
ELL Corner: Linguistic Modifications in Action (Part 3 of 3)

Brooke Norval & Martha Castañeda, Miami University

Abstract: The following is the third installment in a three-part series exploring ways to modify mathematics tasks to make them more equitable for English language learners (ELLs). In this third installment, the authors show explicit examples of how to linguistically modify mathematics test items.

Keywords: English language learners, National Assessment for Educational Progress (NAEP), test items

1 Introduction

In the previous two issues of the *Ohio Journal of School Mathematics*, I explored the gap in mathematics test performance between English Language Learners (ELLs) and non-English language learners (Norval, 2019a) and how that gap can be narrowed by linguistically modifying mathematics test items (Norval, 2019b). In this issue, I provide examples of original NAEP mathematics test items and provide modifications that three different researchers made to the items to make their language more equitable to ELLs. At the time of this writing, I served as a graduate student, lead author, and researcher. I made the initial and final modifications of the test items, while two professors, Dr. Martha Castañeda and Dr. Michael Todd Edwards, served as my mentors, subject-area experts, and researchers. They modified the test items independently.

2 Methodology

For this paper, I selected three questions from the NAEP website's *NAEP Questions Tool*, a database of hundreds of questions from past NAEP assessments. Of the items, one was from the 2009 exam and two were from the 2013 exam. One item was constructed for 4th-grade students, and two were for 8th-grade students. Two were multiple choice, while one was a short constructed-response item. The content classification included two algebra problems and one measurement problem. The items had a variety of complexity level ratings, including low, moderate, and high complexity.

Next, four sets of modifications were made independently. I wanted to gain multiple different perspectives regarding the modification of mathematics problems for ELLs. Modifications reflected (a) a research and literature-based perspective (Norval), (b) a perspective from someone who has extensive experience in teaching mathematics (Edwards), and (c) a perspective from someone who has experience and research knowledge of teaching ELLs (Castañeda). Finally, after reviewing all sets of modifications, a fourth set of modifications was made, taking all perspectives into account.

For the first set of modifications, I reviewed the literature available on linguistically modifying math questions to make them more accessible to ELLs. After paring down the suggestions into the finalized list in Table 1 (see Appendix), I began making modifications to the test questions. Each

item had a description provided by NAEP which stated the “goal” of the problem—for example, Question 2’s description was “Solve arithmetic problem involving time.” In making changes to the problem, I made sure that the new problem retained the goal given by NAEP to ensure construct validity.

For the second and third sets of modifications, I consulted Dr. Todd Edwards, a mathematics education professor in the Department of Teacher Education at Miami University, and Dr. Martha Castañeda, a foreign language education professor in the same department. When making linguistic modifications, I consulted with each faculty member one-on-one. During the consultations, I gave each professor the original NAEP mathematics test items and sat with them while they made their own modifications. I did not show them any modified test items or give them any direction regarding modifications. I recorded their thought processes as field notes as they discussed aloud their proposed modifications. Next, I typed up these field notes into paragraph form, explaining why each faculty member chose to make the modifications that they did.

Lastly, I made a fourth and final set of modifications after taking the perspectives of all three reviewers—research-based, mathematics-based, and foreign language education-based—into account to ensure that each of the final questions were well-rounded, equitable, and valid.

3 Data Collection and Analysis

Each original test item, along with initial modifications by Norval, Edwards, and Castañeda, and my final modifications, are shown in Figures 1 through 15. In addition, a discussion of the rationale for each of the modifications is provided. The collected data for each test item is summarized in Tables 2 through 4 in the Appendix. These tables include a row for each recommendation from Table 1. For each, an “X” is marked in the column for each reviewer who employed that recommendation. Note that I deliberately used the list of recommendations when making my modifications. By contrast, faculty members relied on their expertise in the respective fields as they made modifications. Next, the number of times each recommendation was used was counted in the “Total” column. Tables 5 and 6 summarize these results, with Table 5 sorted by test item and Table 6 sorted by reviewer.

