Applied Measurement in Education, 27: 261–272, 2014

ISSN: 0895-7347 print/1532-4818 online
DOI: 10.1080/08957347.2014.944310

Routledge
Taylor & Francis Group

The Use of Computer Technology in Designing Appropriate Test Accommodations for English Language Learners

Jamal Abedi
School of Education, University of California, Davis

Among the several forms of accommodations used in the assessment of English language learners (ELLs), language-based accommodations are the most effective in making assessments linguistically accessible to these students. However, there are significant challenges associated with the implementation of many of these accommodations. This article reviews studies that explain the impact of language factors on the assessment of ELLs, presents major forms of language-based accommodations, and discusses how computers can assist in the proper administration of these accommodations.

Among multiple sources of construct-irrelevant variance (Haladyna & Downing, 2004), linguistic complexity deserves special consideration (e.g., Abedi, 2006). Large performance gaps between English language learners (ELLs) and their native English speaking peers are observed for items with high levels of language demand. Analyses of standardized test data from multiple states in the nation revealed that ELLs performed between 40 to 60% lower than non-ELLs on English language arts (ELA). The performance gap was substantially smaller (8% to 25%) in mathematics problem-solving items—which had lower levels of language demand—and almost null (0% to 10%) in mathematics computation—which had a minimal level of language demand (Abedi, Leon, & Mirocha, 2003).

Relevant to examining this performance gap is the distinction between language that is related to the focal construct (construct-relevant) and language that is not part of the focal construct (construct-irrelevant). A large performance gap between ELL and non-ELL students in ELA can be explained by the ELL students' lack of English language literacy. However, a large gap in mathematics, science, and social science may not necessarily be explained by ELL students' lack of content knowledge in those areas. Language factors unrelated to the focal construct may at least partly explain the gap (Solano-Flores, 2008), which affects the psychometric properties of assessments for ELLs. Internal consistency reliability coefficients are generally lower for ELLs and particularly evident in assessments with high levels of language demand (Abedi et al., 2003).

Such a gap in reliability may be partly explained by the impact of unnecessary linguistic complexity as a source of systematic measurement error, which violates one of the fundamental assumptions of classical test theory—that measurement error should have random distribution. As the number of measurement increases, the observed scores should become closer to the true scores (Allen & Yen, 1979; Thorndike, 2005). However, for ELL students, error of measurement

Correspondence should be addressed to Jamal Abedi, School of Education, University of California, Davis, One Shields Avenue, Davis, CA 95616-9150. E-mail: jabedi@ucdavis.edu

TABLE 1
Language-Based Accommodations

Accommodation	*Supporting Literature*
Bilingual (dual language) version of the test	Abedi, Courtney, Leon, Kao, & Azzam (2006); Anderson, Liu, Swierzbin, Thurlow, & Bielinski (2000); Duncan et al. (2005); Pennock-Roman & Rivera (2011)
Commercial bilingual dictionary	Abedi et al. (2005); Abedi, Hofstetter, & Lord (2004)
Commercial English dictionary	Abedi et al. (2005); Abedi, Hofstetter, & Lord (2004)
Customized bilingual dictionary	Abedi, et al. (2004); Young et al. (2008)
Customized English dictionary (content-related terms removed)	Abedi et al. (2004); Albus, Bielinski, Thurlow, & Liu (2001).
Modified English (also called simplified English and linguistic modification in the literature)	Abedi (2006); Abedi (2012); Sato, Rabinowitz, Gallagher, & Huang (2010)
Native language version of the test (e.g., Spanish)	Aguirre-Munoz (2000); Hofstetter (2003); Kieffer, Lesaux, Rivera, & Franci (2009)
Pop-up glossary (content-related terms excluded)	Abedi (2009); Kopriva, Emick, Hipolito-Delgado, & Cameron (2007)
Read-aloud test directions in student's native language	Young et al. (2008)
Read-aloud test questions	Acosta, Rivera, & Shafer Willner, (2008); Sireci, Li, & Scarpati (2003)

due to unnecessary linguistic complexity systematically impacts their performance regardless of the number of measurements and is correlated with the observed scores. Therefore, classical test theory, which is based on a unidimensional true score assumption, may not be applied meaningfully in ELL assessment (Abedi, 2006). Language-based accommodations are recommended to control for unnecessary linguistic complexity of assessment and to provide equal opportunity and access on standardized assessments for ELL students (see Table 1).

