**Targeted Linguistic Simplification of Science Test Items for English Learners**

**Date:** June 10, 2020

**Author(s):** Tracy Noble1, Stephen G. Sireci2, Craig S. Wells2, Rachel R. Kachchaf3, Ann S. Rosebery4, and Yang Caroline Wang5

**Affiliation:** 1TERC, 2University of Massachusetts Amherst, 3Smarter Balanced, 4Chèche Konnen Center, 5Education Analytics

**Link:** <https://journals.sagepub.com/doi/full/10.3102/0002831220905562>

# Key Findings

Science tests written in English, particularly when they contain unnecessary linguistic complexity, may be measuring English Learners’ (ELs’) levels of English proficiency rather than their science knowledge. In such cases, ELs’ science test scores may not be valid measures of ELs’ science knowledge. This study investigated whether linguistic simplifications of multiple-choice (MC) test items from the Grade 5 Science, Technology, and Engineering Massachusetts Comprehensive Assessment System (STE MCAS) could reduce unnecessary linguistic complexity and improve EL student performance.

*We asked:* Which specific linguistic simplifications to individual items have the greatest effects on the performance of ELs, but not on the performance of non-ELs?

*Findings showed:*

* ELs performed significantly better on two linguistically simplified (vs. original) test items when visual images were added, and on one item that was simplified based on interviews with EL students about the item.
* The three effective changes to individual items all included adding visual images to the *answer choices* for MC items that contained English words likely to be unfamiliar to ELs. Without visuals, ELs performed significantly worse on these items.

# Data

We administered an experimental test consisting of original and linguistically simplified released MC items from years 2004-2010 of the Grade 5 STE MCAS to 310 Grade 5 ELs and 1580 Grade 5 Non-ELs in four MA urban public school districts. We used SIMS, MCAS, and ACCESS data to characterize our sample and to group students.

# Research Methods

A total of four types of linguistic simplification were applied to released MC Grade 5 STE MCAS test items. The first two simplification types involved removing linguistic features found in prior research to interfere with ELs’ performance. The last two simplifications were the most effective. They involved adding visual images to items and incorporating suggestions from EL students who were interviewed about the items.

1. *Low-Frequency Non-Technical Words & Forced Comparison*: Two linguistic features were removed: 1) Low Frequency Non-Technical words, that is, words that occur infrequently in fifth-grade texts and are *not* STE content words (e.g., *hose*, *repeatedly*) and 2) Forced Comparison, a common MC item feature that forces test-takers to compare all the answer choices and select the one that is the “best” or “most likely” (e.g., “Which of the following is *most likely* the result of heating water in a pan?”).
2. *Forced Comparison & Reference Back*: Two linguistic features were removed: 1) Forced Comparison as described above, and 2) Reference Back, which requires students to reread an item to identify information referred to in the question sentence (e.g., “How would the earthworm respond to *this change*?”).
3. *Visual*: One or more visual images (pictures, tables, or diagrams) were added to either the answer choices or the stem of the item. Figure 1 below shows an original item and its simplified version with Visuals added to the answer choices.
4. *Interview-Based*: Linguistic and visual features of an item were changed based on both prior research and on feedback on items from interviews with Grade 5 EL students.



Figure 1. Example of An Original Item and Its Simplified Version

We used a systematic and principled approach to apply the simplifications to 20 original Grade 5 STE MCAS test items. Each item was simplified using one of the four types listed above. Scientists reviewed the simplified items to make sure that the difficulty of the science content was unchanged. Each student was given an experimental test consisting of 10 original and 10 simplified items. No student saw both the original and the simplified version of the same item. Each test also included six “anchor” items, chosen because ELs scored similarly to non-ELs on them. Anchor items appeared in their original form and students’ performance on these items was used as a measure of STE proficiency.

To determine whether our simplifications worked, we conducted statistical analyses of test data at multiple levels. We focus here on our use of Differential Item Functioning (DIF), a measure typically used to determine whether two groups score differently on a specific test item, even though they are matched for proficiency. In this study, we used an innovative application of DIF to determine whether the simplified versions of the items were more or less difficult than their original counterparts, by comparing scores on two different *versions* of an item (original vs. simplified) for *one* group of students (e.g., ELs), matched for STE proficiency using their scores on the anchor items. We expected that the simplified items would be easier than the original items for ELs, but that the simplifications would not affect the scores of non-ELs because they did not impact the STE difficulty of the items. To detect DIF, we used both logistic regression and Lord’s chi-square, an IRT-based method.

# Detailed Results

DIF analyses comparing EL student scores on original versus simplified test items demonstrated that three out of the 20 simplified items led to significant improvements in ELs’ scores (i.e., *R*2 values greater than .07) and had a negligible effect on non-ELs’ scores. These three items included two with the Visual simplification and one with the Interview-Based simplification. All three of these successful simplified items had visual images added to illustrate words in the answer choices, including words that had been previously identified as unfamiliar to ELs (e.g., *ceramic*, *rod* and *hose,* as shown in Figure 1.). We hypothesize that these visuals clarified the meanings of the answer choices of these items for ELs, and thus allowed them the opportunity to choose the correct answer. The two simplification types that involved changes to the language of the item only (Low Frequency Non-Technical & Forced Comparison and Forced Comparison & Reference Back) had no effect on ELs’ scores. This finding is in keeping with evidence from prior research that indicates that linguistic simplifications are most successful when they are comprehensive and involve features across the word, sentence, and item levels, rather than targeting a small number of item features. The importance of adding visuals to answer choices in particular has not been previously reported, perhaps because of a focus of prior research on mathematics test items, which are more likely than science test items to include numbers and symbols, rather than words, in answer choices.

# Implications for Policy and Practice

This study provides support for the idea that linguistically complex STE test items can interfere with the ability of EL students to demonstrate their STE knowledge and skills when answering these items. In particular, the findings suggest that the addition of visuals to MC test items can be most effective when they are added to illustrate answer choices, especially answer choices that include words in English that are likely to be unfamiliar to ELs. Test designers can use these findings to improve test fairness and validity for ELs.

*Authors’ note.* This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grant No. R305A110122 and the Education Research Collaborative at TERC. The opinions expressed herein are those of the authors and do not reflect the opinions of the funding agency. The authors thank Catherine Bowler and Carrie Conaway of the Massachusetts Department of Elementary and Secondary Education for their partnership with this research effort. The authors also thank the students, teachers, and district and school administrators who generously chose to participate in this research. This article is dedicated to them.