of the assessments.

### **Accommodations for Language Arts and Math Assessments during the 3-5 Year Window**

- There are two broad categories of accommodations--changes in the assessment itself and changes in the standard testing conditions.
- Many of the accommodations in use are questionable because they are not theoretically defensible, do not treat the language difficulties of English-language learners, or lack research evidence.
- Three important factors in considering the merits of accommodations are effectiveness, validity, and feasibility and cost of implementation.
- The only effective and valid accommodation found for English language learners is the use of English glossaries and dictionaries; native-language assessments are only useful for students who received content-area instruction in the native language
- Additional research is needed to fully examine the efficacy of any particular accommodation.

## **Valid and Reliable Language Arts and Math Assessments after 3 Years When Accommodations are No Longer Permitted – Issues**

- Validity and reliability of assessments administered in English to language-minority students may be seriously compromised when the students are not sufficiently proficient in English.
- If the content area assessments ultimately used for school and district accountability purposes are to be administered in English, districts will tend to educate students solely in English. This may be problematic, in that the evidence suggests that bilingual instruction is more effective in building literacy in second language learners.

## **ACCOUNTABILITY**

Several of the important issues related to accountability include:

- Inconsistencies across districts in how English-language learners are defined. States—and districts within the same state—may use different criteria to classify students as ELLs. These inconsistencies in classification affect the accuracy of reporting adequate yearly progress for the ELL subgroup.
- As of 2006, not all states had set all three types of objectives for State Annual Measurable Achievement Objectives (AMAOs) for Title III, and many state AMAOs were vague about how progress or proficiency would be measured.
- The English-language learner subgroup is unlike all other subgroups identified under NCLB in that group membership is temporary. More proficient students leave the subgroup while less proficient students enter it leading to the underestimation of the academic performance of this subgroup.
- The ELL subgroup often overlaps with the racial/ethnic subgroups and with the low-income subgroup. Thus, a low-performing ELL student who comes from a low-income family will affect AYP for up to three subgroups.

## **Assessment and Accountability Provisions of No Child Left Behind for English-Language Learners and Related Issues**

Many organizations support strong academic achievement goals, closing the achievement gap, periodic assessment of students learning, and public reporting of achievement results by population subgroups<sup>1</sup>. Advocacy groups underscore the importance of full inclusion in assessment: “One benefit of an all-inclusive assessment system is that it gives us a more accurate picture of the status of the educational system. When any group of students is systematically excluded from the measurement system, we have a biased picture of education, particularly if the group excluded tends to be low-performing students. Moreover, there is real danger that students who are exempted or excused from assessments will not get the attention and services they need, because nobody is accountable for their progress” (Thurlow & Liu, Harvard Civil Rights Project, 2000).

The following sections describe assessment and accountability issues related English-language learners. See Appendix A for Title III and Title I provisions of the law regarding assessment and accountability.

### **Assessment Issues**

The assessment issues germane to English-language learners relate to English language proficiency testing, the use of accommodated assessments, and validly assessing English-language learners after accommodations are no longer permitted.

***Language Proficiency Assessments.*** The English language proficiency assessments are to be aligned with state standards for English proficiency development in four domains—speaking, reading, writing, and listening. Comprehension, as exhibited through listening and reading, must also be considered. The English language proficiency standards are to be different than English language arts standards yet aligned with them.

The kinds of language proficiency tests currently in use by states vary; some use off the shelf tests that commercial publishers say meet the letter of the law even though they have not been customized for individual states (e.g. Language Assessment Scale as used in Nevada developed by CTB/ McGraw Hill); some have adapted off the shelf tests to their state standards (e.g., Stanford English Language Proficiency Test as used in Arizona and developed by Harcourt Assessment, Inc); some have created their own assessments (e.g. Oregon’s online assessment); and others have used tests developed by consortia financed by the US Department of Education (e.g., the English Language Development Assessment (ELDA) developed through the Council of Chief State School Officers or Assessing Comprehension and Communication in English State-to-State (ACCESS) for ELLs developed by Center for Applied Linguistics under contract with the World-class Instructional Design and Assessment (WIDA))<sup>2</sup>.