This article discusses research that examines the effectiveness and validity of accommodations intended to address the fundamental needs of ELL assessments related to their academic language proficiency and the use of computers in facilitating provision of these accommodations.

RESEARCH ON THE EFFECTIVENESS AND VALIDITY OF LANGUAGE-BASED ACCOMMODATIONS

Many different accommodations are used for ELL students (see, e.g., Abedi, 2012; Abedi & Ewers, 2013; Acosta, Rivera, & Shafer Willner, 2008; Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Pennock-Roman & Rivera, 2011). It is important to know how effective these accommodations are in making assessments more accessible for ELL students and whether these accommodations provide valid assessment outcomes. Below is a summary of research on some of the most commonly used accommodations.

Dual Language Version of the Test

Dual language test versions present a content-based assessment side by side in two languages, English and students' native language. This accommodation may include an aural presentation of

the test questions and answer choices in students' home language on a cassette (Pennock-Roman & Rivera, 2011). The increased length of a dual language translation necessitates generous time limits. There are few studies of this accommodation and results are mixed on its effectiveness (Duncan et al., 2005; Pennock-Roman & Rivera, 2011).

Commercial Bilingual Dictionary

Available bilingual dictionary content is typically directed toward more conversational language and as such may contain less academic language than a commercial English dictionary. The number of test words found in different commercial language dictionaries also may vary making the use of commercial bilingual dictionaries impractical. Abedi, Courtney, Mirocha, Leon, and Goldberg (2005) reported that a bilingual dictionary was not effective for grade 8 ELL students taking a science assessment; however, validity was not threatened since this accommodation did not have any impact on the focal construct.

Commercial English Dictionary

Providing a commercial English dictionary as an accommodation requires that the test administration package include a dictionary for every student. Cost, excessive dictionary size, unfamiliarity with the dictionary provided, and student discouragement with the accommodation are factors that contribute to its impracticality. Providing this accommodation varied in its effectiveness according to student characteristics. For example, this approach was not effective on grade 8 science and mathematics assessments although it was effective for students on a grade 4 science assessment (Abedi et al., 2005; Abedi, Hofstetter, & Lord, 2004).

Customized Bilingual Glossary

A customized bilingual glossary provides English glosses with home language translations tailored to match key non-content words found in the assessment. Several studies with science tests resulted in accommodated students scoring higher than non-accommodated students, although the difference did not reach statistical significance (Abedi et al., 2004). Students with a higher level of English proficiency benefited more than those with lower proficiency. Non-accommodated students' scores were not affected indicating that this accommodation did not compromise the validity of the assessment.

Customized English Dictionary

A customized English dictionary can be more user friendly for students and test administrators due to its reduced size; it contains words and terms from the test that are not content-related—they are not considered to be part of the focal construct. For example, terms such as "whole number," "decimal," or "fraction" are considered related to "Number Sense" mathematics standards. In contrast, complex linguistic structures such as passive voice, conditional clause and unfamiliar vocabulary can be considered as unnecessary linguistic complexity and should be avoided to the extent possible. Several studies report differing effectiveness results when students were provided with customized English dictionaries with content-related terms removed (Abedi et al., 2004).

Linguistically Modified English

Linguistic modification has been introduced as a methodology to increase the linguistic accessibility of test items for ELL students (Abedi, 2006, 2012; Sato, Rabinowitz, Gallagher, & Huang, 2010). Linguistic modification involves identifying and reducing complex linguistic structures in assessments that are not content-related and can lead to misunderstanding and misinterpretation of item content. Features such as passive voice construction, conditional/adverbial clauses, and unfamiliar vocabulary are considered for revisions/modifications (for a more detailed list of these features, see Abedi, 2006).

