

Smarter Balanced Assessment Consortium: Guidelines for Accessibility for English Language Learners

Developed by Measured Progress/ETS Collaborative
April 16, 2012

Principle Authors: John W. Young, Mary J. Pitoniak,
Teresa C. King, and Elizabeth Ayad



Smarter Balanced Assessment Consortium: Guidelines for Accessibility for English Language Learners

General Principles

For English language learner students (ELLs) who take large-scale content assessments, the most significant accessibility concern is associated with the nature of the language used in the assessments. Because ELLs have not yet acquired complete proficiency in English, the use of language that is not fully accessible to them in assessments will degrade the validity of the test score interpretations that can be inferred from their results. In extreme cases the use of language on an assessment that is not accessible to ELLs will lead to test scores that have limited to no validity as indicators of the students' content knowledge. These guidelines are intended primarily to inform assessment developers who will be developing Smarter Balanced Assessment Consortium (Smarter Balanced) assessments. Other educational practitioners, including content specialists and testing coordinators, may also find the information contained in this document useful. Note that these guidelines are not intended to guide the development of English language proficiency assessments.

Although there are many validity issues related to the assessment of ELLs, the main threat to validity when assessing content knowledge stems from language factors that are not relevant to the construct of interest. The goal of these guidelines is to minimize or eliminate these factors that contribute to such construct-irrelevant variance. Adherence to these guidelines will help ensure that, to the greatest extent possible, Smarter Balanced assessments administered to ELLs will measure only what they are intended to measure.

In a discussion of language used on assessments, it is important to distinguish between language that is content-related versus language that is not content-related. Language that is content-related includes terminology and wording that is assumed to be covered as part of instruction. For example, the use of words with specific content meanings, such as "slope" when used in algebra or "population" when used in biology, can and should be used to assess content knowledge for all students. In contrast, greater caution should be exercised when including words that are not directly content-related on a content assessment. Because ELLs may have had cultural and social experiences that differ from those of other students, one should be cautious in assuming that ELLs have the same degree of familiarity with concepts or objects in everyday use. Thus, whenever possible, use contexts or objects based on classroom or school experiences rather than ones that are based outside of school. For example, in constructing mathematics items, it is preferable to use common school objects, such as books and pencils, rather than objects in the home, such as kitchen appliances. For example, although most students, including ELLs, will likely be familiar with a refrigerator, the fact that any student may not be is sufficient to increase the potential for construct-irrelevant variance to be associated with a test item that includes reference to a refrigerator.

Smarter Balanced Guidelines for Accessibility for English Language Learners

In situations where the construct of interest includes a language component, the decisions regarding the proper use of language become more nuanced. For example, if the construct being assessed is the ability to explain a mathematical concept to another student, then the decisions must rest on how this construct is defined. If the construct includes the use of certain language skills, such as the ability to explain a concept using an innovative context, then it is quite appropriate—and would enhance validity—to include the assessment of these skills on test items. For assessments in English language arts, there can be uncertainty as to how to properly develop test items that faithfully measure the construct while avoiding the use of inaccessible language for ELLs. As with other assessments, the decisions rest upon the content standards, definition of the construct, and the interpretation of the claims and assessment targets, since these factors will determine what forms a test item can validly assume. For example, if the skill being assessed is interpreting the meanings in a literary text, then the use of original source materials is acceptable. However, the test item itself—as distinct from the passage or stimulus—should be written so that the task presented to a student is clearly defined using accessible language.

The following sections expand upon the main issues for assessing ELLs when they are administered content assessments. Whenever possible, the guidance and recommendations we provide are based on an evaluation of the research evidence regarding that particular issue. At present, however, it is important to understand that, for many issues related to the assessment of ELLs, the current state of research-based understanding regarding best practices is limited. Keeping this caveat in mind, our goal in this document is to provide guidelines and recommendations based on our best understanding of the issues that bear upon the valid assessment of ELLs.