4 Test Items and Modifications

4.1 Test Item 1

NAEP Question ID: 2009-4M5 #13 (Grade 4)

Content Classifications: Measurement (Low Complexity), Difficulty: Hard

A turkey is put in the oven at 10:30 a.m. If the turkey takes 2 hours to cook, at what time should it be taken out of the oven?

- A. 12:15 p.m.
- B. 12:45 p.m.
- C. 1:15 p.m.
- D. 1:45 p.m.

Fig. 1: Original copy of Test Item 1.

Jenna puts a turkey in the oven at 10:30 a.m. She cooks the turkey for $2\frac{3}{4}$ hours. What time should she take the turkey out of the oven?

- A. 12:15 p.m. B. 12:45 p.m. C. 1:15 p.m. D. 1:45 p.m.

Fig. 2: Lead author's initial modifications of Test Item 1.

I changed everything from passive voice to active voice. In doing this, I introduced "Jenna," as it was recommended by Sato et al. (2010) to remove abstract or impersonal presentations. I also modified the question phrase from a conditional form to two separate sentences. The past tense verb was also changed to present tense. In addition, I felt that "at what time...?" should be replaced with "what time...?" for clarity and to simplify a more complex question phrase. The following guidelines from Table 1 were taken into account.

- **Recommendation 2.** Change past, conditional, or future tense verbs to present tense when possible (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010)
- **Recommendation 3.** Construct sentences using active voice instead of passive voice (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 14.** Break conditional clauses into separate sentences (Abedi, 1995; Sato et al., 2010)
- **Recommendation 17.** Change complex question phrases to simple question words (Abedi, 1995)
- **Recommendation 25.** Remove abstract or impersonal presentations (Abedi, 1995)

It takes $2\frac{3}{4}$ hours to cook a turkey. If you start cooking at 10:30 a.m., when will the turkey be done?

- A. 12:15 p.m. B. 12:45 p.m. C. 1:15 p.m. D. 1:45 p.m.

Fig. 3: Dr. Edwards' modifications of Test Item 1.

Dr. Edwards changed the phrasing of "The turkey takes $2\frac{3}{4}$ hours to cook" because the turkey does not "take time," the cooking "takes time." He also changed the question phrase to remove "at what time" to omit extraneous words. Finally, he changed the phrasing of the second sentence to "If you start cooking ..." to put "you" as the subject of the sentence to put the reader in the test item.

Mary's mom dropped her off at school at 10:30 a.m. Mary stayed at school $2\frac{3}{4}$ hours and her mom came back to school to pick her up. What time did Mary's mom pick her up from school?

- A. 12:15 p.m. B. 12:45 p.m. C. 1:15 p.m. D. 1:45 p.m.

Fig. 4: Dr. Castañeda's modifications of Test Item 1.

Dr. Castañeda changed the context of the question to be more general and not specific to certain cultures. She felt that this context, a school setting, drew upon students' background knowledge

more readily than the original topic, noting that people in some cultures and countries don't have ovens in their houses (e.g., parts of China). She also separated the conditional clause into two separate sentences. Lastly, she started the sentences with the subject, removing the passive voice while making the item more personal and less abstract.

Jenna starts cooking a chicken at 10:30 a.m. She cooks the chicken for $2\frac{3}{4}$. What time does she stop cooking the chicken?

- A. 12:15 p.m. B. 12:45 p.m. C. 1:15 p.m. D. 1:45 p.m.

Fig. 5: Lead author's final modifications of Test Item 1.

All three reviewers chose to remove passive voice construction so that the turkey was no longer the subject of the sentence. All three reviewers also simplified the "at what time...?" question phrase to "what time...?" Both myself and Dr. Castañeda removed the conditional clause from the test item and replaced it with two separate sentences. We noted that the test item could be culturally biased. As earlier noted, ovens are not a household appliance in all cultures. In addition, chicken is a much more common food in most cultures than turkey, so I modified the wording accordingly.

4.2 Test Item 2

NAEP Question ID: 2013-8M3 #14 (Grade 8)

Content Classifications: Algebra (Moderate Complexity), Difficulty: Medium

Archaeologists measure the lengths of certain bones to estimate a dinosaur's height. When the length of the tibia, or leg bone, is known, a dinosaur's height can be estimated by the following formula, where t and h are in centimeters.

$$h = 73 + \frac{2}{5}t$$

If the length of the tibia of a certain dinosaur is 400 centimeters, what is its estimated height in centimeters?