Native Language Version of the Test

Several studies have found that the language of instruction is an important moderator of the effectiveness of this accommodation (Hofstetter, 2003; Kieffer, Lesaux, Rivera, & Francis, 2009). These studies concluded that this accommodation is effective when the language of the test is aligned with the language of instruction. However, Robinson (2010) found that for students in grades K–1, a student's home language is a better indicator of the effectiveness of translation. Spanish speaking students with lower English proficiency also tended to perform higher with this accommodation than intermediate Spanish speakers.

Pop-Up Glossary

During a computer-based test administration, a pop-up glossary is activated when a student places his or her cursor over a glossed word, causing the gloss to appear on the computer screen. Several studies have found this to be an effective accommodation (Abedi et al., 2003; Kopriva et al., 2007). Kopriva et al. (2007) suggested that this accommodation is more effective when selected based on student's level of language proficiency and cultural proximity (i.e., time in U.S. school and native country schooling).

Read-Aloud Test Directions

Although there is little research on the use of read-aloud test directions as an accommodation, Kieffer et al. (2009) suggest that it is likely to be responsive to the needs of ELLs. Young et al. (2008) rarely observed significant differential item functioning (DIF) on test items where this accommodation was used.

Read-Aloud Test Questions

This accommodation is typically addressed by content area. Acosta et al. (2008) indicated the reading aloud test questions is effective for students at the lower level of English proficiency in mathematics, science, and history/social science. ELA are considered separately when using this accommodation because of the concern that reading aloud the test question may alter the construct (Sireci, Li, & Scarpati, 2003).

USE OF COMPUTERS IN FACILITATING APPLICATION OF LANGUAGE-BASED ACCOMMODATIONS

The widespread use of computer-based tests (CBT) in educational assessments is burgeoning. A survey of state websites in 2010 revealed that computer-based tests were implemented in 26 states with a total of 51 assessments either operationalized or in a field-test status (Thurlow, Lazarus, Albus, & Hodgson, 2010). Most notably, the computer-based testing approach adopted by the two national Race to the Top consortia, the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Partnership for Assessment of Readiness for College and Careers (PARCC) brings the concept of CBT and its applications and methodological advancement and issues into the forefront of the national assessment and accountability systems. As a new assessment mode, CBT has the potential to facilitate the effective implementation of language-based accommodations.

CBT allows test accommodations to be embedded in the programming to permit consistency or uniformity in their delivery (Russell, Hoffmann, & Higgins, 2009). For example, variability is reduced by having a screen reader present a read-aloud accommodation on a CBT versus administration by individual teachers. Russell and colleagues (2009) suggested that this consistency results in a greater chance for increased test validity. Whereas paper-and-pencil tests (PPT) require multiple copies of test booklets and an increased number of test administrators with special skills (e.g., dual language translations, read aloud of test directions and/or items in student's native language), a CBT allows "students' needs to be accommodated within the framework of the test" (Russel et al., 2009, p. 2). Dolan et al. (2009) also addressed the role of variance in computer-based tests using embedded accommodations. They noted the importance of restricting variance by designing items with "built in supports that reduce or remove sources of variance" (Dolan et al., 2009, p. 6).

The use of a CBT to deliver test accommodations is closely tied to Universal Design, defined as, "an approach that involves developing assessments for the widest range of students from the beginning while maintaining the validity of results from the assessment" (Thurlow et al., 2010, p. 10). Accommodations designed and built into the fundamental structure of computer-based system are likely to interact with all other tools which allows for multiple accommodations to be used simultaneously (Russell et al., 2009). For example, a read-aloud accommodation could be combined with a pop-up glossary or a bilingual test to meet the individual needs of a student.

Russell et al. (2009) described a computer-based system employing principles of Universal Design as one that:

> Requires developers to build features into the architecture of a system that allow accommodation tools to be accessed flexibly to meet the needs of each individual user. In a universally designed test-delivery system, all students across a testing program use the same standard interface and have access to high-quality tools and accommodations delivered in a controlled, standardized, and equitable manner. (p. 3)

Dolan et al. (2009) stated that although Universal Design of text is almost unattainable using a print format, digital means vastly simplify the process. Several upcoming digital formats (e.g., the National Instructional Materials Accessibility Standard, NIMAS, and the Digital Accessible Information System, DAISY) will result in "predictable and automated transformations of text into alternative formats" (p. 25) making flexible digital text routine.

