

June 2023

## Effectiveness of English Learners Computer-Based Testing Accommodations: A Meta-Analysis

Tuba Gezer Dr.  
*The Ohio State University, gezer.3@osu.edu*

Claudia Flowers  
*University of North Carolina at Charlotte, claudiaflowers@uncc.edu*

Richard Lambert  
*University of North Carolina at Charlotte, rglander@uncc.edu*



Part of the Bilingual, Multilingual, and Multicultural Education Commons, Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, Elementary Education Commons, and the Language and Literacy Education Commons

Find similar works at: <https://stars.library.ucf.edu/jele>

University of Central Florida Libraries <http://library.ucf.edu>

This Article is brought to you for free and open access by STARS. It has been accepted for inclusion in Journal of English Learner Education by an authorized editor of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

### Recommended Citation

Gezer, T., Flowers, C., & Lambert, R. (2023). Effectiveness of English Learners Computer-Based Testing Accommodations: A Meta-Analysis. *Journal of English Learner Education*. (15)1.

Retrieved from <https://stars.library.ucf.edu/vol15/iss1/2>

### Introduction

In the USA, the federal government reauthorized the Elementary and Secondary Education Act (ESEA) in 2001, which increased state accountability for students with disabilities and students who are English learners (ELs). The Every Student Succeeds Act (ESSA) required states to create goals and interim measures for English language proficiency (United States Department of Education, 2015). With approximately 10% of all school children being ELs (McFarland et al., 2018) and with the increased number of ELs participating in state testing systems, there has been an increased focus on fairness in testing EL students.

The Standards for Educational and Psychological Testing emphasizes fairness as a fundamental validity issue that should be addressed in all phases of the testing process (AERA, APA, & NCME, 2014). Designing fair tests requires developers to reduce barriers for all examinees during test development, administration, scoring, interpretation, and uses (Thurlow et al., 2009). An essential concept in fairness is accessibility. Accessibility focuses on eliminating barriers and providing equal access for all examinees, allowing score interpretations to have comparable meaning for individuals or groups in the intended population of test-takers (Stone & Cook, 2018). Methods for increasing accessibility and fairness in testing include the careful design of assessments and the use of accommodations (Thurlow et al., 2009). In this study, we are using accessibility and accommodation as defined in the Standards for Educational and Psychological Testing. Accessibility is the notion of providing unobstructed opportunities to all students to demonstrate their ability on the measured construct (AERA, APA, & NCME, 2014, p. 49), and accommodations refer to changes in the test format, test administration, or response procedure while maintaining the original construct (AERA, APA, & NCME, 2014, p.58).

Computer-based testing (CBT) has made it easier and cheaper to accommodate specific test takers' needs for presenting test items during an assessment by reducing the number of paper-based test forms required to meet specific students' needs. In the U.S. Department of Education's major initiatives (e.g., Race to the Top Assessment Program), the development of CBT was encouraged because of the many positive merits, including built-in accommodations (Thurlow, Lazarus, Albus, & Hodgson, 2010). Unlike paper-based tests, CBT allows individual customization of tests. Given the ability to customize the presentation and response modes in the CBT environment, test developers can build accommodations into the test design to eliminate construct irrelevant variance.

While EL students are increasing in the US educational system, ELs' academic performance continues to lag behind their English-speaking peers. Based on 2017 data from the National Assessment of Educational Progress

(NAEP), approximately half of all states reported a decrease in the percentage of grade 4 proficiency rates in mathematics, and similar results are noted in reading proficiency rates (NAEP, 2017). The academic under-performing of EL students is due, in part, to the challenges for ELs in communicating in the English language, especially on core subject tasks (Abedi & Levine, 2013). Test accommodations are one support that has been shown to be effective at narrowing the achievement gap. Research suggests that ELs who received testing accommodations outperform ELs who do not receive accommodation or do not receive appropriate accommodations (Abedi, 2009; Kopriva et al., 2007).

Research has examined EL test accommodations' effectiveness and validity (Abedi, 2009; Abedi et al., 2020; Albus et al., 2005; Johnson & Monroe, 2004; Kopriva et al., 2007). Several meta-analyses summarized the fairness and effectiveness of EL test accommodations (Li & Suen, 2012a; 2012b; Pennock-Roman & Rivera, 2011; Rios et al., 2020) and EL test accommodations in large-scale testing settings (Kieffer et al., 2009). However, little is known about computer-based testing (CBT) accommodations for ELs, even though CBT is the predominant mode of delivering large-scale summative tests in the U.S. educational system. Therefore, the purpose of this meta-analysis is to aggregate data from CBT accommodation studies to evaluate the validity and effectiveness of accommodations for ELs.

### **Test Accommodations**

EL test accommodations are designed to reduce construct irrelevance variance while increasing content accessibility for ELs. Common test accommodations on paper-based tests are extended time, dictionaries (e.g., glossary, English dictionary, bilingual dictionary), and linguistically simplified test items (Rivera, 2003; Shafer Willner et al., 2008).

Dictionary accommodations provide the definition of the words, but it does not include content-related terms. Dictionary accommodations on CBT provide a simple definition of a word when the cursor is on a word. Dictionary accommodation can provide visual support or the translation of the word.

Linguistic modification accommodation provides simplifying the language complexity of the test and linguistic modifications are used interchangeably with simplified English version in this study. Translation accommodation is administering the test in students' home language, and it is offered when students are proficient in their home language. The Spanish version of the tests is the most common one.