- A. 402.5
B. 473
C. 475.5
D. 1,000
E. 1,073

Fig. 6: Original copy of Test Item 2.

Scientists measure the length of an animal's bone to estimate the animal's height. The length of the bone is b . Scientists estimate the height, h , of the animal with the equation below:

$$h = 73 + \frac{2}{5}t$$

If the length of the bone is 400 centimeters, what is the height of the animal in centimeters?

- A. 402.5 B. 473 C. 475.5 D. 1,000 E. 1,073

Fig. 7: Lead author's initial modifications of Test Item 2.

I replaced the word "archaeologist" with "scientist," "dinosaur" with "animal," and "tibia" with "bone" in order to avoid low-frequency words. Much of the text in the second sentence was removed as it was irrelevant. Most of the sentences were shortened. Passive voice in the second sentence was changed to active voice. The second sentence was compound, so it was separated into two sentences for simplicity. Finally, the last conditional sentence was broken down into two separate sentences. The following guidelines from Table 1 were taken into account.

- **Recommendation 1.** Reduce sentence length and complexity (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 3.** Construct sentences using active voice instead of passive voice (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 4.** Avoid irrelevant words or phrases (Sato et al., 2010)
- **Recommendation 5.** Vocabulary should be at or below grade level (Fairbairn & Fox, 2009; Sato et al., 2010)
- **Recommendation 6.** Use high-frequency words (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 13.** Break compound sentences into two separate sentences (Abedi, 1995; Sato et al., 2010)
- **Recommendation 14.** Break conditional clauses into separate sentences (Abedi, 1995; Sato et al., 2010)

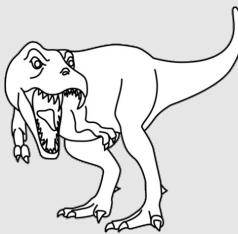
Dr. Edwards added a graphic cue to help readers who may not know the word "dinosaur." In addition, he deemed the phrase "tibia, or leg bone," was unnecessarily wordy since the reader does not have to know that the tibia is a leg bone to solve the task. The word "tibia" adds extra technical, uncommon words that make the test item less accessible. Leg length was used since those words are more common. The word "archaeologist" was omitted since it didn't change the mathematics of the task. Also, he kept the units of "centimeters" in the task, but removed many of its occurrences to eliminate redundant words. In addition, he used the variable l instead of t to reflect the replacement of the word "tibia" with "leg length."

Dr. Castañeda defined "archaeologist" since it is not a commonly used word. She also changed "tibia" to "leg bone" because "tibia" is an academic word, and the test item is not meant to assess academic language or science. She added "...centimeters high" to the answer choices for clarification. In addition, she added a person "Jamie" to the context to make it more relatable and personal for the reader. Finally, she provided the synonym "tall" for "high" to make the definition of "high" clearer to the reader.

All three reviewers reduced the length of the test item by removing unnecessary wording. We also all removed the word "tibia," as it was unnecessary and uncommon academic language. Dr.

The height of a dinosaur can be estimated using the formula

$$h = 73 + \frac{2}{5}l$$



where h is height and l is the length of a leg. If length is 400 centimeters, what is the estimated height?

- A 402.5
- B 473
- C 475.5
- D 1,000
- E 1,073

Fig. 8: Dr. Edwards' modifications of Test Item 2.

Jamie, an archeologist who studies dinosaurs, found a dinosaur leg. Jamie knows that he can use the length of the leg bone (l) to calculate the dinosaur height (h). If the leg (l) of the dinosaur is 400 centimeters long, how tall or high (h) was the dinosaur?

$$h = 73 + 2.5l$$

- A 402.5 centimeters high
- B 473 centimeters high
- C 475.5 centimeters high
- D 1,000 centimeters high
- E 1,073 centimeters high

Fig. 9: Dr. Castañeda's modifications of Test Item 2.