EXAMPLES OF COMPUTER-BASED ASSESSMENTS WITH ACCOMMODATIONS

UCLA/CRESST Computer Assessment/Accommodation System

The National Center for Research on Evaluation, Standards and Student Testing (CRESST) developed an ELL assessment system in which several accommodations were incorporated into the system (Abedi, 2009). The performance of ELL and their native English speaking peers in Grades 4 and 8 in mathematics was compared under accommodated and non-accommodated conditions. The accommodations used in the study were: (1) a pop-up glossary, (2) a customized English dictionary, (3) extra testing time, and (4) small-group testing. The mathematics test for both grades 4 and 8 consisted of released items from NAEP and TIMSS. The study found that computer testing with embedded language-based accommodations was the most effective approach in making assessments more linguistically accessible for ELL students. The accommodations used in this study did not impact the focal construct; therefore, the use of accommodations did not compromise the validity of assessments.

UC Davis/CRESST Computer Assessment/Accommodation System

The study used a quasi-experimental design to compare the results of a computer-based mathematics test with the results of a paper-and-pencil (PPT) version (Abedi et al., 2012). A total of 618 ELL and non-ELL students in grade 4 were randomly assigned to take a mathematics test under CBT and PPT formats. Neither formats had time limits. Students under CBT had access to three accommodations: a pop-up glossary, a read aloud, and the option to change fonts. A pop-up glossary was prepared for 71 non-content words by two experienced grade 4 teachers. During the actual CBT administration, the glossary was activated when students placed their cursor over a glossed word, causing the gloss to appear on the computer screen. The second option was to have the test questions read aloud. A speaker icon allowed students to listen to a reading of the problem as many times as they selected the icon. Another CBT option permitted students to adjust the font size on the computer screen. Although this is not typically an ELL responsive accommodation, it was offered to facilitate possible vision difficulties based on computer screen size, font size, and the physical set-up of computer monitor, desk, and chair. Results indicated that the accommodations did not alter the focal construct. Results also indicated that students were interested in the computerized assessment with accommodations; they understood directions and used the accommodation options efficiently.

The Selection Taxonomy for English Language Learner Accommodations (STELLA)

Kopriva et al. (2007) developed STELLA—a computerized theory-driven decision process that informs the selection of accommodations for ELLs based on the data from multiple sources—and found that accommodations assigned to students based on their academic needs made assessments more accessible for ELL students. Results of the study showed that grade 3 and 4 students who received the accommodations recommended by STELLA scored significantly higher on a mathematics test than ELLs who were tested under no accommodations or those who received incomplete or not recommended accommodations.

Nimble Tools

Another example of a CBT with embedded accommodations is a system developed by Nimble Tools (Russell et al., 2009). This company, in collaboration with several state departments of education and other state and national organizations, developed a computer-based test accommodation system that incorporated a full range of accessibility and accommodation tools. A study conducted across 21 Florida schools that included 444 students found the computer-based Florida state standardized test with embedded accommodations feasible, useful, and effective (Russell, Johnstone, Higgins, & Hoffmann, 2008).

ONPAR

ONPAR methodology is a technology-based approach designed to communicate challenging content to students with limited English and/or literacy skills. The system uses multisemiotic representations and text-supported accommodation tools to substantially minimize the language load without altering the focal construct. A randomized study found that grade 4 and 8 ELLs with low English proficiency scored significantly better under ONPAR science items than under traditional released items measuring the same constructs, while the control group of non-ELLs scored similarly under both approaches (Kopriva, Gabel, & Cameron, 2009; see also Kopriva & Winter, 2013).

SELECTION OF APPROPRIATE ACCOMMODATIONS USING A COMPUTER-BASED TESTING SYSTEM

Among the most salient characteristics needed for accommodations to provide valid assessment outcomes for ELL students are effectiveness, validity, feasibility, and differential impact (Abedi, 2012). Accommodations are effective to the extent they make assessments more accessible to ELLs. Effectiveness is examined experimentally by randomly assigning ELLs to either a treatment group (in which they receive a particular accommodation) or a control group (in which they are tested under a standard condition with no accommodations). An accommodation is considered effective if ELL students in the treatment group show, on average, significant improvements in their performance over ELL students in the control group. To meet the validity requirements, accommodations should have a minimum or null impact on the performance of non-ELL students. In other words, accommodations should not alter the focal construct.