Abedi (2006a) examined the difficulties of measuring EL students' content knowledge due to the linguistic complexity in many academic content area tests. He found a statistically significant difference between EL and non-EL student groups' measurement error resulting in disadvantages for EL students. Although a large amount of research highlights the performance gap between EL and non-EL

students (Abedi, 2006a; Miley & Farmer, 2017; Polat et al., 2016; Solano-Flores & Trumbull, 2003; Wolf et al., 2008), the source of this achievement gap could be language proficiency, not the content knowledge (Abedi, 2006b). An examination of standardized test scores demonstrated that ELs and native speakers' achievement gap was higher in reading and writing in high language demand items than in lower language demands in mathematics tests (Abedi, Leon, & Mirocha, 2003). Since most content assessments are not designed to measure student linguistic abilities, the language demand on content assessment is a threat to construct validity. Therefore, providing effective and valid test accommodations is critical to reveal EL students' actual performance on content assessments. Because CBT large-scale testing often carries high stakes that impact students, identifying successful practices that support EL students' linguistic and academic needs is necessary to demonstrate what they know and can do.

#### **Validity of EL Test Accommodations**

There have been multiple studies examining the validity of EL test accommodations. In an experimental study that compared non-EL students' performance with accommodations and non-EL students' performance without accommodations (Abedi, 2009; Abedi et al., 2020), the math performance of non-EL students without accommodations was not statistically significantly different from those students receiving EL test accommodations (i.e., linguistic modification, English read aloud, and English glossary) (Abedi et al., 2020). However, a bilingual glossary improved non-EL accommodated students' math performance compared to non-EL without accommodations. Li and Suen (2012b) conducted a meta-analysis with 21 studies. They found that the accommodations do not influence non-EL students' performance. Although there are mixed results regarding the validity of EL test accommodations, overall the evidence suggests that EL test accommodations do not cause unfair advantages to EL students.

#### **Effectiveness of EL Test Accommodations**

Despite the theoretical support of the validity of test accommodations for EL students in terms of maintain focal construct, the research presents mixed results regarding test accommodations' effectiveness. EL test accommodations could be divided into dictionaries and linguistic modification (e.g., translation, linguistic simplification). Dictionaries may help reduce the linguistic complexity so EL students can understand unknown words. However, defining the words may create a threat to the assessment's validity due to giving unfair advantages to EL students (Abedi, Courtney & Leon, 2003; Acosta et al., 2008). On the other hand, Albus et al. (2001) concluded that there was no statistically significant difference between EL students who used a dictionary and those who did not use a dictionary in reading tests. Alternatively, testing the effectiveness of pop-up glossaries with 4<sup>th</sup> and 8<sup>th</sup>-grade students demonstrated that pop-up glossary test accommodation could increase ELs' test scores by a .50 standard deviation unit

(Abedi, 2009). In addition, there are bilingual dictionaries and picture dictionaries that provide visual support for English words. The previous meta-analysis studies combined customized dictionaries, glossaries, bilingual dictionaries, picture dictionaries, and pop-up glossaries, and they did not find a statistically significant effect of dictionary and glossary accommodations (Kieffer et al., 2012; Liu & Suen, 2012a; Rios et al., 2020).

Translation and linguistically modified tests are also EL test accommodations. Translation accommodation can take different forms, such as taking the test in the home language or taking only the test instructions in the home language. Turkan and Oliveri (2014) articulated that 12 out of 50 states offer translation accommodations to EL students, and they highlight that the effectiveness of translation accommodations depends on the quality of the test translation. There is limited research on home-language accommodations for EL (Kieffer et al., 2012), but it is important to match the language of assessment with the language of instruction (Abedi et al., 2004; Kieffer et al., 2012). A combination of 12 studies' effect sizes demonstrates that linguistic simplification accommodation can improve ELs performance and decrease the achievement gap between ELs and non-ELs from 9% to 19% (Kieffer et al., 2012).

### **Computer-Based Testing for EL Students**

In 2015, most K-8 students took a high-stakes state summative assessment on the computer; in fact, only 15% of the assessments were offered only using a paper-based format (Strategies, 2015). While many assessments were available in hybrid (both computer administration and paper-and-pencil format), the most test administrations used computer administration. CBT can simultaneously allow multiple accommodations, combining pop-up glossary and read-aloud accommodations (Russell et al., 2009). According to Abedi (2014), CBT allows effective test accommodations, which are not readily available for paper-based testing, to provide an efficient assessment mode for EL students. Considering the importance of providing test accommodations and CBT's benefits for creating accessible and fair assessments, examining the validity and effectiveness of test accommodations on CBT becomes critical for an appropriate assessment process.

Comparing test accommodations, pop-up glossary on a computer, customized dictionary, extra time, small-group testing reveals that computer testing and extra time effectively improve assessment accessibility to EL students without raising any validity concern (Abedi, 2009). Abedi et al. (2020) examined the effectiveness of EL test accommodations on CBT using an experimental design and surprisingly found that there were no significant gains for participants who used CBT accommodations, and in some cases, there was a negative impact. The study found statistically significantly lower scores for EL participants who used Spanish math tests and bilingual glossary accommodations than EL students' who did not use any accommodations. Furthermore, there were no differences

between the experimental and control groups for linguistic modification, English read-aloud, and English glossary.

### **Previous Meta-analyses on EL Test Accommodations**

As EL test accommodation literature grew, there have been multiple meta-analyses to summarize EL test accommodations literature. This section will summarize the main findings of previous meta-analysis on EL test accommodations. Test accommodations can increase students with disabilities' academic testing scores up to .16 standard deviation units (Chiu & Pearson, 1999). While this study was first meta-analysis that include EL test accommodations, out of 30 studies, only seven studies included EL students. Thus, the study provided limited evidence for EL test accommodations. Kieffer et al. (2009) conducted a meta-analysis examining the effectiveness of test accommodations for EL students on large-scale assessments using 11 studies and found that English dictionaries and glossaries have a small but statistically significant effect on ELs performance. In a meta-analysis by Pennock-Roman and Rivera (2011), which included 14 studies, simplified English test accommodation was more beneficial if EL students' language proficiencies were intermediate. This meta-analysis also revealed that computer-administered glossaries are effective regardless of time restrictions.