---

<sup>1</sup> Forum on Educational Accountability is a coalition of more than eighty national education, civil rights, religious, children's, disability, and civic organizations.

<sup>2</sup> Originally established through a federal grant, the WIDA Consortium consists of fourteen partner states and the District of Columbia: Alabama, Delaware, Georgia, Illinois, Kentucky, New Hampshire, New Jersey, Maine, North Dakota, Pennsylvania, Oklahoma, Rhode Island, Vermont, and Wisconsin<sup>2</sup>. Prior to developing an assessment, the WIDA Consortium developed English language proficiency standards. WIDA has developed Spanish language arts standards and is planning a system of alternate academic assessments for beginning English language learners (ONPAR™). Research and professional

Issues with many of the language proficiency tests currently available include: (1) the degree to which off the shelf tests are aligned with state English language proficiency standards; (2) the feasibility and practicality of giving an individually administered assessment that requires a substantial amount of time to administer (e.g., the ELDA developed with consortia funds requires 4-6 hours to administer); (3) how well any one assessment can meet the needs of students from diverse language backgrounds (e.g., in Montana, 81% of the LEP students are American Indians); (4) and the level of resources (both financial and technical) available to states for development, for validation, and for technical assistance in the use of the assessments. A recent estimate of state spending on No Child Left Behind-related testing was less than \$750 million dollars, out of a total K-12 spending of more than \$500 billion (Merrow, 2007).<sup>3</sup>

A fifth issue relates to how well the English proficiency standards define proficiency levels that will help English-language learners acquire the language skills necessary to meet academic content and achievement standards<sup>4</sup>. While preliminary research from one state suggests that it is possible to use English language proficiency assessments in a meaningful way to index English-language learners' progress towards proficiency on English-language arts and math assessments, there is virtually no research into how scores on state language proficiency assessments link to scores on language arts and other content assessments.

A sixth issue is fundamental to developing sound assessments of language proficiency and relates to how we conceptualize language arts and language proficiency, especially in the areas of reading and writing. In developing linked tests, it is important to consider whether there should be considerable overlap in the distribution of language proficiency as measured on the two kinds of tests or whether there should be an abrupt demarcation between the proficiency levels as measured on the two types of tests. Such an abrupt demarcation in the distributions of ability measured by the two tests might be expected if the two tests measured qualitatively different dimensions of language competence. However, such an abrupt demarcation would be difficult if not impossible to achieve if the two tests measured a single underlying trait that is continuously distributed. Figuring out how to design sound language proficiency assessments is challenging given the current knowledge base about what constitutes knowledge and skills in English language arts independent of proficiency in English, especially in the areas of reading and writing.

#### ***Accommodations for Language Arts and Math Assessments during the 3-5 Year Window.***

The validity and reliability of assessments administered in English to language-minority students may be seriously compromised when the students are not sufficiently proficient in English (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Durán, 1989). This is because the language challenge for second-language learners may interfere with the assessment of the content (e.g., mathematics) they have learned, making their test scores invalid as indicators of content knowledge or achievement (Butler & Stevens, 2001). Recent research on the assessment of

---

development activities complement the WIDA standards and assessment products:

<https://uwosh.courses.wisconsin.edu/d21/orgTools/ouHome/ouHome.asp?ou=7061>

<sup>3</sup> According to John Merrow, this contrasts with the 3 to 4 percent on research and evaluation spent by chemical engineering companies.