Feasibility and Ease of Administration

With access to student background data in a computerized assessment system, a teacher using a settings tool can assign accommodations that fit individual student needs and backgrounds. As discussed above, certain language-based accommodations can provide linguistically accessible assessments but are logistically difficult to implement with paper-and-pencil tests. Below we elaborate on feasibility and differential impact in a computerized accommodation system.

Commercial dictionary. One of the most commonly used language-based accommodations is a commercial dictionary. There are major logistical issues involved in the provision of this

accommodation under a paper-and-pencil testing condition. This accommodation can be administered to students in a paper-and-pencil mode in two forms. First, students bring their own dictionary. However, the different dictionaries that students may bring and the type of dictionary used may be important sources of measurement error. Second, a dictionary is provided as part of test administration. However, it would be a burden to test administrators to provide standard dictionaries to individual students. Computers help resolve these feasibility issues by providing electronic version of the same dictionary to all test takers and by allowing students to have easy access to different parts of the dictionary.

Customized dictionary. While confined to only those words found in the assessment, customized dictionaries add paper bulk to a paper-and-pencil test package. This accommodation often requires students to turn their attention to a separate page or set of papers. Several studies have shown that students do not regularly refer to them (see, e.g., Abedi, 2009). A computer based test eliminates the need for a customized dictionary by replacing it with a glossary (detailed below). Computers can improve feasibility of this accommodation by providing easy access and assigning the same customized dictionary to all the examinees.

Glossary. Glossaries can be implemented efficiently by computers. First, glossaries can be presented by computers in a "pop-up" form, next to the word in question rather than at the end of the test booklet. It is, therefore, more convenient for test takers to access glossaries while reading the text without having to leave a page to turn to the customized dictionary. Abedi (2009) found that students who had access to a pop-up glossary format by computer used glossaries substantially more than those who took the paper version of the test with the glosses at the end of the booklet. Second, and more importantly, students have the option of easily accessing glosses of the terms that they need rather than searching for the gloss from an end-of-test-booklet list of all glosses supplied by the test developers. Computers can make this accommodation more feasible to the test takers by providing access through the pop-up approach, as opposed to placing the glossary at the end of the test booklet.

Linguistic modification. Linguistic modification can be implemented more efficiently under a computerized assessment system than with paper versions of the tests. Typically, with the paper version, a linguistically modified version of an assessment is given to ELLs with varying levels of proficiency in English. Clearly, some ELL students at a lower level of English proficiency may benefit from more simplified language than others (Hofstetter, 2003). Therefore, a single form of linguistically modified assessments under the paper version may not work for all ELL students.

A computer adaptive testing (CAT) approach can offer linguistic modification opportunities to ELLs at different levels of English proficiency. Under the CAT testing environment, the linguistic modification approach can be adapted based on students' levels of English proficiency. This is a significant advantage, given the diversity among ELL (Solano-Flores, 2014 [this issue]) and linguistic minority groups in general (Ercikan, Roth, Simon, Sandilands, & Lyons-Thomas, 2014 [this issue]). Assuming that sufficiently large numbers of items are available and that all items are rated on their level of linguistic complexity, the system can start with presenting a test item with a certain level of linguistic complexity to the student. An incorrect response to this item could branch to another item measuring the same content but with less complex linguistic structure. This process can continue until the possibility of an incorrect response due to unnecessary linguistic

complexity of assessment is diminished. In fact, it might be a good practice to always start with the version of the test that has been linguistically modified to reduce or eliminate unnecessary linguistic complexities and then add more accommodations as they deemed helpful.

Differential Impact of Accommodations

The level of effectiveness of accommodations in making assessments accessible to ELLs largely depends on individual students' academic backgrounds (Abedi & Ewers, 2013). For example, for ELL students who are at a higher level of proficiency in English, accommodations such as a customized English dictionary or English glossary might be quite effective in making an assessment more linguistically accessible to them. On the other hand, for ELL students who are at a lower level of English proficiency but are quite fluent in their native language and are instructed in their native language, accommodations such as native language testing might be more beneficial.

