ELL Corner: Can we change mathematics test items to be more equitable to ELLs? (Part 1 of 3)

Brooke Norval, *Miami University*

Abstract: *Throughout the next three issues of the* Ohio Journal of School Mathematics*, we will be including one article in each issue about English language learners (ELLs) and how to modify mathematics test items to make them more equitable for ELLs. In this first installment, the author summarizes data indicating that ELL students score significantly lower on National Assessment for Educational Progress (NAEP) mathematics items than non-ELL students. Moreover, the author provides a review of literature supporting language modification of mathematics test items.*

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

1 Introduction

In each of the next three issues of the *Ohio Journal of School Mathematics*, we will include one article about English language learners (ELLs) and how to modify mathematics test items to make their wording more equitable for ELLs. English language learners are a quickly-growing demographic in the United States that should not be overlooked. While the perception may exist that mathematics is a universal language, test scores show that this is not the case. In this article, I summarize ELL performance on mathematics items from the National Assessment for Educational Progress (NAEP). I also review literature that supports changing the language of NAEP test items to reduce bias against ELL students. In future articles, I will discuss specific strategies for modifying math items that preserve construct validity and share example modifications. Throughout the three installments of the ELL Corner, I aim to answer the fundamental research question—*How can mathematics test items be linguistically modified to make them more equitable for ELLs?*

2 Who Are English Language Learners?

ELLs are an important and rapidly-growing population of students in the United States. Formerly called ESL students, or English as a Second Language students, ELLs are students whose native language is not English and are currently learning English. From 1979 to 2003, the general student population increased by 19%. At the same time, the ELL population increased by 124% (Department of Education, 2005b). According to the National Center for Education Statistics, ELLs made up about 9.4% of public school students in the 2014-2015 school year (Department of Education, 2017b). Because ELLs are such a prevalent group, they cannot be ignored or treated as an afterthought. They must be taken into account in our education system as students with unique challenges and considerations. Treating students as a monoculture without accommodating for individual needs impedes efforts of ELLs to thrive academically.

English language learners come from a variety of cultural, ethnic, and linguistic backgrounds. Hispanic students made up 77.8% of ELL students in 2014-2015; Asians, 10.7%; Whites, 5.9%; and Black students, 3.6%, shown in Figure 1.

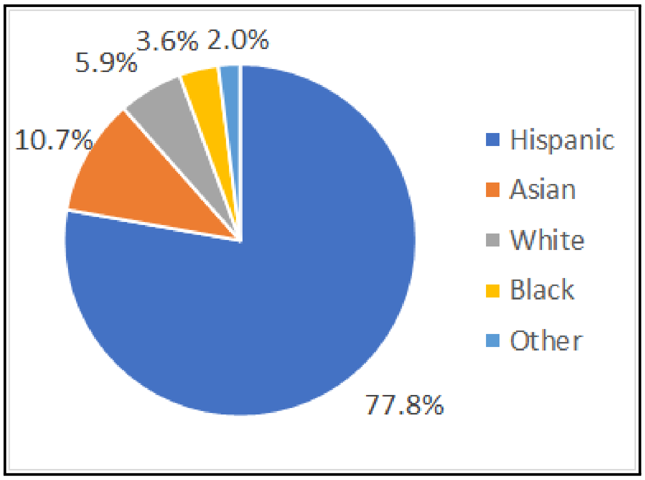

Fig. 1: *Ethnic background of ELL students (Department of Education, 2017b).*

Similarly, most ELLs—about 77.1%—speak Spanish as their home language. Arabic was the next most-spoken home language at 2.3%, followed by Chinese at 2.2% and Vietnamese at 1.8%, shown in Figure 2 (Department of Education, 2017b).

