This policy brief addresses the inclusion of English language learners (ELLs) in large-scale assessments and ELL assessment accommodations. The inclusion of ELL students creates specific accountability policy challenges. States differ in the students they include and their inclusion policies and accommodation practices, and, at present, inclusion does not address the different language and knowledge skills within the ELL population. Another problem is that ELL test results may be used inappropriately to sort or retain students. Issues of validity, effectiveness, differential impact, and feasibility must be considered in planning the inclusion of ELL students. Research from the Center for Research on Evaluation, Standards, and Student Testing has shown that the only accommodation that narrowed the gap between ELL and non-ELL students was linguistic modification of questions with excessive language demands. Providing extra time, using a glossary of key terms on the test plus extra time, or reducing the language complexity of the test questions resulted in substantially higher scores for ELL and non-ELL students. The brief provides eight recommendations for policymakers involved in the inclusion of ELL students. These include developing a common definition of ELL students, taking students' backgrounds into account, modifying tests or providing accommodations in accordance with educational research, and monitoring and evaluating accommodations closely. (SLD)
Assessment and Accommodations for English Language Learners: Issues and Recommendations

Jamal Abedi

National Center for Research on Evaluation, Standards, and Student Testing
Assessment and Accommodations for English Language Learners: Issues and Recommendations

Jamal Abedi

Today, one out of seven children in the U.S. speaks a language other than English at home. In California, Florida, New Mexico, and Texas, English language learners (ELLs) exceed 10% of the student population. This policy brief addresses the inclusion of ELL students in large-scale assessments and ELL assessment accommodations.

Both federal and state legislation now require inclusion of all students, including English language learners (ELLs), in large-scale assessments based on a number of key factors:

- Inclusion provides a more accurate picture of overall student achievement and growth.
- Inclusion makes individual diagnostic information available to parents of ELL students, their teachers, and school administrators.
- Inclusion can provide evidence that ELL students have reached proficiency and therefore should no longer be considered English language learners.
- Inclusion may allow for specific policies and funding to improve the performance of ELL students.

CRESST partner Lorrie Shepard and colleagues point out that “inclusion signals the commitment of the educational system to support the academic progress of all its students; and it ensures the representativeness of the data reported.” However, inclusion of ELL students creates specific accountability policy challenges:

1. Which ELL students should be included in large-scale assessments is highly variable from state to state as well as between states and other major assessments such as the National Assessment of Education Progress (NAEP).
2. States differ on inclusion/exclusion policies, type and use of accommodations, reporting of accommodated assessments, and related issues.
3. National norm-referenced standardized tests may have been normed using mostly native English speakers. Accuracy of results may be affected for states with large ELL student populations.
4. If ELL performance is low, we may not know whether the cause is due to limited language skills, low content knowledge, or a combination of both.
5. Inclusion does not address different language and knowledge skills within the ELL population. Although ELL students as a group usually underperform compared to non-ELL students, their range of performance is quite broad.
6. ELL test results may be used inappropriately, either to sort or retain students, or to pull children out of their regular classrooms and into less than ideal instructional programs.
The complex ELL assessment environment is made even more challenging with the introduction of accommodations for ELL students. (See Butler and Stevens for a comprehensive list of frequently used accommodations.) Accommodations are intended to level the playing field, that is, to make language less of a factor, or ideally a non-factor, when measuring performance. However, the use of accommodations requires a complex set of practical and technical decisions, and the research in support of these choices is thin. Such decisions should be informed by the following:

- **Validity**: Does provision of accommodation alter the construct of the assessment?
- **Effectiveness**: What accommodation strategies would be the most effective in reducing performance gaps between ELL students and non-ELL students that are due to language factors?
- **Differential impact**: Which student background characteristics impact accommodated assessment?
- **Feasibility**: Which accommodation(s) are more feasible, particularly in large-scale assessments?

Among these guidelines, the most important is validity. Accommodations should reduce the impact of language but not give LEP students an "unfair advantage" over students not receiving accommodations. However, CRESST research using randomized assignment methods suggests that accommodations may threaten the validity of an assessment, either through over-accommodation or by significantly changing the standardized administration conditions under which the assessment was developed.