On the other hand, Li and Suen's (2012a) meta-analysis, including 19 studies, suggested English proficiency level and time restriction influence the effectiveness of the EL test accommodations. Rios et al. (2020) conducted a meta-analysis, which included 26 studies and 95 effect sizes, and concluded that accommodations could improve ELs test performance by .16 standard deviations. This meta-analysis also demonstrated that EL test accommodations are less effective for math/science content than non-math/science test content.

### ***The Rationale for the Current Study***

Rios et al. (2020) highlighted the disparity between EL test accommodations research and practice. Similarly, previous meta-analyses (Kieffer et al., 2009; Li & Suen, 2012; Pennock-Roman & Rivera, 2011; Rios et al., 2020) combined paper-based test accommodations and CBT accommodations even though the majority of state accountability assessment is delivered on CBT. Thus, the validity and effectiveness of EL test accommodations for only CBT could not be clearly disentangled in the previous meta-analyses. Considering the expansion of CBT and the disparity between EL test accommodation research and practices, it is necessary to summarize the research on EL test accommodation on CBT using a meta-analysis.

The purpose of this study is to investigate the validity and effectiveness of EL accommodations on CBT. This meta-analysis uses the random-effect model to quantify the average effects of EL test accommodations on CBT. The following research questions have guided this study:

1. Do test accommodations on CBT provide an unfair advantage to ELs? In other words, do EL test accommodations on CBT influence non-EL students' academic performance?
2. To what extent are CBT test accommodations effective in improving EL students' academic performance?
3. What factors (i.e., grade, accommodation type) influence the effectiveness of EL test accommodations?

### **Method**

A meta-analysis approach, which aggregates quantitative research findings to uncover the patterns of the literature and build new theories, was introduced by Glass (1976; 1977). The general structure of conducting a meta-analysis includes a statement of the problem, literature search for relevant studies, quality evaluation, analyzing the outcomes, interpreting the evidence, and displaying the results (Cooper, 2017, p.25).

### **Literature Search**

Electronic and manual literature searches were used to capture all the relevant research regarding EL test accommodations in CBT. An electronic literature search was conducted targeting the major databases, Education Resources Information Center (ERIC), Educational Administration Abstracts, Journal Storage (JSTOR), ProQuest, and PsycINFO. The electronic literature search was conducted in May 2020, and different combinations of the following keywords were used to find related studies: *accommodations, test accommodations, English language learners, computer, computer-based assessments*. Search results included peer-reviewed articles, technical reports, theses/dissertations, and conference proposals. Peer-reviewed journal articles and technical reports were included because of the quality of the research. Theses/dissertations and conference proposals were included to eliminate the publication bias because studies with statistically significant effects are more likely to be published (Glass, 1977). This electronic literature search for research was published between 1997 to 2020, and the location was limited to the United States.

### **Inclusion Criteria**

The following inclusion criteria used to select studies included (a) empirical quantitative studies that analyze EL test accommodation on CBT in K-12 settings, (b) studies should report effect sizes or enough data to compute effect sizes, and (c) studies with at least one EL test accommodation on CBT. EL accommodation studies using paper-based format were excluded (e.g., Abedi et al., 2001; Deysson, 2013; Fairbairn, 2006). Studies were excluded if they did not include test performance (i.e., Roohr & Sireci, 2017) or report enough information to compute effect sizes (i.e., Cohen et al., 2017).

In addition to published articles and technical reports, conference presentations, dissertations, and theses were included to eliminate the probability of publication bias. Borenstein et al. (2009) stated that comprehensive research is ideal for dealing with publication bias so that multiple databases (ERIC, JSTOR, ProQuest, PsycINFO) were included in the literature search, and the time frame was from 1997 to 2020.

### Data Coding

The following variables were included in data analysis based on the literature: grade, content, accommodations (Dictionary/Glossary, Linguistic Modification, Translation, Read-aloud), and the use of multiple accommodations. Like Li and Suen (2012a), the grade was coded as 0 for K-6 and 1 for 7-12. It was hypothesized that test accommodations would be statistically more effective for K-6 EL students (Rios et al., 2020). Test content was coded Reading, and Math/Science as Li and Suen (2012) suggested. Because all eligible studies were in Math/Science content, this variable was not included in the analysis (see Table 1). Accommodations were dictionaries, including a bilingual glossary, pop-up glossary and picture dictionary, translation, linguistic simplification, and English read-aloud. We grouped the accommodations into three groups based on previous studies (Li & Suen, 2012; Rios et al., 2020). Two dummy variables were coded for dictionary and translation accommodations, and others (i.e., linguistic simplification and read-aloud), with the others accommodation serving as the reference/comparison group of accommodations.

To examine the quality of the coding, a second reader independently coded over half of the studies. The average agreement between the two raters was 85%. The coders met to examine the disagreements and resolved all disagreements.

### Data Analytical Procedure

Cohen's  $d$  effect size was used to compute the standardized mean difference between the treatment and control groups according to the following formula:

$$d = \frac{X_1 - X_2}{SD_{pooled}} \quad SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}} \quad (1)$$

$$V_d = \frac{n_1 + n_2}{n_1 n_2} S_{pooled}^2$$

$X_1$  is the mean of the treatment group (receiving accommodations),  $X_2$  is the control group (no accommodations) mean, and  $SD_{pooled}$  is the pooled standard deviation where  $n_1$  is the sample size of the treatment group, and  $n_2$  is the control group's sample size and  $SD_1$  and  $SD_2$  are the standard deviations of treatment and control group, respectively.  $V_d$  is variance. Since Cohen's  $d$  tend to overestimate effect sizes from small samples, Cohen's  $d$  was converted to Hedges'  $g$  by using the following formulas:



$$g = j * d \quad \text{and} \quad j = 1 - \frac{3}{4(n_1 + n_2 - 2) - 1} \quad (2)$$

In this formula,  $g$  is Hedges'  $g$ ,  $d$  is Cohen's  $d$ ,  $n_1$  is the sample size of the treatment group, and  $n_2$  is the control group's sample size.