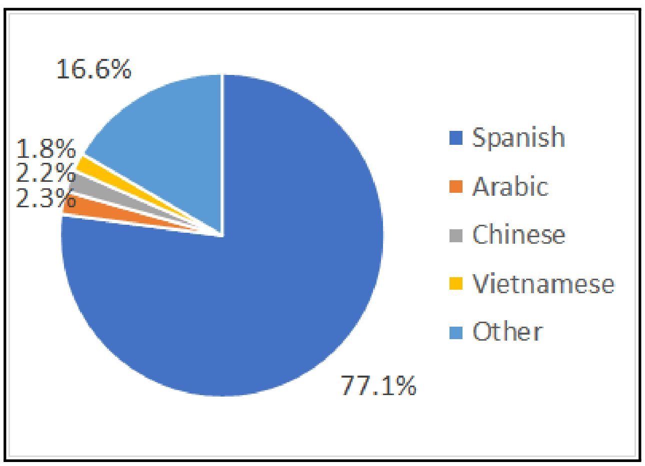

Fig. 2: *Home language of ELL students (Department of Education, 2017b).*

3 The National Assessment for Educational Progress (NAEP)

Although some may view mathematics as a universal language, this is not the case. According to Wolf and Leon (2009), "One notable finding was that the math tests, which are typically assumed to possess lower language demands than science, contained a wide variety of general academic vocabulary. This finding suggests that a math test can also be linguistically demanding" (p. 155). Thus, the perception that students of any native language should be able to perform math problems presented in English equally well is not only incorrect, but damaging. NAEP testing results show that ELLs have significantly lower mathematics test scores than native English speakers.

The National Assessment for Educational Progress (NAEP) is administered to representative samples of students in grades 4, 8, and 12. NAEP tests students in a variety of subjects, including math, reading, science, writing, economics, and geography. In this study, I analyze math results exclusively. NAEP uses the same set of test booklets for all subjects, and the assessment has only minor changes from year to year, so the results may be compared across time (Department of Education, 2017a). Accommodations are provided for English language learners. ELLs are encouraged to be included in the NAEP assessment if the student "(a) participated in the regular state academic assessment in the subject being tested, and (b) if that student can participate in NAEP with the accommodations NAEP allows" (Department of Education, 2017d). Even if students do not meet those two aforementioned requirements, they may still participate in the NAEP assessment if school staff decide that NAEP-provided accommodations will be enough for students.

In the NAEP mathematics assessment, for each of the three grades tested (i.e., 4, 8, and 12), approximately two-thirds of the questions do not permit calculator use. Each question measures one of five content areas: number properties and operations; measurement; geometry; data analysis, statistics, and probability; and algebra. These classifications describe the wide range of mathematical engagement that students experience but do not separate the mathematics problems into discrete categories. Items are also classified into low, moderate, and high complexity. These classifications are intended to describe level of demand (Department of Education, 2017c). For example, low complexity items mainly involve recalling previously learned concepts, specifying exactly what the student should do. Moderate complexity items cause the student to think more flexibly and decide carefully between several possible choices and outcomes. These problems normally have more than one step. Finally, high complexity items are the most demanding items. These items require more abstract reasoning and analysis and demand an answer with more sophisticated thought (Department of Education, 2005a).

3.1 Analysis of NAEP mathematics scores for English language learners

Data for this analysis was obtained from the Main NAEP Data Explorer tool, which allows users to select data based on subject, grade level, year, jurisdiction, and variable (Department of Education, 2018). Tables and reports are generated from this data. Composite mathematics scores for 4th, 8th, and 12th grade students were included for years 2005, 2009, 2013, and 2015. Accommodations were permitted for tests taken in all of these years. The jurisdiction chosen was "National," which includes students in both public and nonpublic schools. The variable chosen for analysis was "Status as English Language Learner, 3 categories" (NAEP ID: ELL3). Students were classified as ELL, non-ELL, and formerly ELL. Results are shown in the tables and three figures below.

Table 1: *Summary of composite NAEP mathematics scores by ELL status, 2005-2015.*

Year	4th grade scores			8th grade scores			12th grade scores		
	ELL	Non-ELL	Formerly ELL	ELL	Non-ELL	Formerly ELL	ELL	Non-ELL	Formerly ELL
2005	216	240	240	244	281	276	120	151	147
2009	218	242	241	243	286	270	117	155	143
2013	219	244	246	246	287	272	109	155	141
2015	218	243	246	246	285	277	115	153	140

As shown in Table 1 and Figures 2, 3 and 4, ELLs achieve significantly lower scores on NAEP mathematics assessments than non-ELLs or former ELLs. On average for the years shown, ELLs' 4th grade math scores are around 90% of those of both non-ELLs and former ELLs. For 8th grade, ELLs' scores are 86% of those of non-ELLs and 89% of those of former ELLs. For 12th grade, ELLs' average scores are even lower at 75% of those of non-ELLs and 81% of those of former ELLs. Clearly, there appears to be a significant achievement gap in the mathematics content area for English language learners.