Abedi and colleagues found that non-ELL students benefited more from certain forms of accommodation than ELL students did. Dan Koretz found that accommodations provided in Kentucky produced unrealistically high scores for some special needs students. See Abedi et al. for a discussion of challenges to the validity of an assessment through accommodations.

### What CRESST Research Tells Us

- Language proficiency strongly relates to test performance. Students designated ELL by their schools score significantly lower than non-ELL students on many science and math questions. However, the performance gap decreases, or even disappears, on math items that have relatively low language demands, such as math computation.

- ELL students who are better readers, as measured by separate reading tests, perform better on questions with high language demands.

- Translating test items from English to a student's native language does not significantly improve ELL performance when the language of instruction is not the student's native language.

- **The only accommodation that narrowed the gap between ELL and non-ELL students was linguistic modification of those test questions with excessive language demands.**

- In addition to language proficiency, other background factors influence ELL performance. These factors include length of time in the United States, overall grades, and student mobility.

- Many accommodations require substantial amounts of additional administrative time that may increase costs substantially.
Recommendations

CRESST research, supported by other research we have reviewed, leads us to make the following recommendations to policymakers and educators involved in the assessment of ELL students. As we accumulate ongoing research, these recommendations will expand.

1. National clarification is needed on a common definition of ELL students. Even more important are criteria for ELL classifications and appropriate accommodations. Comparability between states, as required in current proposals to reauthorize the Elementary and Secondary Education Act, will otherwise be impossible.

2. Translating test items from English to other languages may not be a successful accommodation when ELL students are taught in English. Our data suggest that the language of the assessment should match the student's primary language of instruction.

3. Student background variables, including language background, are strong indicators of preparedness for participation in large-scale assessments. We recommend that states and other large-scale assessments endeavor to collect background information including length of time living in the United States, type and amount of language spoken in the home, proficiency level in English and student's native language, and number of years taught in both languages.

4. Another suitable readiness indicator is a student's proficiency in academic English. An additional benefit of an external language proficiency measure is that it can be used to suggest appropriate accommodations and monitor language progress over time.

5. Modify test questions to reduce unnecessary language complexity during the development and improvement of all large-scale assessment programs. Reducing language complexity helps to narrow the performance gap between native English speaking students and ELL students.

6. Customized dictionaries are a viable alternative to providing traditional dictionaries as accommodations. A traditional dictionary may provide ELL students an unfair advantage on certain types of tests.

7. Feasibility considerations are important. National and state assessments involve a large number of ELL students, so accommodations have substantial costs. Providing dictionaries or glossaries to all ELL students, administering assessments one-on-one to students, or reducing the language complexity of test items may exceed a school, district, or state's capability. Cost-benefit analyses should be considered. At minimum, accommodations' costs should be tracked and evaluated.

8. Intended and unintended accommodations effects must be monitored and evaluated closely. Ideally, accommodations will have no effect on native English speaking students, while reducing the language barrier for ELL students. With states increasingly moving to reward or sanction schools based on test results, evaluating accommodations effects takes on added, schoolwide importance.

Key CRESST Research Finding

Some accommodations are more effective than others. Providing extra time, a glossary of key terms on the test plus extra time, or reducing the language complexity of the test questions resulted in substantially higher test scores for ELL and non-ELL students. Providing a glossary without extra time did not increase ELL performance, possibly due to information overload.
Recommendations (continued)

9. While we dislike concluding that more research is needed, the complexity of accommodations and their relative newness to education accountability programs require additional investigation. For example, we need to know whether accommodations have different effects by background factors or by subgroups (e.g., Spanish, Vietnamese, and Cambodian). Other areas of research we need to investigate are the effects of multiple accommodations and why some accommodations work better at certain grade levels than others. We also need to improve our reporting of assessments when accommodations are provided.

Contact Jamal Abedi at jabei@cse.ucla.edu

Selected References


Please check the CRESST Web site (www.cse.ucla.edu) for a fully referenced version of this policy brief.
NOTICE

Reproduction Basis

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (3/2000)