All the analyses were conducted in R software with Meta and Metafor packages. Outliers were examined before computing the average effect size using Mahalanobis distance. Two studies (three effect sizes) were detected as outliers. The analyses were conducted with outliers and without outliers; we did not remove the outliers because of minor differences between the results. The random-effect model was used to compute the average effect size and effect size heterogeneity for EL test accommodations' validity and the effectiveness of EL test accommodations on CBT. Heterogeneity of effect size and the average effect size estimates were computed with a random-effects (intercept-only) model according to restricted maximum likelihood estimation. Inverse variance weight was applied to the effect sizes.  $I^2$  statistics were used to measure heterogeneity where  $I^2 < 50\%$  indicates small heterogeneity,  $50\% < I^2 < 75\%$  indicates medium heterogeneity, and  $I^2 > 75\%$  indicates large heterogeneity (Higgins, & Thompson, 2002).

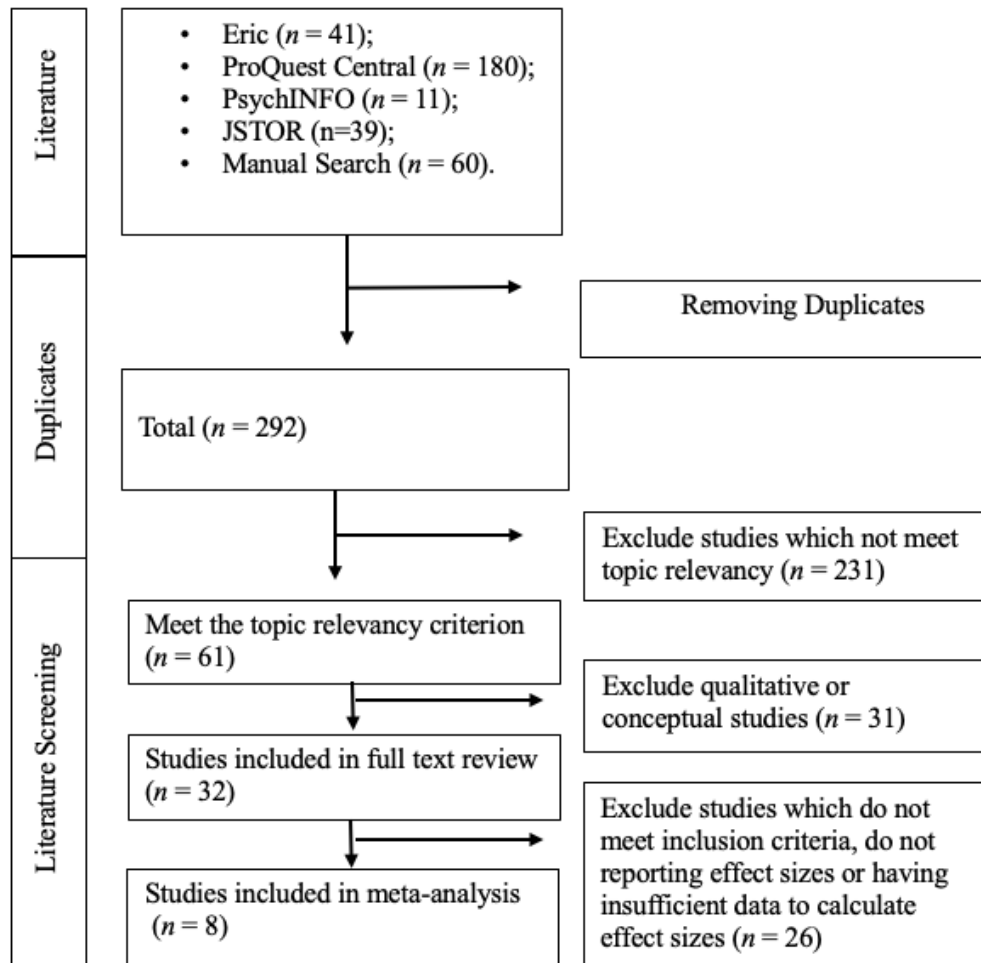
The moderator model for the effectiveness of EL test accommodations on CBT was conducted via restricted maximum likelihood estimation:

$$\begin{aligned} \text{Effectiveness}_{\text{Hedges' } g} &= \beta_0 + \beta_1 * (\text{Grade}) + \beta_2 * (\text{Dictionary}) + \beta_3 \\ &* (\text{Translation}) + e \end{aligned} \quad (3)$$

In this formula,  $\beta_0$  represents the average effect size after controlling for the included variables, and  $e$  was the residual term.

### Results

The PRISMA Flow Diagram for the literature search is shown in Figure 1. Each study's abstract was examined to understand if the study was empirical research about CBT EL test accommodations. In addition, previous EL test accommodation meta-analyses (Kieffer et al., 2009; Li & Suen 2012a, 2012b; Pennock-Roman & Rivera, 2011; Rios et al., 2020) references were reviewed for potential research that met the inclusion criteria. After literature screening, out of 292 studies, 61 studies met the topic relevancy criterion. Out of these, 32 studies were eligible for full-text screening, and finally, eight studies were included in the meta-analysis. One of the eight eligible studies (12.5%) came from unpublished research.