<sup>4</sup> According to guidance from the US Department of Education, the English language proficiency standards and English language arts standards are to be linked so that “the English proficiency standards define proficiency levels that will help LEP students acquire the English language skills necessary to meet academic content and achievement standards” (US Department of Education, 2003, p. 8).

students with limited English proficiency has demonstrated a substantial link between students' language proficiency and their performance on content area tests. Specifically, English-language learners are twice as likely to achieve the minimum English language arts standards if they score at the highest levels in English proficiency rather than at the intermediate level; and very few students with low English proficiency score well on the state language arts and math assessments (Francis, et al, 2006).<sup>5</sup> Assessments which have more linguistically challenging content yield the largest performance gaps between English-language learners and native English speakers (Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Lord & Hofstetter, 1998; Abedi, Lord & Plummer, 1997; Pennock-Román, 2006). In order to attempt to obtain valid and reliable test scores for English-language learners, NCLB stipulates that accommodations can be used for three years and native language assessments can be used for up to five years.

The most common practice used by states to include English-language learners in state language arts and math assessments are accommodations. There are two broad categories of accommodations--changes in the assessment itself including changes in directions, items, student response options (e.g., multiple choice) and changes in the standard testing conditions, including how the assessment is presented to the student, how the student is allowed to respond (e.g., orally rather than in writing), equipment or materials to be used, the extent of time allowed to complete the test, and the changes to the environment in which the student takes the test.

Several states (Virginia, Minnesota, Nebraska, New York, and Texas) began using English language proficiency assessments for some English-language learners as substitutes for the regular language arts tests used for accountability under Title I. While Minnesota, New York, and Texas have dropped this practice because the language proficiency assessments and language arts assessments were not comparable, Virginia had not and is currently negotiating with the Federal government to extend the timeline for creating comparable assessments. Last October the US Department of Education funded a consortium of 14 states and the District of Columbia to develop an alternative language arts test for English language learners. The test is expected to be implemented in 2010.

Rivera, Collum, Shafer Willner, and Sia (2006) provide a table of 75 different accommodations currently in use with English-language learners and an elaborate taxonomy for classifying the accommodations. They note that many of the accommodations in use are questionable because they are not theoretically defensible, do not treat the language difficulties of English-language learners or they lack research evidence.

A recent report by the Center on Instruction (Francis, et al., 2006) provides an analysis of the merits of recent accommodations for English-language learners used in large scale assessments. They use three factors in considering the merits of an accommodation. The first factor is effectiveness; the accommodation must lead to improved test scores for the targeted students. Appropriate accommodations will differ based on student characteristics (e.g. levels of English language proficiency and native language literacy, grade level) and instructional context (e.g., participation in native language instruction and familiarity with the type of accommodation being offered such as practice using dictionaries). The second factor is the validity of the

---

<sup>5</sup> In interpreting these results, the author notes that students may need to reach a particular level of performance on the English language proficiency assessments before English language proficiency performance begins to effect performance on the English language arts and math assessments. Alternatively, the floors of the English language arts and math assessments do not reach low enough.

accommodated assessment<sup>6</sup>. An accommodation is valid if it does a good job of predicting students' performance in the area tested and additionally leads to better test scores for only the targeted population (in the case English-language learners). The third factor is feasibility and cost of implementation.

A recent meta-analysis of 11 studies (involving 37 samples of students) indicates that very few accommodations are effective (Francis, et. al, 2006). Findings indicate that customized English glossaries and dictionaries were the only accommodations that were effective across the studies. Providing native-language assessments are useful only when students can demonstrate their content knowledge more effectively in their native language, typically because they received content-area instruction in the native language. Many English-language learners have not been instructed in their native languages in the content areas, limiting the potential utility of native language and/or bilingual assessments. Extra time in which students are given more time than usual to complete a test showed positive effects but these were not statistically different than zero. Simplified English, in which there are linguistic changes in the vocabulary and grammar of test items to eliminate irrelevant complexity while keeping the content the same (e.g., eliminating non-content related vocabulary, shortening sentences and using simple sentence structures where possible, using familiar or frequently used words, active instead of passive voice, and using present verb tenses where possible) did not help English-language learners.