The length of an animal's leg bone is b . Jamie, a scientist, calculates the height, h , of the animal with the equation below:

$$h = 73 + 2.5b$$

The length of the bone is 400 cm. What is the height of the animal?

- A. 402.5 cm
- B. 473 cm
- C. 475.5 cm
- D. 1,000 cm
- E. 1,073 cm

Fig. 10: Lead author's final modifications of Test Item 2.

Edwards and I both removed the word “archaeologist” completely, while Dr. Castañeda defined it for the reader. Dr. Edwards reduced the length of the test item even further than I had in my first modification without making it harder to understand, so I used this technique as inspiration for the final modification. In addition, Dr. Castañeda replaced the word “estimate” with “calculate” as “calculate” is a more common word, which I included in the new modification. I also took Dr. Castañeda’s suggestion to include Jamie as a character to make the test item more personal. Finally,

I agreed with Dr. Edwards' proposal to remove repeated use of the word "centimeter," and I have replaced it with just "cm." In doing so, I added "cm" to the end of all answer choices as did Dr. Castañeda.

4.3 Test Item 3

NAEP Question ID: 2013-8M7 #12 (Grade 8)

Content Classifications: Algebra (Moderate Complexity), Difficulty: Hard

Old Faithful is one of hundreds of geysers in Yellowstone National Park. Predicting when Old Faithful will erupt next can be done by timing the previous eruption.

If an eruption lasts t minutes, then the next eruption will occur approximately $12.5t + 33$ minutes after the eruption ends. If the previous eruption lasted 6 minutes and ended at 1:23 P.M., when is the next eruption expected to occur?

Answer: _____

Show how you found your answer.

Fig. 11: Original copy of Test Item 3.

An explosion lasts t minutes. The next explosion happens $12.5t + 33$ minutes after the first explosion ends.

- The first explosion lasted 6 minutes.
- The first explosion ended at 1:23 P.M.

What time will the next explosion happen?

Answer: _____

Show how you found your answer.

Fig. 12: Lead author's initial modifications of Test Item 3.

Several introductory phrases were removed since they were irrelevant to the construct being assessed and added to the cognitive load of the reader. Low-frequency words such as "geyser" and "eruption" were removed, as well as construct-irrelevant proper nouns and references to American culture such as "Yellowstone National Park." The sentence length of the remaining text was significantly reduced. In addition, conditional clauses were split into separate sentences. Some information was reconfigured as a bulleted list for easier reading. Finally, the wording of the question phrase was made clearer to the reader. The following guidelines from Table 1 were taken into account.

- **Recommendation 1.** Reduce sentence length and complexity (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 4.** Avoid irrelevant words or phrases (Sato et al., 2010)
- **Recommendation 6.** Use high-frequency words (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 8.** Avoid proper nouns that are construct-irrelevant (Sato et al., 2010)
- **Recommendation 12.** Remove unnecessary introductory phrases (Abedi, 1995; Sato et al., 2010)
- **Recommendation 14.** Break conditional clauses into separate sentences (Abedi, 1995; Sato et al., 2010)
- **Recommendation 15.** Use bulleted lists (Sato et al., 2010)
- **Recommendation 16.** Avoid references to American culture or holidays (Sato et al., 2010; Shaftel et al., 2006)
- **Recommendation 17.** Change complex question phrases to simple question words (Abedi, 1995)

A geyser erupts every $12.5t + 33$ minutes. The last eruption was 6 minutes and ended at 1:23 p.m. At what time will the next eruption begin? Show how you found your answer.

Fig. 13: Dr. Edwards' modifications of Test Item 3.

Edwards removed “Yellowstone National Park,” “Old Faithful,” and facts about them that were not mathematically necessary. “Old Faithful” was replaced with “geyser.” The conditional clause was split into 2 sentences to make the passage more easily understood. Lastly, the phrase “If an eruption lasts t minutes . . .” and the word “approximately” were deemed unnecessary and removed.

Your neighbor down the street plays music every day and there seems to be a pattern to when she plays the music. You can predict when she will play music by calculating when she has played music in the past.