**Figure 1***PRISMA Flow Diagram of Literature Search and Screening*

The eight studies were conducted between 1999 and 2020. Although the sample size is small in this study, 25% of the studies (two studies: Abedi et al., 2020; Ardasheva et al., 2018) were not included in the previous meta-analyses. Also, it is important to note that researchers used paper-based terminology regarding dictionary accommodations for ELs. We presented the description of each accommodation in Table 1 based on the researchers' choice. However, all dictionary accommodations were grouped as one category in the data analysis process.

**Table 1***The Summary of Eligible Studies*

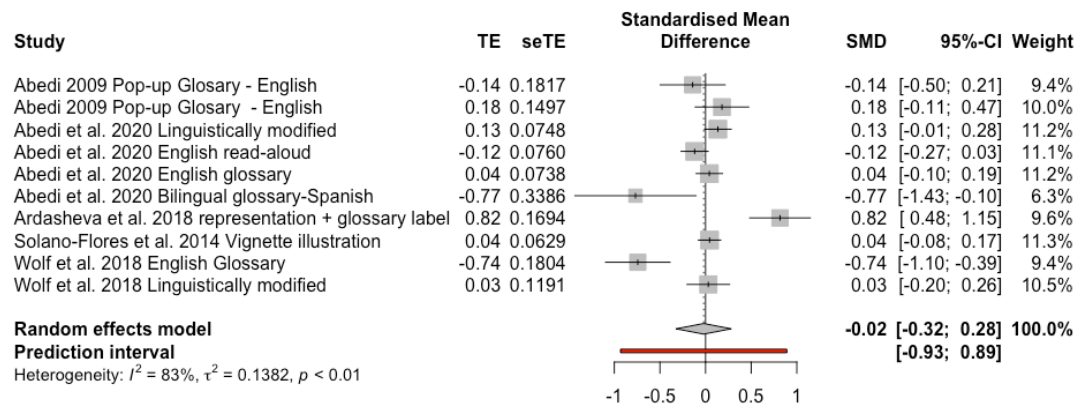
Study	Sample Size	Grade	Content	Accommodations	The implementation
Abedi (2009)	1149	4 & 8	Math	Pop-up Glossary	This study uses PPT and CBT. Pop-up dictionary demonstrates a simple explanation of a word with the touch of the mouse.
Abedi et al. (2020)	1530	9	Math	Linguistic simplification, English read-aloud, English glossary, Translation, Bilingual glossary	Linguistic simplification reduces language demand without threatening the validity of the assessment. Read-aloud have audio files for questions so students can listen to them. English glossaries include definitions of selected words. Students take Spanish Glossary or Spanish Math (translation) test.
Kopriva et al. (2007)	272	3 & 4	Math	Picture dictionary, bilingual dictionary, and English read-aloud, and combination of these	Picture dictionaries present a picture of selected words with a mouse click. Bilingual dictionaries provide Spanish translations of the selected words. Students' Spanish proficiency was determined by their teachers. Read-aloud would read the items to students.
Ardasheva et al. (2018)	174	7	Science	Visual support + glossary	This accommodation provides visual representations of the words through Google images and short definitions of the words.
Robinson (2010)	3273	K & 1	Math & literacy	Translation	Spanish version of the Math and Literacy tests.
Solano-Flores et al. (2014)	728	8	Science	Illustration (Visual support)	This accommodation adds an illustration to the items that included only text.
Alt et al. (2013)	21	2	Math	Translation	Spanish version of the test
Wolf et al. (2018)	513	8 & 9	Math	English Glossary, Linguistic simplification	English glossaries provide a short definition of the words when students click the word. Linguistic simplification provides lexical and syntactic support.

### The Validity of EL Test Accommodations on CBT

Five studies examined the validity of EL test accommodations on CBT by comparing the effects of the accommodations on EL and non-EL students, and these studies included 10 effect sizes and 2779 non-EL students. In Figure 2, the random-effect model results for the validity of EL test accommodations were presented. Effect sizes were treated as independent because effect size were derived from a unique sample. The total effect size for the validity of EL test accommodations on CBT was  $-0.02$   $SD$  ( $SE=0.13$ ; 95% CI:  $-0.3246$ ,  $0.2825$ ;  $p=.88$ ). Although there is a large heterogeneity ( $I^2=83%$ ;  $Q=53.23$   $p < .001$ ), a moderator analysis was not conducted due to the small sample size.

#### Figure 2

*The Forest Plot of the Effect Sizes for the Validity of EL Test Accommodations on CBT*