The authors note that, despite the disappointing evidence on test accommodations, it would be a mistake to close the door on accommodations because of these findings. The current research is insufficient to fully examine the efficacy of any particular accommodation. The studies have not experimentally examined the possibility that accommodations are differentially effective for students at different levels of language proficiency. Some studies have attempted to look for such effects after the fact, but studies have not been designed to look at this important question directly. Equally important to the question of differential effectiveness, few studies have been conducted on the effectiveness of accommodations with state assessments. Most studies have been conducted on the NAEP, or have used NAEP items, or items from international assessments.

***Valid and Reliable Language Arts and Math Assessments after 3-5 Years.*** A third issue relates to the provision which requires academic assessments (using tests written in English) of reading or language arts and math of any student who has attended school in the US (not including Puerto Rico) for three or more consecutive school years. In some cases, states may use native language assessments to assess content knowledge in areas other than language arts for two additional years.

As mentioned in the previous section, a major issue is that the validity and reliability of assessments administered in English to language-minority students may be seriously compromised when the students are not sufficiently proficient in English (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Durán, 1989).

Some English-language learners will not be ready to take reading/language arts assessments or math assessments in English after three years, especially if they arrive in school in the upper

---

<sup>6</sup> Validity is defined as the extent to which the accommodation preserves the nature of the construct being measured and thus allows for valid inferences about students' standing on the construct of interest when based on a test score obtained under accommodated testing conditions.

grades and have had very limited schooling in their home country. While states are allowed to give native language assessments for up to five years, developing good native language assessments for students from less-common language backgrounds is very challenging for states given the lack of available resources.

In addition, although it is important to encourage students to learn English, if the content area assessments used for school and district accountability purposes are to be administered in English (rather than in the students native language), districts will tend to educate students solely in English. This may be problematic, in that the evidence suggests that bilingual instruction is more effective in building English literacy in second language learners than is English-only instruction (Francis, Lesaux & August, 2006). There are many other reasons why bilingualism might be valued independent of helping students read well in English including personal fulfillment and economic gain. While states are allowed to use native language assessments for up to five years, they may still be inclined to teach students in English because after the five year period ends many students instructed in content in their native language will have a difficult time demonstrating grade level competence in content in English given their instructional history.

Regardless of the language of testing, if English-language learners are not assessed in content area knowledge, aligned with state standards, it will be much less likely that these students are provided with strong content area instruction.

### **Accountability Issues**

There are several important issues related to accountability. The first relates to inconsistencies across districts in how English-language learners are defined. The No Child Left Behind Act defines English language learners as students who (a) are 3-21 years of age, (b) are enrolled or preparing to enroll in elementary or secondary school, (c) were either not born in the United States or speak a language other than English, and (d) owing to difficulty in speaking, reading, writing, or understanding English, do not meet the state's proficient level of performance to successfully achieve in English-only classrooms. However, states and districts may narrow or broaden this definition, so there is no uniform definition across the country. Consequently, states—and districts within the same state—may use different criteria to classify students as ELLs. These inconsistencies in classification affect the accuracy of reporting adequate yearly progress for the ELL subgroup (Abedi, 2004).

A second issue regards the number of states that had not set all three types of objectives for State Annual Measurable Achievement Objectives (AMAOs) for Title III, and the number of state AMAOs that were vague about how progress or proficiency would be measured (Center on Education Policy, 2006). This makes it very difficult, if not impossible, to make state-to-state comparisons of annual outcomes. States have also been slow in reporting data to districts showing their progress in meeting AMAOs; only 38% of districts with ELLs have received AMAO data from the state.

A third issue is that the English-language learner subgroup is unlike all other subgroups identified under NCLB in that group membership is temporary. As students acquire English proficiency they leave the subgroup leading to the most academically proficient members exiting the subgroup. Concurrently, language-minority students new to US schools who are generally more limited proficient enter the subgroup. The effect of the loss of more proficient students and the gain of less proficient students results in the underestimation of the academic performance of this subgroup. While recent provisions allow district to include English-language learners for

two years after they have become 'English proficient', this provision is probably not sufficient because the majority of students in this subgroup, by definition, will not meet English language arts standards.