As indicated above, students' level of English proficiency is one of the most important criteria in selecting appropriate accommodations for ELL students. ELL students' level of proficiency is typically measured by an English language proficiency (ELP) test, as required under No Child Left Behind (NCLB) Title III. The outcome of the ELP assessments is reported as a single composite score, which is the sum of four domain scores of reading, writing, listening, and speaking. ELP scores of the four domains can provide additional information in informing proper assignment of accommodations to ELL students. For example, the ELP listening domain score can help decide whether a read-aloud accommodation can be helpful. Students with lower listening ELP scores may not benefit from a read-aloud test direction or read-aloud test question accommodation as much as those with higher listening scores. Similarly, students with low writing scores may not do well with open-ended questions where extensive writing is required.

A computer-based testing system can greatly facilitate proper assignment of accommodations to ELL students based on the academic background information including their language background, both in English and native language. All the relevant student background data can be entered into the computer for accommodation decision making. ELP test scores (raw and scale scores) as well as ELP levels (typically, five) can be obtained from students' records in all four language domains. A research-based decision algorithm can be built into the computer system to select appropriate accommodations based on the existing student background data.

SUMMARY AND CONCLUSION

Several language-based accommodations have been proposed and are currently in use with the intent to reduce the academic achievement performance gap between ELLs and non-ELLs in large-scale assessment due to unnecessary linguistic complexity. Studies on the effectiveness of some of these accommodations show promising results (see, e.g, Abedi, 2007; Acosta et al., 2008; Kieffer et al., 2009). However, major issues concerning standardization, validity, differential impact, and feasibility or ease of implementation arise in PPT.

Computer-based testing approach uses modern technology to resolve obstacles in the provision of accommodations. Its most important advantage over paper-and-pencil testing is that it allows students to choose among many accommodations. For example, depending on their level

of proficiency in English and their native language, a student may have a choice of a bilingual dictionary or glossary. This decision can be easily implemented into the computer-based assessment system.

This article has outlined the challenges associated with implementation of many accommodations in a paper-and-pencil test administration and elaborated on how computer-based assessments can help in the administration of accommodations. A computer-based assessment system can provide feasible accommodations that are reliable and valid, standardized, differentially selected according to important student characteristics.

The new generations of assessments developed by the two Race-to-the-Top assessment consortia (Smarter Balanced and PARCC) are mainly computer-based and require accommodations that can be implemented validly and efficiently by computers. Abedi and Ewers (2013) developed a research-based system to help with their decisions selecting accommodations that are also implementable under computer-based assessment environment. The system provides recommendations based on existing research as well as experts' view of which accommodations are effective in making assessments more accessible to ELLs and students with disabilities and are implementable under computer testing environment. The system also identifies accommodations that are used in paper-based format but no longer effective and valid in computer-based testing.

In sum, computer-based assessment system brings more efficiency into the assessment of ELLs by providing accommodations that are effective in making assessments more accessible to these students and do not impact the focal construct. These effective and valid accommodations are extremely difficult to implement with fidelity in traditional, paper-and-pencil testing. Computer-based assessment can effectively incorporate accommodations that are shown to help ELL students without altering the focal construct.

ACKNOWLEDGMENTS

The author thanks Dr. Nancy Ewers for her excellent suggestions for revisions as well as for her contribution to the literature review section of the article. The author also thanks Kimberly Mundhenk for her very useful suggestions and revisions.

REFERENCES

Abedi, J. (2006). 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. (2007). Utilizing accommodations in the assessment of English language learners. In N. H. Hornberger (Ed.), *Encyclopedia of language and education: Vol. 7: Language testing and assessment* (pp. 331–348). Heidelberg, Germany: Springer Science + Business Media.

Abedi, J. (2009). Computer testing as a form of accommodation for English language learners. *Educational Assessment*, *14*, 195–211.

Abedi, J. (2012). Validity issues in designing accommodations. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 48–62). Florence, KY: Routledge, Taylor & Francis Group.