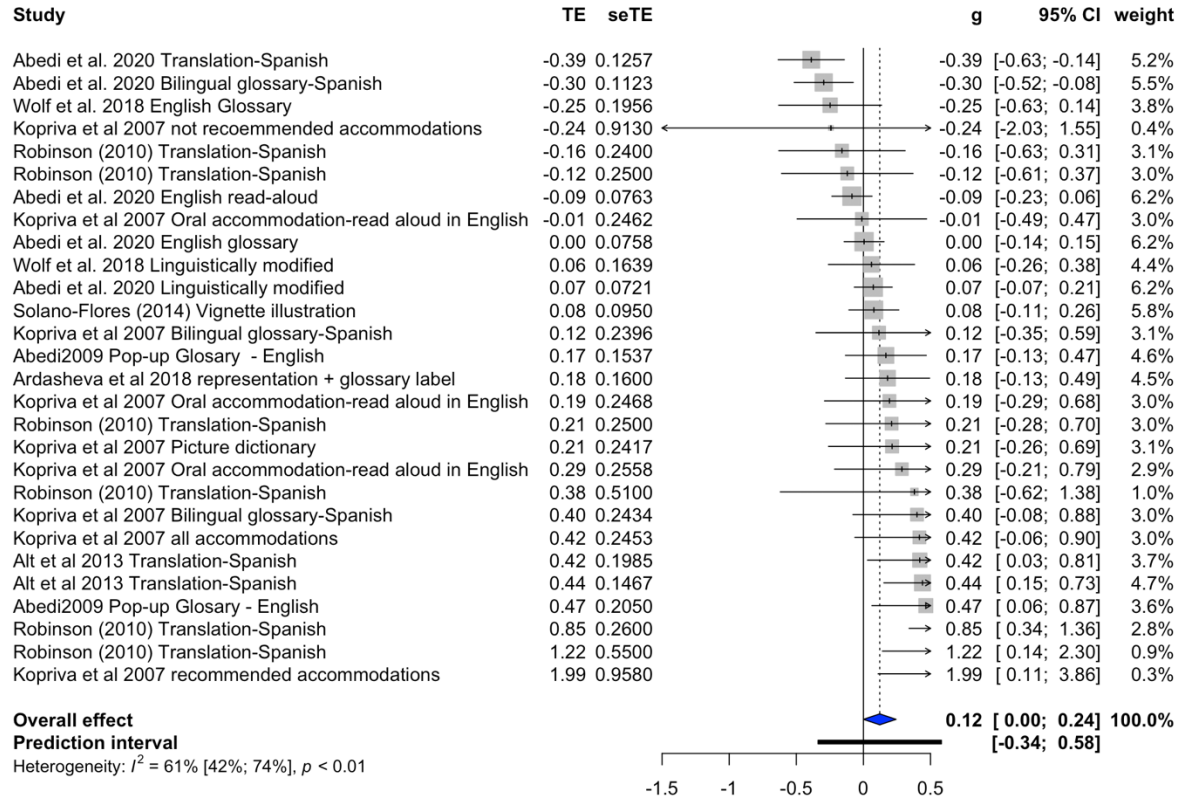
*Note.* In the forest plot, the center of the box represents the size of the treatment effect, and the black line shows the confidence interval. While diamond shows the summary of the treatment effect, the confidence interval is represented with right and left extremes.

### Effectiveness of EL Test Accommodations on CBT

In total, eight studies examined the effectiveness of EL test accommodations on CBT, and these studies included 28 effect sizes and 5987 EL students. In Figure 3, the random-effect model results for the effectiveness of EL test accommodations on CBT were presented. The total effect size of EL test accommodations on CBT was  $.12$   $SD$  ( $SE=0.06$ ; 95% CI:  $0.0002$ ,  $0.2433$ ;  $p < .05$ ). A moderator analysis was conducted because of the heterogeneity ( $I^2=61%$ ;  $Q=69.86$   $p < .01$ ).

#### Figure 3

*The Forest Plot of the Effect Sizes for the Effectiveness of EL Test Accommodations on CBT*



*Note.* In the forest plot, the center of the box represents the size of the treatment effect, and the black line shows the confidence interval. While diamond shows the summary of the treatment effect, the confidence interval is represented with right and left extremes.

According to 28 effect sizes, grade, dictionary, and translation variables were included in the model. Table 2 presents the model results. Four out of 8 studies included at least one K-6 sample, which was about 65% of the effect sizes. The results demonstrated a statistically significant difference in accommodation effectiveness when comparing samples in K-6 and Grade 7-12 ( $\beta = -0.43$ ;  $p < .001$ ), which indicated that accommodations' effectiveness was .43 SD lower in grade 7-12 compared to K-6. The estimated effect size for K-6 was .34 when the average values of the dictionary and translation variables were added to the regression equation, while the estimated effect size for grades 7-12 was -.09. Although dictionaries, including bilingual glossaries and picture dictionaries, were the most common accommodations and accounted for 46% of the effect

sizes, there was no statistically significant difference between a dictionary and other accommodations ( $\beta = -0.124$ ,  $p = .30$ ). Similarly, translation accommodation (26%) did not have a statistically significant effect on EL test performance compared to other test accommodations ( $\beta = -0.201$ ;  $p = .20$ ).

**Table 2**

The Moderator Analysis for EL Test Accommodations

Moderator	Moderator Model (k=8, n=28)			
	$I^2 = 37.45$ ; $\tau^2 = 0.016$ (SE=0.0131)			
	Estimate	SE.	95% CI	<i>p</i>
Intercept	0.462	0.134	0.174, 0.750	0.003**
Grade	-0.431	0.113	-0.662, -0.200	0.001***
Dictionary	-0.124	0.117	-0.365, 0.117	0.298
Translation	-0.201	0.151	-0.512, 0.110	0.195

Note. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

### Discussion

The findings of this study are similar to those found for paper-based accommodations for ELs. This study suggests that non-ELs' performance was not improved by CBT accommodations ( $-.02$  SD), which is consistent with Li and Suen's (2012b) meta-analysis. The evidence indicates that EL test accommodations on CBT do not provide unfair advantages to EL students.

This meta-analysis suggested that EL test accommodations on CBT have a small and statistically significant effect on improving EL students' academic performance. EL students with accommodations on CBT received .12 SD higher scores than EL students without test accommodations on CBT. Rios et al. (2020) meta-analysis found the average effect of test accommodations as .16 SD, which is the between confidence interval of this study. Even though Rios et al. (2020) and Li and Suen's (2012a) meta-analyses did not report a statistically significant difference between grade levels, this study showed that EL test accommodations on CBT were more effective for K-6 compared to grades 7-12. One explanation of this result could be the age difference in the grade level because younger learners are fast language learners. Hartshorne et al. (2016) indicated that language ultimate attainment is consistent when learners start before age of 10-12 (p.272) which is the similar age of 6 graders. Another reason could be that the language proficiency variation was higher among grades 7-12. Future studies should focus on how the use of accommodations and other factors (e.g., English language proficiency level) may vary across grade levels.