Additionally, the ELL subgroup often overlaps with the racial/ethnic subgroups that must be tracked under NCLB, and with the low-income subgroup. Under NCLB, each student must be included in every subgroup for which he or she qualifies. So a low-performing ELL student who comes from a low-income family will affect AYP for up to three subgroups (the ELL subgroup, one racial/ethnic subgroup, and the low-income subgroup). This leads to greater pressure on schools, districts and states to rapidly increase the English proficiency and academic performance of English language learners in order to improve the performance of three subgroups.

## References

Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance* (CSE Technical Report #429). Los Angeles, CA: Center for the Study of Evaluation.

Abedi, J. Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance*. Los Angeles, CA: UCLA Center for the Study of Evaluation/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 Measurements: Issues and Practice*, 19 (3), 16-26.

Abedi, J. (2004). *Considering English language Learners in the No Child Left Behind Act: Reporting Adequate Yearly Progress*. Paper prepared for CEP forum on ideas to improve the *No Child Left Behind Act* accountability provisions for students with disabilities and English-language learners.

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

American Federation of Teachers (2004) English-language Learners and AYP calculations: A catch 22. <http://www.aft.org/topics/nclb/downloads/ellayp.pdf>

August, D. & Shanahan, T. (2007). *Developing literacy in second-language learners*. Mahway, NJ: Lawrence Erlbaum Associates.

Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K–12: Current trends and old dilemmas. *Language Testing*, 18(4), 409–427

Center on Education Policy (2006). *From the Capital to the Classroom: Year 4 of the No Child Left Behind Act*.

Durán, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 573–587). New York: American Council on Education.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., Rivera, H. (2006). *Research-based recommendations for the use of accommodations in large-scale assessments*. Center for Instruction.

Francis, D. Lesaux, N. & August, D. (2006). Language of Instruction. In D.L. August and T. Shanahan (Eds.). *Developing literacy in second-language learners*. Mahwah, NJ: Lawrence Erlbaum Associates.

Merrow, J. (2007, February 14). A Surge Strategy for No Child Left Behind? *Education Week*, pp. 44,32.

Pennock-Román, M (2006). Language and cultural issues in the educational measurement of Latinos. In Lourdes Diaz Soto (Ed.), *The Praeger Handbook of Latino Education*, Portsmouth, NH: Greenwood.

Rivera, C., Collum, E., Shafner Willner, L. & Sia, J.K. (2006). An analysis of state assessment policies regarding the accommodation of English language learners. In C. Rivera and E. Collum (Eds.), *State assessment policy and practice for English-language learners: A national perspective*. Mahway, NJ: Lawrence Erlbaum Associates.

Thurlow, M. & Liu, K (October 2000). *State and District Assessments as an Avenue to Equity and Excellence for English Language Learners with Disabilities*, Unpublished manuscript, University of Minnesota.

U.S. Department of Education *Standards and Assessments, Non-Regulatory Draft Guidance*, March 10, 2003.

U.S. Department of Education. (2004, February 19). Secretary Paige announces new policies to help English language learners (Press release). Retrieved September 28, 2005, from <http://www.ed.gov/news/pressreleases/2004/02/02192004.html>

## **Appendix A**

### **Assessment and Accountability Provisions<sup>7</sup>**

The standards-based accountability reform movement, with its emphasis on high academic standards and high expectations for all students and the use of assessments to measure students' attainment of such standards, has led to mandates to include English-language learners in assessments sooner and more broadly than in the past. Further, it requires that all children make progress toward meeting high standards and that outcome data be disaggregated by language proficiency status so that it is possible to monitor the progress of this subgroup of children.