Accessibility Considerations

Using clear and accessible language is a key strategy that can serve to minimize construct-irrelevant variance in test items. As stated previously, one should not simplify language that is part of the construct being assessed. For non-content-specific language, the language of presentation should be as clear and as simple as is practical. These guidelines for the use of accessible language are intended to serve as guidance in the development of test items and are not intended to violate other principles of good item construction. In addition, these guidelines are not intended to replace the professional expertise and judgment of experienced item writers and test developers. Some general guidelines for the use of accessible language are provided below:

- Design test directions to maximize clarity and to minimize the potential for confusion.
- Use vocabulary in test items that is widely accessible to all students, and avoid unfamiliar vocabulary that is not directly related to the construct (August, Carlo, & Snow, 2005; Bailey, Huang, Shin, Farnsworth, & Butler, 2007).
- Avoid the use of syntax or vocabulary that is above the test's target grade level (Borgioli, 2008). The test item should be written at a vocabulary level no higher than the target grade level, and preferably at a slightly lower grade level, to ensure that all students understand the task presented (Young, 2008).
- Keep sentence structures as simple as is possible while expressing the intended meaning. In general, ELLs will find a series of simpler, shorter sentences to be more accessible than longer, more complex sentences (Pitoniak, Young, Martiniello, King, Buteux, & Ginsburgh, 2009).

Smarter Balanced Guidelines for Accessibility for English Language Learners

- Consider the impact of cognates (words with a common etymological origin) when developing test items. More importantly, be particularly aware of false cognates (or more precisely, false friends), which are word pairs or phrases that appear to have the same meaning in two or more languages, but in fact, do not. Spanish and English share literally thousands of cognates, and because the large majority of ELLs speak Spanish as their first language (nationally, more than 75%), the presence of cognates can inadvertently confuse students and alter the skills being assessed by a test item. Examples of false cognates include: billion (the correct Spanish word is mil millones; not billón, which means *trillion*); deception (engaño; not decepción, which means disappointment); large (grande; not largo, which means long); library (biblioteca; not librería, which means bookstore).
- Do not use cultural references or idiomatic expressions (such as “being on the ball”) that are not equally familiar to all students (Bernhardt, 2005).
- Avoid sentence structures that may be confusing or difficult to follow, such as the use of passive voice or sentences with multiple clauses (Abedi & Lord, 2001; Forster & Olbrei, 1973; Schachter, 1983).
- Do not use syntax that may be confusing or ambiguous, such as using negation or double negatives in constructing test items (Abedi, 2006; Cummins, Kintsch, Reusser, & Weimer, 1988).
- Minimize the use of low-frequency, long, or morphologically complex words and long sentences (Abedi, 2006; Abedi, Lord & Plummer, 1995).

In the same way that good content teachers use multiple semiotic representations to convey meaning to students in their classrooms, assessment developers should also consider ways to create test questions using multi-semiotic methods so that students can better understand what is being asked (Kopriva, 2010). This might include greater use of graphical, schematic, or other visual representations to supplement information provided in written form. In addition, if the assessment delivery system allows for the use of audio or dynamic visual representations, these methods should be considered if it will enhance the accessibility of test items for ELLs. Bear in mind that because ELLs taking Smarter Balanced content assessments will have a wide range of English proficiency skills, it is important to consider the accessibility needs of ELLs across the entire spectrum of English language proficiency.

Again, as a reminder, because ELLs by definition have not attained complete proficiency in English, the foremost concern in developing test items for a content assessment is ensuring that the language used is as accessible as possible. Note that the use of accessible language does not guarantee that construct-irrelevant variance will be completely eliminated, but it is the best strategy for ensuring that the scores of ELLs on content assessments will have the same valid interpretations as for other students.

Smarter Balanced Guidelines for Accessibility for English Language Learners

Examples of Linguistically Accessible Items

In this section, we have provided as examples two sets of test items that were previously used in an ETS research study in 2007. The first set consists of original versions of the items, which are released items from a state assessment program. The second set contains modified versions created by ETS assessment developers and researchers with the intent of making the items more linguistically accessible to ELLs. Two item pairs are presented, one for 3rd grade mathematics and the other for 5th grade mathematics. The first item in each pair is from the original set, and the second is from the modified set. Note that these example items are only for the purposes of illustrating these guidelines, as they have not been vetted through the Smarter Balanced content expectations or item specifications.

3rd Grade Mathematics: Original Version

The town of Milburg has 5256 grown-ups and 2987 children. How many people live in Milburg?

- A 7133
- B 8133
- C 8243
- D 8343

3rd Grade Mathematics: Modified Version

One town has 5256 adults and 2987 children. How many people live in this town?

- A 7133
- B 8133
- C 8243
- D 8343

Smarter Balanced Guidelines for Accessibility for English Language Learners

5th Grade Mathematics: Original Version

Sharice scored the following numbers of points in 5 dart games.

88, 96, 112, 135, 144

What is the median of these numbers?

- A 56
- B 88
- C 112
- D 115

5th Grade Mathematics: Modified Version

Look at the 5 numbers below.