If the music lasts t minutes, then the next time she will play music will occur approximately $12.5t + 33$ minutes after the music ended. If the previous time she played music lasted 6 minutes and ended at 1:23 P.M., when is the next time she will play music?

Answer: _____

Show the formulas and work you did to find your answer.

Fig. 14: Dr. Castañeda's modifications of Test Item 3.

Dr. Castañeda replaced the context of a geyser at a national park with an everyday context of someone playing music. She believes that the average English Language learner will not have background knowledge regarding geysers and possibly national parks as these may not exist in their countries of origin. In the process, she removed the academic language such as “geyser” and “eruption.” The original context contained too many references to American culture, as many people

from other cultures may not be familiar with Yellowstone National Park. In addition, many people of low socioeconomic status cannot afford to travel across the country to go to Yellowstone National Park, which was a further layer of bias in the test item. She also considered using the variable m for minutes instead of t , but decided to stick with t . Finally, she changed the instructions to show work at the end of the test item to be more specific.

Lorena sings a song every day. Her song is m minutes long. She plays the next song $12.5m + 33$ minutes after the first song ends. The first song lasted 6 minutes. It ended at 1:23 P.M. What time will the next song start?

Answer: _____

Show how you found your answer.

Fig. 15: Lead author's final modifications of Test Item 3.

Both Dr. Castañeda and I removed the word "geyser," and Dr. Castañeda changed the context of the test item completely to something more relatable. I took this as inspiration for my final modification, using someone playing a song as opposed to an explosion. All three reviewers removed the context of Yellowstone National Park, as it included unnecessary wording and was not accessible or understandable for all students. I also used the name "Lorena" as had been suggested by Dr. Castañeda in her modification for Question 3. In addition, although Dr. Castañeda considered using the variable m in place of t , she ultimately decided not to in order to preserve the integrity of the original test item. However, I ended up using m instead of t for the final modification, as I felt that it made more sense for the reader.

5 Discussion

Looking at Table 5 in the appendix, it is clear that there is some variance in the popularity of each linguistic recommendation. The most popular one, number 14, *Break conditional clauses into separate sentences* (Abedi, 1995; Sato et al., 2010), was used a total of 9 times among reviewers. Other popular linguistic recommendations include number 4, *Reduce sentence length and complexity* (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006), which was used a total of eight times, as well as number 3, *Construct sentences using active voice instead of passive voice* (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006), and number 6, *Use high-frequency words* (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006), both of which were used seven times each.

There were ten linguistic recommendations that were not used by any reviewers—namely 7, 9, 10, 11, 18, 19, 21, 22, 23, and 24. This does not mean that the recommendations weren't worthwhile. Only three test items were analyzed, and not all linguistic features that were referenced in the list of linguistic recommendations appeared in the test items. For example, there were no gerunds, so number 11 could not be used as a modification strategy with the selected items. However, the unused recommendations are important because many other items contain gerunds, hyphenated words, negative questions, and so on, and the reader should be aware of these linguistic features and approach them with caution.

Finally, some techniques employed by faculty members were not specifically listed as a recommendation in the literature that was reviewed. However, these techniques still bring valuable insight in modifying test questions for ELLs, and all of the faculty members' reasoning for modifications

is included. For example, Dr. Castañeda sought to provide a real-world context in some of the problems to make them relatable to ELLs. This was not a technique directly suggested by the literature, but it provides valuable insight and may be crucial in helping ELLs comprehend test questions.

6 Conclusion

Although some may assume that mathematics is a universal language, the gap in mathematics standardized test performance between ELLs and non-ELLs is significant. This may be the case when tests inadvertently assess reading rather than math concepts. Linguistically modifying test items is an effective method of reducing this gap. This research and similar research must be used to inform practice, teacher preparation, and test development. Current and future educators must understand that language affects performance on tests and must learn to create tests that are equitable to ELLs in order to ensure equality in their own classrooms. Universities and colleges must make sure that their teacher preparation programs bring up these issues to their students and teach students how to modify and create test items that ELLs can easily read and access. Standardized test creators must critically analyze their test items from an ELL perspective and ensure that they are equitable for students of all cultures and languages. It is clear that standardized test creators have a long way to go before this is achieved, as the test items selected for this paper have all been from the past eleven years and still include linguistic aspects that are unnecessarily challenging for ELLs.