Although the dictionary is a common EL test accommodation, the results did not demonstrate a statistically significant effect of dictionary accommodation on CBT. This result is similar to Rios et al.'s (2020) and Li and Suen's (2012a) meta-analysis results. A small sample size could be a potential reason for the

statistically insignificant result, while limited evidence is available whether dictionaries increase content accessibility (Rios et al., 2020). On the other hand, lack of alignment between the language of instruction, including textbooks and teachers' instructional language, and EL test accommodations could be the reason for the statistically insignificant effect of some EL test accommodations (Abedi et al., 2020).

This meta-analysis provides some evidence about the validity and effectiveness of EL test accommodations on CBT. However, there are some limitations. First, the sample size of the meta-analysis was small. Even though CBT is a common practice for states' high-stakes academic testing, there is limited research that examines EL accommodations on CBT. Second, the small sample size could be the reason for the statistically insignificant effect of dictionary and translation test accommodation. Third, the EL population is a heterogeneous group of students, and accommodations should be provided based on an individual student's needs. This important feature of accommodations could not examine the effect of the EL population's diversity in terms of language proficiency level and student background information because the studies did not provide enough background information on participating EL students. Lastly, the quality of a meta-analysis based on the studies included we acknowledged that these eight studies are different in terms of sample size, grades, the number of accommodations used, types of accommodations, and how these accommodations were implemented.

### **Future Studies**

As the use of technology increases in educational settings, CBT will become the common assessment practice. The results of this study are important because the majority of states' large-scale assessments are administered via CBT. This meta-analysis provides some evidence that suggests EL test accommodations on CBT have a small effect on improving EL students' academic performance. In addition, this study revealed the need for more experimental studies examining CBT accommodations. CBT allows the seamless integration of accommodations into the testing process and permits gathering extensive data, such as response time and the frequency of accommodation use (Roohr & Sireci, 2017). Therefore, examining the effect of the frequency of accommodations on ELs academic achievement is a much-needed research area.

While Rios et al.'s (2020) meta-analysis included 26 experimental design studies about EL test accommodations, there were eight eligible studies on CBT for this meta-analysis. This study agreed with Rios et al.'s findings regarding EL test accommodation having limited evidence to support the effectiveness of accommodation. All studies in this meta-analysis were in math and science content areas so that future EL test accommodation research should include other content areas. Since this study indicated that EL test accommodations are more

effective in K-6 grades, future research should examine which accommodations are effective for different grade levels. ELs are a diverse group of students, which requires additional research to examine effective accommodations based on students' English language proficiency and other student characteristics. Since the benefits of accommodations are dependent on students' backgrounds and needs (Abedi, 2013), examining the moderating effects of these factors could provide insight into increasing the effectiveness of accommodations. Until additional empirical studies are conducted examining student characteristics, specific types of accommodations, and specific content areas, we will not fully understand the impact of CBT accommodations and potential methods for improving fairness and accessibility for all students.



### References

References marked with an asterisk indicate studies included in the meta-analysis.

- Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4–14.
- Abedi, J. (2006a). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, 108(11), 2282.
- Abedi, J. (2006b). Language issues in item-development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Erlbaum.
- \*Abedi, J. (2009). Computer testing as a form of accommodation for English language learners. *Educational Assessment*, 14(3-4), 195-211.
- Abedi, J. (2013). Validity issues in designing accommodations for English language learners. In *The Routledge handbook of language testing* (pp. 62-76). Routledge. Accessed on: 16 Jan 2020  
<https://www.routledgehandbooks.com/doi/10.4324/9780203181287.ch3>
- Abedi, J. (2014). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology*, 10(2), 1-30.
- Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Tech. Rep. No. 586). Los Angeles, CA: University of California, National Center for Research on Evaluation Standards and Student Testing.
- Abedi, J., & Levine, H. G. (2013). Fairness in assessment of English learners. *Leadership*, 42(3), 26–38. <https://eric.ed.gov/?id=EJ1004517>
- \*Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence - based computer accommodation decision system. *Educational Measurement: Issues and Practice*.
- 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.
- Abedi, J., Lord, C., Kim, C., & Miyoshi, J. (2001). *The effects of accommodations on the assessment of LEP students in NAEP* (CSE Tech. Rep. No. 537). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Acosta, B. D., Rivera, C., & Willner, L. S. (2008). *Best practices in state assessment policies for accommodating English language learners: A delphi study*. George Washington University Center for Equity and Excellence in Education.

- Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Projects Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Albus, D., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research*, 98(4), 245-256.
- \*Alt, M., Arizmendi, G. D., Beal, C. R., & Hurtado, J. S. (2013). The effect of test translation on the performance of second grade English learners on the KeyMath-3. *Psychology in the Schools*, 50, 27-36.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association and National Council on Measurement in Education.
- \*Ardasheva, Y., Wang, Z., Roo, A. K., Adesope, O. O., & Morrison, J. A. (2018). Representation visuals' impacts on science interest and reading comprehension of adolescent English learners. *Journal of Educational Research*, 111(5), 631–643. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1182786&site=ehost-live&scope=site>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, UK: John Wiley & Sons.
- Chiu, C. W. T., & Pearson, P. D. (1999, June). *Synthesizing the effects of test accommodations for special education and limited English proficiency students*. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.
- Cohen, D., Tracy, R., & Cohen, J. (2017). On the effectiveness of pop-up English language glossary accommodations for EL students in large-scale assessments. *Applied Measurement in Education*, 30(4), 259-272.
- Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage publications.
- Deysson, S. L. (2013). *Equity for limited English proficient students regarding assessment and effectiveness of testing accommodations: A study of third graders* (Order No. 3550425). [Doctoral Dissertation, The George Washington University]. Available from Education Database; ProQuest Central. (1287056937). Retrieved from <https://search-proquest-com.librarylink.uncc.edu/docview/1287056937?accountid=14605>
- Dolan, R. P., Burling, K. S., Rose, D., Beck, R., Murray, E., Strangman, N., Jude, J., Harms, M., Way, W. D., Hanna, E., Strain-Seymour, E., & Nichols, A.