Abedi, J., Bayley, R., Ewers, N., Mundhenk, K., Leon, S., & Kao, J. (2012). Accessible reading assessments for students with disabilities. *International Journal of Disability, Development and Education*, *59*(1), 81–95.

Abedi, J., Courtney, M., Leon, S., Kao, J., and Azzam, T. (2006). *English language learners and math achievement: A study of opportunity to learn and language accommodation* (CSE Report 702, 2006). Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification* (CSE Report 666). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., & Ewers, N. (2013). *Accommodations for English language learners and students with disabilities: A research-based decision algorithm*. Smarter Balanced Assessment Consortium. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/08/Accomodations-for-under-represented-students.pdf

Abedi, J., Hofstetter, C., & Lord, C. (2004) Assessment accommodations for English language learners: Implications for policy based research. *Review of Educational Research. 74*(1), 1–28.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students' language background on content based assessment: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Acosta, B. D., Rivera, C., & Shafer Willner, L. (2008). *Best practices in state assessment policies for accommodating English language learners: A delphi study*. Arlington, VA: George Washington University Center for Equity and Excellence in Education.

Aguirre-Muñoz, Z. (2000). The impact of language proficiency on complex performance assessments: Examining linguistic accommodation strategies for English language learners (Doctoral dissertation, University of California at Los Angeles). Retrieved from Proquest Dissertations and Theses Full Text (Publication no. AAT 9973171).

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. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/LEP1.html

Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., & Bielinski, J. (2000). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase 2* (Minnesota Report No. 31). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Dolan, R. P., Burling, K. S., Harms, M., Beck, R., Hanna, E., Jude, J., . . . Way, W. (2009). *Universal design for computer-based testing guidelines*. Iowa City, IA: Pearson.

Duncan, T., Parent, L., Chen, W., Ferrara, S., Johnson, E., Oppler, S., & Shieh, Y. (2005). Study of a dual-language test booklet in eighth grade mathematics. *Applied Measurement in Education*, *18*(2), 129–161.

Ercikan, K., Roth, W.-M., Simon, M., Sandilands, D., & Lyons-Thomas, J. (2014). Inconsistencies in DIF Detection for Sub-Groups in Heterogeneous Language Groups. *Applied Measurement in Education*, *27*, 273–285.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27.

Hofstetter, C. (2003). Contextual and mathematics accommodation test effects for English-language learners, *Applied Measurement in Education*, *16*(2), 159–188.

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.

Kopriva, R., Emick, J., Hipolito-Delgado, C., & Cameron, C. (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.

Kopriva, R. J., Gabel, D., & Cameron, C. (2009). *Overview of results from the ONPAR elementary and middle school science experimental study with ELs and non-ELs: A promising new approach for measuring complex content knowledge of English learners with lower proficiency levels*. Retrieved from http://onpar.us/research/reports.html

Kopriva, R. J., & Winter, P. C. (2013). *Assessing the knowledge, skills, and abilities of ELs, selected SwDs and controls on challenging high school science content: Results from randomized trials of ONPAR and technology-enhanced traditional end-of-course biology and chemistry tests*. Retrieved from http://onpar.us/research/reports.html

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.

Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, *39*, 582–590.

Russell, M., Hoffmann, T., & Higgins, J. (2009). Meeting the needs of all students: A universal design approach to computer-based testing. *Innovate*, *5*(4).

Russell, M., Johnstone, C., Higgins, J., & Hoffmann, T. (2008). *FCAT computer accommodations pilot study final report*. Tallahassee, FL: Florida Department of Education.

Sato, E., Rabinowitz, S., Gallagher, C., & Huang, C.W. (2010). *Accommodations for English language learner students: The effect of linguistic modification of math test item sets*. (NCEE 2009-4079). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Sireci, S., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature* (Report No. 485). Amherst, MA: University of Massachusetts Amherst, Center for Educational Assessment Research.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, *37*(4), 189–199.

Solano-Flores, G. (2014). Probabilistic approaches to examining linguistic features of test items and their effect on the performance of English language learners. *Applied Measurement in Education*, *27*, 236–247.

Thorndike, R. (2005). *Measurement and evaluation in psychology and education*. Upper Saddle River, NJ: Pearson.

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.

Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., & Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment*, *13*, 170–192.