*Title III.* Title III requires State Education Agencies (SEAs) receiving funds to establish English language proficiency standards, identify or develop, and implement English language proficiency assessments, and define annual measurable achievement objectives (AMAOs) for increasing and measuring the level of 'English development and attainment of English proficiency (US Department of Education, 2003, p. 5). Testing requirements under Title III of NCLB (and Title I) require school districts and states to provide for an annual assessment of English proficiency in the four domains of reading, writing, speaking, and listening. Title III also requires the assessment of comprehension. Although Title III requires SEAs and Local Education Agencies (LEAs) to report a separate score for the domain of comprehension, a separate assessment instrument is not required. The Title III integrated system of standards and assessments requires that language proficiency assessments be aligned with state language proficiency standards, which in turn are to be aligned with state challenging academic content and achievement standards. States are not required to use one particular language proficiency assessment for all school districts in the state. However, if states use multiple English language proficiency assessments, states should "set technical requirements for the assessments, ensure the assessments are equivalent to one another in their content difficulty and quality, review and approve each assessment, and ensure that data from all the assessments can be aggregated for comparison and reporting purposes, as well as disaggregated by English proficiency levels and grade levels" (US Department of Education, 2003, p. 27).

Title III "requires SEAs to hold LEAs accountable for meeting annual measurable achievement objectives (AMAOs). These are state defined achievement targets that states use to evaluate the effectiveness of language instructional programs. The objectives are based on the English language proficiency standards and related to 'LEP' students' development and attainment of English language proficiency. The requirements for annual measurable achievement objectives must: reflect the amount of time an 'LEP' student has been enrolled in a language instruction program, and expected attainment of English language proficiency; set targets for annual increases in the number or percentage of children making progress in learning English, and annual increases in the number or percentage of children attaining English language proficiency by the end of each school year; and set targets for schools and LEAs making adequate yearly progress with respect to 'LEP' students, on assessments in the academic areas" (US Department of Education, 2003, p. 14). Unlike Title I where states for the most part use cross-sectional data to evaluate student progress in meeting standards, under Title III states use

---

<sup>7</sup> Portions of this section have been excerpted from the report of the National Literacy Panel on Language Minority Children and Youth (August & Shanahan, 2006). They are reproduced with permission of Lawrence Erlbaum Associates.

longitudinal data to assess student progress in meeting the annual measurable achievement objectives in the development of English language proficiency.

**Title I.** Title I stipulates that all students, including English-language learners, take academic content assessments in reading/language arts and math in grades 3-8 and one time in High School.. Title I also stipulates that

States may not exempt [limited English-proficient] students from participating in the State assessment system, in the grades the assessment is given, in their first three years of attending schools in the United States. Inclusion in the State academic assessment system must immediately begin when the student enrolls in school (however, see exemption clause for recent arrivals). No exemptions are permitted based on level of English proficiency (U.S. Department of Education, 2003, p.19).

English-language learners who have been in U.S. schools fewer than 3 years, may take the academic content assessments in language arts/reading and math in their native language. The native language assessments must be aligned with the state content and achievement standards. They may also take an accommodated version of the content area assessment. Accommodations on state academic assessments “include, for example, extra time, small group administration, flexible scheduling, simplified instructions, and allowing the use of dictionaries. Other accommodations listed in the non-regulatory guidance include providing audio-taped instructions in the native language, allowing students to respond in either their native language or English using audiotape, and providing additional clarifying information at the end of the test booklet or throughout the test (e.g., synonyms for unclear idiomatic words and phrases)” (US Department of Education, 2003 p. 20). English-language learners who have attended schools in the United States for at least 3 consecutive years (except those living in Puerto Rico) are subject to the same types of assessments, including literacy assessments, in English as native-English-speaking students. In specific situations, districts may use an assessment in a language other than English for up to 2 additional years.