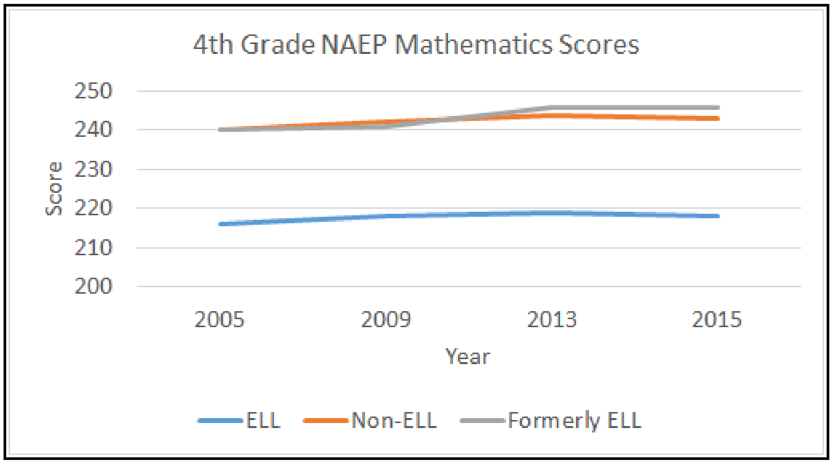

Fig. 3: *Composite 4th grade NAEP mathematics scores by ELL status, 2005-2015.*

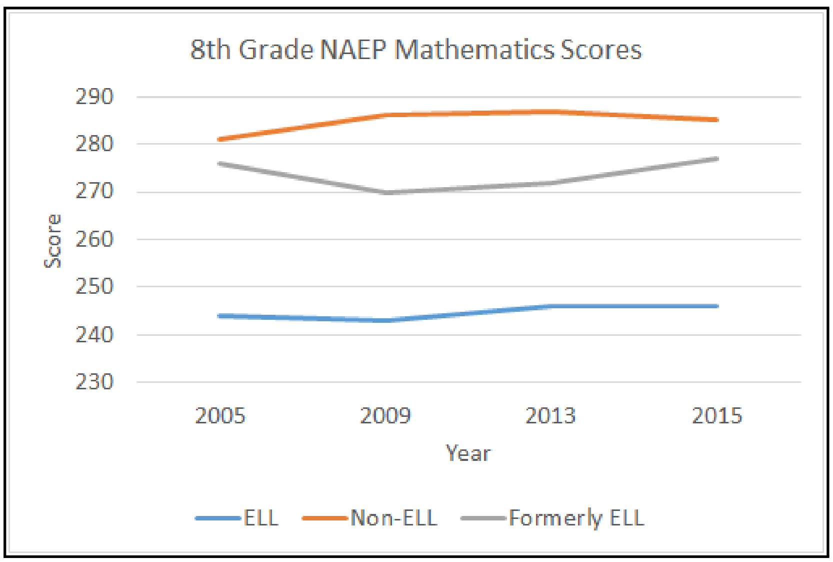

Fig. 4: *Composite 8th grade NAEP mathematics scores by ELL status, 2005-2015.*

What are the implications for the achievement gap in these test scores on the lives of English language learners? A poor performance on the mathematics NAEP assessment may predict a poor performance on the mathematics portions of high-stakes tests, such as the American College Test (ACT) or the SAT. These tests are taken in large numbers by those students who wish to apply to colleges and universities. Universities use these test scores to help determine admissions decisions. Students with high ACT and SAT scores are more competitive candidates to get into their school of choice—or any school at all. Furthermore, ACT and SAT scores are also often used by colleges to determine admissions into honors programs, eligibility for scholarships, and so on. Students who score poorly on these high-stakes exams are at a disadvantage when it comes to college admissions and eligibility for scholarships.