Further research could be done in several ways. For example, more faculty members could be interviewed other than the two who were interviewed for this paper. In addition to university educators, other perspectives could be taken into account—test makers, practicing secondary teachers, experts in linguistics, current and former ELLs, and so on. Furthermore, actual testing of the original items versus the modified items could be done on a sample of ELL and non-ELL students to accurately measure the efficacy of this particular set of linguistic modifications. From this, further analysis could be done on an item-by-item basis to determine the efficacy of certain types of linguistic modification over others.

Overall, we hope that this series of three articles in the ELL Corner will cause classroom math teachers and test creators to think critically about the way they craft their test items to ensure that the test is equitable to ELLs. The ELL population continues to grow in number and become a greater percentage of students in American classrooms year after year. Thus, teachers must be sure that they thoughtfully consider ELLs in the classroom to ensure that they have the same opportunities to succeed as non-ELL students. Not only will ELLs perform better on tests, “simplification of linguistic structures . . . results in better performance by English language learners and other students who are not good readers” (Kiplinger, Haug, & Abedi, 2000, p. 13).

References

- Abedi, J. (1995, July). Language background as a variable in NAEP mathematics performance. *NAEP TRP Task 3d—Language background study. Final deliverable.* (Rep.). Retrieved <https://eric.ed.gov/?id=ED404176> (ERIC Document Reproduction Service No. ED404176).
- Fairbairn, S. B., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers - Essential considerations for test developers and decision makers. *Educational Measurement: Issues and Practice*, 28(1), 10–24.

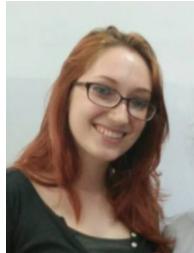
Kiplinger, V. L., Haug, C. A., & Abedi, J. (2000). *Measuring math—not reading—on a math assessment: A language accommodations study of English language learners and other special populations* (Rep.). (ERIC Document Reproduction Service No. ED441813)

Norval, B. (2019a). ELL Corner: Can we change mathematics test items to be more equitable to ELLs? (Part 1 of 3). *Ohio Journal of School Mathematics*, 82, 29–35.

Norval, B. (2019b). ELL Corner: How to modify test items for ELLs: What research says (Part 2 of 3). *Ohio Journal of School Mathematics*, 83, 16–24.

Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C. (2010, June). *Accommodations for English language learner students: The effect of linguistic modification of math test item sets* (Rep. No. NCEE 2009-4079). Retrieved from <https://eric.ed.gov/?id=ED510556> (ERIC Document Reproduction Service)

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, 11(2), 105–126.



Brooke Norval, norvalbe@miamioh.edu, is currently a Graduate Associate in the Department of Teacher Education at Miami University with an undergraduate degree in Mechanical Engineering from The Ohio State University. Brooke's research interests include international perspectives of mathematics education and teaching and learning with English language learners.



Martha Castañeda, castanme@miamioh.edu, is a Professor of Foreign Language Education in the Department of Teacher Education. She has served as the president of the Ohio Foreign Language Association (OFLA). Dr. Castañeda's research interests focus on best practices in language acquisition, instructional design, and cognitive skills as they impact foreign and second language learners.

APPENDIX

Table 1: Linguistic modification recommendations (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2011).