For purposes of Title I, schools are required to show adequate yearly progress (AYP) in making sure that all students achieve academic proficiency in order to close the achievement gap in reading and math by 2014. NCLB requires states to include the academic achievement results of all students, including English-language learners, in school and state AYP calculations.

In 2004, then Secretary Paige announced two new policies related to the implementation of NCLB. The new policies allow limited-English-proficient (LEP) students who are new arrivals to US public schools during their first year of enrollment in US schools to have the option of taking the reading/language arts content assessment in addition to the English-language proficiency assessment. Previously, they were required to take both assessments. They are required to take the mathematics assessment, with accommodations as appropriate. In addition, states are now permitted to exclude for 1 year or one time the results from the mathematics and, if given, the reading/language arts content assessments in AYP calculations. The other new policy change allows states, for the purpose of AYP calculations, for up to 2 years to include in the LEP subgroup students who have attained English proficiency and are no longer considered LEP according to the district/state’s definition. According to the press release (U.S. Department of Education, 2004), the intent of this change is to “give states the flexibility to allow schools

and local education agencies (LEAs) to get credit for improving English language proficiency from year to year” (p. 1).

## **Appendix B**

### **Research on Accommodations**

A recent report from the Center for Instruction (Francis, et al., 2006) explored the effectiveness of seven different types of accommodations: simplified English (15 studies); English dictionary/glossary use (11 studies); bilingual dictionary/glossary use (5 studies); extra time (2 studies); Spanish language tests (2 studies); dual language questions (1 study); and dual language booklets (1 study). Some studies involved multiple accommodations: extra time plus simplified English (2 studies); extra time plus English dictionaries (3 studies); and extra time plus bilingual dictionaries (2 studies).

The results indicated only one type of accommodation had an overall positive, but small effect on ELL outcomes—customized English language glossaries and dictionaries. Assessments that made use of this accommodation added definitions or simple paraphrases for potentially unfamiliar or difficult words to the test booklet (usually on the margins). Another variation of this accommodation consisted of providing computerized tests with built in English glossaries; the computer provided a simple and item-appropriate synonym for each difficult word on the test. The results were based on studies of math and science in either grade 4 or grade 8. Interestingly, studies that did not allow extra time showed a small, but higher effect size (.238 compared with .074). Additionally, there was a slightly higher effect for computerized tests (.284) compared with paper and pencil (.161).

Spanish language assessments showed considerable variability indicating they might be effective for some students but not others. The assessments were not translated, but adapted to preserve the meaning of the original assessment through back translation. Of the two studies that examined native language assessments, there was a large positive effect if students had been instructed in Spanish (1.064), but a negative effect if they had not been. Dual language test booklets, in which a test booklet contains English items on one side and the corresponding native language items on the other side showed similar results, with students instructed bilingually benefiting from the booklets, but not students who were instructed solely in English.

The use of bilingual dictionaries and glossaries where words were written in English and the students' native language or computerized versions with bilingual glossaries built in, did not help overall. However, the results differed by study. According to the authors, "all five effects involved 4<sup>th</sup> or 8<sup>th</sup> grade science, but the two largest effects were of opposite sign, and both came from studies with 4<sup>th</sup> grade English-language learners" (p. 24). Thus the results indicate "the effect of the accommodation may be very different in different contexts and among different populations of students, and may reflect unobserved differences in instruction" (p. 24).

Extra time in which students were given more time than usual to complete a test showed positive effects but these were not statistically different than zero. Simplified English, in which there were linguistic changes in the vocabulary and grammar of test items to eliminate irrelevant complexity while keeping the content the same (e.g., eliminating non-content related vocabulary, shortening sentences and using simple sentence structures where possible, using familiar or frequently used words, active instead of passive voice, and using present verb tenses where possible) did not help English-language learners. Moreover, the effect sizes were consistent across the collection of studies, indicating that context did not matter. However, it is important to note that the pool of students was small and somewhat narrowly focused in terms of grades, content areas, and type of assessment.