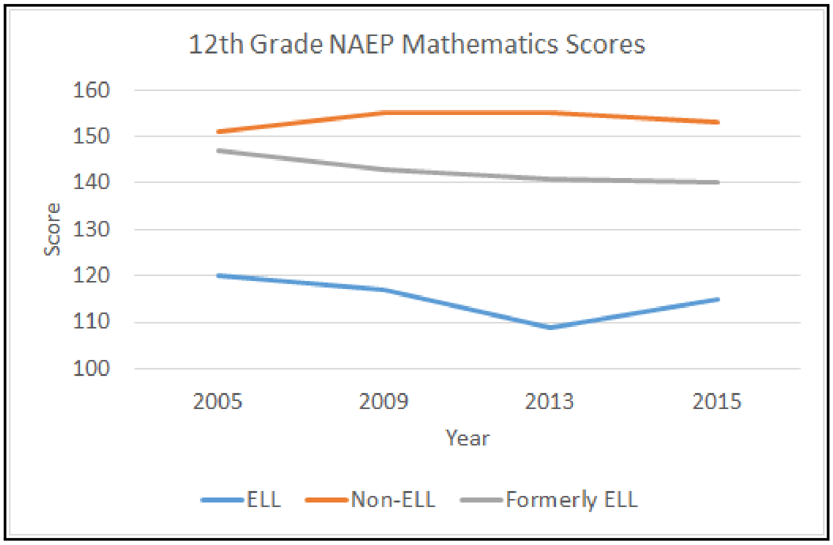

Fig. 5: *Composite 12th grade NAEP mathematics scores by ELL status, 2005-2015.*

4 Linguistic Modification of Mathematics Test Items

It is critical to ensure that tests measure the construct being learned. In this case, mathematics test items must measure students' mathematical abilities, not construct-irrelevant abilities such as language skills, in order to maintain construct validity. If mathematics test items measure language skills instead of math skills, then the test does not demonstrate construct validity. Mahoney (2008) expands on this, stating, "Of particular concern is that test scores will not accurately reflect knowledge of a specific subject area if the student is not yet proficient in the language of the achievement test. For tests in English, a student's English proficiency may introduce construct-irrelevant variance to the test score as error, making the test score and interpretations of test scores less reliable and less valid" (p. 17). Sato et al. (2010) builds upon this idea, stating, "To assess English language learner students' knowledge of academic content, it is critical to determine whether their academic performance reflects their understanding of the targeted content or their lack of English language proficiency" (p. 80). Therefore, when creating test items, language differences among students must be taken into account to ensure that tests are only measuring the construct being assessed and not additional irrelevant constructs.

Studies show that language has an impact on ELL students' performance on mathematics test questions. For example, Abedi, Hofstetter, & Lord (2004) found that "as the language demands of individual test items decrease, the performance gap between English learners and English-proficient students decreases" (p. 6). It follows that performance gaps in math scores between native English speakers and ELLs aren't simply a result of ELLs being less intelligent or less proficient in math than native English speakers; instead, language demands result in increased difficulty and understanding for ELLs. The authors also found that "the performance gap virtually disappeared in math computation, where test item language demands were minimal" (Abedi, Hofstetter, & Lord, 2004, p. 6). When mathematics questions solely involve computation with little to no language, ELLs appear just as competent as native English speakers in mathematics performance. ELLs are often provided with accommodations during tests in order to make the tests more equitable. According to Butler & Stevens (1997), accommodations are intended to provide support for "students for a given testing event either through modification of the test itself or through modification of the testing procedure to help students access the content in English and better demonstrate what they know" (p. 5). Accommodations are meant to reduce construct-irrelevant variance in student testing scores and maintain a test's construct validity. Accommodations for ELLs may include dictionaries, extended time, translations, and linguistically modified test items. Linguistically modified test items, sometimes referred to as linguistically simplified test items, use plain grammar, shorter sentence structure, more common vocabulary words, and so on in order to improve comprehension for ELLs. According to Sato et al. (2010), "Linguistic modification of test items is an approach for addressing the particular access needs of English language learner students so that test performance is attributable less to English language proficiency and more to knowledge and skills related to the tested content" (p. 80). Using linguistic modification as a test accommodation has been widely researched. Some studies show that linguistically modified test questions aid ELLs, helping to close the test scores gap.