88, 96, 112, 135, 144

What is the median of these numbers?

- A 56
- B 88
- C 112
- D 115

Smarter Balanced Guidelines for Accessibility for English Language Learners

Language Translation Issues

Providing an assessment translated into a student's native language is a testing accommodation used by many states. Translated assessments are typically only used for the content areas of mathematics and science, since the use of translated assessments for English language arts would lead to different constructs being assessed in the original and translated versions. A recent review by Pennock-Roman & Rivera (2011) found that, for students with low levels of English proficiency, the use of an assessment translated into Spanish was the most effective testing accommodation. Because of the lack of sufficiently large samples for other native languages, almost all studies of native language versions of content assessments have been limited to Spanish. Note, however, that for students who are instructed in English and not in Spanish, the use of translated assessments seems to have limited effectiveness. The main reason appears to be that unless students are initially exposed to concepts and content vocabulary through instruction, the use of an assessment in a language different from that of their instruction creates problems that mitigate the effectiveness of this approach.

Because the language used in assessments often has specialized meanings, the direct translation of test items can be problematic and is no longer considered to be a good approach for developing translated versions of an assessment. The primary concern is that direct translation could lead to a version that differs in difficulty from the original version of the test item or that the test item does not assess the target construct in the same manner. For example, some common English words, such as "evidence" as used in "Give evidence from the table...," are difficult to translate.

In contrast, *transadaptation*, a combination of translation and adaptation, is a process of adapting the content in an assessment to account for cultural and linguistic differences between two languages (Stansfield, 2003). Assessment developers and translators should take these differences into account in order to ensure that the test item, in either language, provides the same valid and reliable measurement of a student's knowledge. High quality in the transadapted versions of a content assessment can be achieved by having assessment developers work closely in conjunction with professional translators.

At present, one preferred approach to creating translated versions of an assessment can be carried out if a team of assessment developers who are completely bilingual in both languages is available. The content assessment is developed simultaneously in both languages using common item and test specifications. The two versions of the assessment are then compared to ensure that the content, context, skills, and cognitive demands are as identical as possible. In developing these translated versions of content assessments, the major challenge is identifying teams of assessment developers who are fully bilingual for all the languages being considered. This approach may likely be feasible for only very common languages (i.e., Spanish) as the likelihood of identifying assessment developers who are fully bilingual and also experienced in developing translated assessments may be quite difficult.



Smarter Balanced Guidelines for Accessibility for English Language Learners

Concluding Remarks

The primary purpose of these guidelines is to provide assessment developers and educational practitioners with a framework by which to make appropriate decisions in developing content assessments that are fair and valid for ELLs. Within these guidelines, we have provided guidance that is research-based to the greatest extent possible. Lastly, we want to emphasize that these guidelines are not intended to replace the professional expertise and judgment of experienced item writers and test developers, but rather is intended to add to the principles of good item construction that they already know and use.

Smarter Balanced Guidelines for Accessibility for English Language Learners

References

- Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. Mahwah, NJ: Erlbaum.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*, 219-234.
- Abedi, J., Lord, C., & Plummer, J. (1995). *Language background as a variable in NAEP mathematics performance* (CSE Technical Report 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disability Research and Practice, 20*(1), 50-57.
- Bailey, A. L., Huang, B. H., Shin, H W., Farnsworth, T., & Butler, F. A., (2007) *Developing academic English language proficiency prototypes for 5th grade reading: Psychometric and linguistic profiles of tasks* (CSE Technical Report 727). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics, 25*, 133–150.
- Borgioli, G. (2008). Equity for English language learners in mathematics classrooms. *Teaching Children Mathematics, 15*, 185-191.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*, 405-438.
- Forster, K. I. & Olbrei, I. (1973). Semantic heuristics and syntactic trial. *Cognition, 2*, 319-347.
- Kopriva, R. (2010, September). *Building on student strengths or how to test ELs against challenging math (and science) standards when they don't have the English yet*. Common Core State Standards Implementation Conference.
- Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice, 30*, 10-28.
- Pitoniak, M., Young, J. W., Martiniello, M., King, T., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.
- Schachter, P. (1983). *On syntactic categories*. Bloomington, IN: Indiana University Linguistics Club.
- Stansfield, C. W. (2003). Test translation and adaption in public education in the USA. *Language Testing, 20*, 189 – 207.
- Young, J. W. (2008, December). Ensuring valid content tests for English language learners. *R&D Connections, No. 8*. Princeton, NJ: Educational Testing Service.