- (2010). *Universal Design for Computer-Based Testing (UD-CBT) guidelines*. Iowa City, IA: Pearson.
- Every Student Succeeds Act. (2015). *Every Student Succeeds Act of 2015*, Pub. L. No 114 - 95 (2015 - 2016).
- Fairbairn, M. B. (2006). *English language learners' performance on modified science test item formats: A pilot study*. (Unpublished doctoral dissertation, University of Iowa, IA.)
- Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from <http://centeroninstruction.org/files/ELL3-Assessments.pdf>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5, 3–8.
- Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351–379.
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558.
- Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention*, 29(3), 35-45.
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79(3), 1168-1201.
- Kieffer, M. J., Rivera, M., & Francis, D. J. (2012). Based recommendations for the use of accommodations in large-scale assessments: 2012 Update. *Practical Guidelines for the Education of English Language Learners*. Book 4. Center on Instruction.
- \*Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, 26(3), 11-20.
- Li, H., & Suen, H. K. (2012a). The effects of test accommodations for English language learners: A meta-analysis. *Applied Measurement in Education*, 25(4), 327-346.
- Li, H., & Suen, H. K. (2012b). Are test accommodations for English language learners fair? *Language Assessment Quarterly*, 9(3), 293-309.
- McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). *The Condition of Education 2017* (NCES 2017- 144).

- US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144>.
- McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., and Bullock Mann, F. (2018). *The Condition of Education 2018* (NCES 2018-144). US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144>.
- Miley, S. K., & Farmer, A. (2017). English Language Proficiency and Content Assessment Performance: A Comparison of English Learners and Native English Speakers Achievement. *English Language Teaching*, 10(9), 198–207.
- National Assessment of Educational Progress (NAEP) (2017). *Academic performance and outcomes for English learners: Performance on national assessments and on-time graduation rates*. Washington, DC: US Department of Education. <https://www2.ed.gov/datastory/el-outcomes/index.html>
- National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs. (2007). Retrieved from [http://sbo.nn.k12.va.us/esl/documents/ncela\\_fast\\_faqs.pdf](http://sbo.nn.k12.va.us/esl/documents/ncela_fast_faqs.pdf)
- 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(3), 10-28.
- Polat, N., Zarecky-Hodge, A., & Schreiber, J. B. (2016). Academic growth trajectories of ELLs in NAEP data: The case of fourth- and eighth-grade ELLs and non-ELLs on mathematics and reading tests. *Journal of Educational Research*, 109(5), 541-553. doi:10.1080/00220671.2014.993461
- Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence - based? A multistudy systematic review and meta - analysis. *Educational Measurement: Issues and Practice*.
- Rivera, C. (2003). State assessment policies for English language learners. Paper presented at the Presented at the 2003 Large-Scale Assessment Conference.
- \*Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, 39(8), 582-590.
- Roohr, K. C., & Sireci, S. G. (2017). Evaluating computer-based test accommodations for English Learners. *Educational Assessment*, 22(1), 35-53.

- Russell, M., Hoffmann, T., & Higgins, J. (2009). Meeting the needs of all students: A universal design approach to computer-based testing. *Innovate: Journal of Online Education*, 5(4).
- Shafer Willner, L. S., Rivera, C., & Acosta, B. D. (2008). *Descriptive study of state assessment policies for accommodating English Language learners*. Arlington, VA; The George Washington University Center for Equity and Excellence in Education. Retrieved from <http://ceee.gwu.edu>
- Shepard, L. A., Taylor, G., & Betebenner, D. W. (1998). *Inclusion of limited-English-proficient students in Rhode Island's grade 4 mathematics performance assessment*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.
- Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3–13.
- \*Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., & Nguyen-Le, K. (2014). Developing testing accommodations for English language learners: Illustrations as visual supports for item accessibility. *Educational Assessment*, 19(4), 267-283.
- Stone, E. A. & Cook, L. L. (2018). Fair testing and the role of accessibility. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices*, Second edition, (pp. 59-73). Springer.
- Strategies, E. (2015, November 5). *Pencils down: The shift to online & computer-based testing*. EdTech Strategies. [https://www.edtechstrategies.com/wp-content/uploads/2015/11/PencilsDownK-8\\_EdTech-StrategiesLLC.pdf](https://www.edtechstrategies.com/wp-content/uploads/2015/11/PencilsDownK-8_EdTech-StrategiesLLC.pdf)
- Thurlow, M., Laitusis, C. C., Dillon, D. R., Cook, L. L., Moen, R. E., Abedi, J., & O'Brien, D. G. (2009). *Accessibility principles for reading assessments*. Minneapolis, MN: National Accessible Reading Assessments Projects.
- Thurlow, M., Lazarus, S. S., Albus, D., & Hodgson, J. (2010). *Computer-based testing: Practices and considerations (Synthesis Report 78)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Turkan, S., & Oliveri, M. E. (2014). *Considerations for providing test translation accommodations to English language learners on common core standards - based assessments*. ETS Research Report Series, 2014(1), 1-13.
- United States Department of Education (2015). *Every Student Succeeds Act*. Retrieved from <https://www.ed.gov/essa?src=rn>
- Wolf, M. K., Herman, J. L., Kim, J., Abedi, J., Leon, S., Griffin, N., et al. (2008). *Providing validity evidence to improve the assessment of English language*

*learners* (CSE Tech. Rep. No. 738). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

- \*Wolf, M. K., Yoo, H., Guzman-Orth, D., & Abedi, J. (2022). Investigating the Effects of Test Accommodations with Process Data for English Learners in a Mathematics Assessment. *Educational Assessment*, 27(1), 27-45.  
<https://doi.org/10.1080/10627197.2021.1982693>