For example, Abedi, Lord, Hofstetter, & Baker (2000) analyzed tests with four different accommodations and compared them to tests with no accommodations. The four accommodations were modified English, extra time only, glossary only, and extra time with a glossary. The only accommodation that led to increased ELL test scores but not to increased native English speaker test scores was the modified English accommodation. If an accommodation helps both ELLs and native English speakers, then it simply serves to make the test easier for everyone, and does not remove bias against ELLs (Abedi, Lord, Hofstetter, & Baker, 2000).

In addition, Sato et al. (2010) conducted a study in which 7th and 8th grade students were randomly assigned either a regular math test or a math test with linguistically modified items. They found that "As a whole, the items in the linguistically modified set more reliably measured math understanding of students" (p. 53) in the ELL subgroup than did the items in the original item set. They also found that linguistically modified problems improved the reliability of that item set due to the fact that some items were more closely tied to math understanding for ELLs, but not for native English speakers. In addition, the average difference in math scores for the original vs. linguistically modified item sets was greatest for ELLs and lowest for native English speakers. This suggests that linguistic modification encourages language understanding. Results suggest that linguistically modifying the math problems did not alter the math constructs being assessed; that is, the mathematical aspects of the problem and the math skills used to solve the problems were not changed, only the linguistic aspects of the problem were (Sato et al., 2010). Altogether, these findings make linguistic modification a highly effective, valid, and useful test accommodation.

5 Conclusion

Throughout this paper, I have shown how linguistic modifications are effective in reducing testing bias against ELLs. However, several questions still remain. How, exactly, do other researchers suggest mathematics test items should be modified to become more equitable to ELLs? What does a modified mathematics test item look like? How would researchers of different backgrounds modify the same test items, and how would their modifications differ? These questions will be answered in the next two issues of the *Ohio Journal of School Mathematics*.

References

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment Accommodations for English Language Learners: Implications for Policy-Based Empirical Research. *Review of Educational Research, 74*(**1**), 1–28. doi:10.3102/00346543074001001

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(**3**), 16–26.

Butler, F. A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (CSE Tech. Rep. No. 448). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Department of Education, National Center for Education Statistics. (2018). *Main NAEP Data Explorer*. Retrieved from https://www.nationsreportcard.gov/ndecore/landing

Department of Education, National Center for Education Statistics. (2005a, September 14). 2005 Mathematical Complexity. Retrieved April 05, 2018, from https://nces.ed.gov/nationsreportcard/mathematics/complexity2005.aspx

Department of Education, National Center for Education Statistics. (2005b, June). *The Condition of Education 2005* (Rep. No. NCES 2005-094). Retrieved April 05, 2018, from https://nces.ed.gov/pubs2005/2005094.pdf

Department of Education, National Center for Education Statistics. (2017a, May 10). *NAEP Overview*. Retrieved April 05, 2018, from https://nces.ed.gov/nationsreportcard/about/

Department of Education, National Center for Education Statistics. (2017b). *The Condition of Education 2017* (Rep. No. NCES 2017-144). Retrieved April 05, 2018, from https://nces.ed.gov/pubs2017/2017144.pdf

Department of Education, National Center for Education Statistics. (2017c, November 16). *What Does the NAEP Mathematics Assessment Measure?* Retrieved April 05, 2018, from https://nces.ed.gov/nationsreportcard/mathematics/whatmeasure.aspx

Department of Education, National Center for Education Statistics. (2017d, November 22). *Frequently Asked Questions*. Retrieved April 05, 2018, from https://nces.ed.gov/nationsreportcard/faq.aspx

Mahoney, K. (2008). Linguistic Influences on Differential Item Functioning For Second Language Learners on the National Assessment of Educational Progress. *International Journal of Testing, 8*(**1**), 14–33. doi:10.1080/15305050701808615

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)

Wolf, M. K., & Leon, S. (2009). An Investigation of the Language Demands in Content Assessments for English Language Learners. *Educational Assessment, 14*(**3–4**), 139–159. doi:10.1080/10627190903425883 Language Learners. Educational Assessment, 14(3-4), 139-159. doi:10.1080/10627190903425883

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.