1. Reduce sentence length and complexity (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
2. Change past, conditional, or future tense verbs to present tense when possible (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010)
3. Construct sentences using active voice instead of passive voice (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
4. Avoid irrelevant words or phrases (Sato et al., 2010)
5. Vocabulary should be at or below grade level (Fairbairn & Fox, 2009; Sato et al., 2010)
6. Use high-frequency words (Abedi, 1995; Fairbairn & Fox, 2009; Sato et al., 2010; Shaftel et al., 2006)
7. Avoid ambiguous words, unnecessary words, or words that have multiple meanings (Sato et al., 2010; Shaftel et al., 2006)
8. Avoid proper nouns that are construct-irrelevant (Sato et al., 2010)
9. Avoid words that function as both nouns and verbs (Sato et al., 2010)
10. Avoid hyphenated words and compound words (Sato et al., 2010)
11. Avoid gerunds (Sato et al., 2010)
12. Remove unnecessary introductory phrases (Abedi, 1995; Sato et al., 2010)
13. Break compound sentences into two separate sentences (Abedi, 1995; Sato et al., 2010)
14. Break conditional clauses into separate sentences (Abedi, 1995; Sato et al., 2010)
15. Use bulleted lists (Sato et al., 2010)
16. Avoid references to American culture or holidays (Sato et al., 2010; Shaftel et al., 2006)
17. Change complex question phrases to simple question words (Abedi, 1995)
18. Reduce the number of linguistic elements for lower item difficulty (Shaftel et al., 2006)
19. Avoid relative pronouns that do not have a clear antecedent (Sato et al., 2010)
20. Use visuals that mirror the wording of the text (Fairbairn & Fox, 2009)
21. Avoid colloquialisms or slang (Fairbairn & Fox, 2009; Shaftel et al., 2006)
22. Rephrase negative questions into positive questions (Sato et al., 2010)
23. Shorten unnecessarily long nominals (Abedi, 1995; Sato et al., 2010)
24. Remove relative clauses (Abedi, 1995; Sato et al., 2010)
25. Remove abstract or impersonal presentations (Abedi, 1995)

Table 2: Linguistic recommendation analysis for Test Item 1.

Recommendation Number	Lead Author (Initial)	Todd Edwards	Martha Castañeda	Lead Author (Final)	Total
1					
2	X			X	2
3	X	X	X	X	4
4		X		X	2
5					
6					
7					
8					
9					
10					
11					
12					
13					
14	X		X	X	3
15					
16			X	X	2
17	X	X	X	X	4
18					
19					
20					
21					
22					
23					
24					
25	X		X	X	X

Table 3: *Linguistic recommendation analysis for Test Item 2.*

Recommendation Number	Lead Author (Initial)	Todd Edwards	Martha Castañeda	Lead Author (Final)	Total
1	X	X		X	3
2					
3	X		X	X	3
4	X	X		X	3
5	X			X	2
6	X	X	X	X	4
7					
8					
9					
10					
11					
12					
13	X			X	2
14	X		X	X	3
15					
16					
17					
18					
19					
20		X			1
21					
22					
23					
24					
25			X	X	2

Table 4: Linguistic recommendation analysis for Test Item 3.

Recommendation Number	Lead Author (Initial)	Todd Edwards	Martha Castañeda	Lead Author (Final)	Total
1	X	X		X	3
2					
3					
4	X	X		X	3
5					
6	X		X	X	3
7					
8	X	X	X	X	4
9					
10					
11					
12	X	X		X	3
13					
14	X	X		X	3
15	X				1
16	X		X	X	3
17	X			X	2
18					
19					
20					
21					
22					
23					
24					
25					

Table 5: Total times each recommendation was used sorted by test item.

Recommendation Number	Item 1	Item 2	Item 3	Total
1		3	3	6
2	2			2
3	4	3		7
4	2	3	3	8
5		2		2
6		4	3	7
7				0
8			4	4
9				0
10				0
11				0
12			3	3
13		2		2
14	3	3	3	9
15			1	1
16	2		3	5
17	4		2	6
18				0
19				0
20		1		1
21				0
22				0
23				0
24				0
25	3	2		5

Table 6: Total times each recommendation was used sorted by reviewer.

Recommendation Number	Lead Author (Initial)	Todd Edwards	Martha Castañeda	Lead Author (Final)	Total
1	2	2		2	6
2	1			1	2
3	2	1	2	2	7
4	2	3		3	8
5	1			1	2
6	2	1	2	2	7
7					0
8	1	1	1	1	4
9					0
10					0
11					0
12	1	1		1	3
13	1			1	2
14	3	1	2	3	9
15	1				1
16	1		2	2	5
17	2	1	1	2	6
18					0
19					0
20		1			1
21					0
22					0
23					0
24					0
25	1		2	2	5