

MEANINGFUL MEASUREMENT

The Role of Assessments in Improving High School Education in the Twenty-First Century

June 2009



© 2009 Alliance for Excellent Education. All rights reserved.

Suggested citation:

L. M. Pinkus, ed., *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* (Washington, DC: Alliance for Excellent Education, 2009).

Ordering information:

Copies of *Meaningful Measurement: The Role of Assessments in Improving High School Education in the Twenty-First Century* can be downloaded from the Alliance's website at www.all4ed.org. To request print copies of the report, please visit http://www.all4ed.org/publication_material/order_form. The first copy of the report is complimentary. Additional copies are available at a charge of \$1 per copy to cover shipping and handling costs.

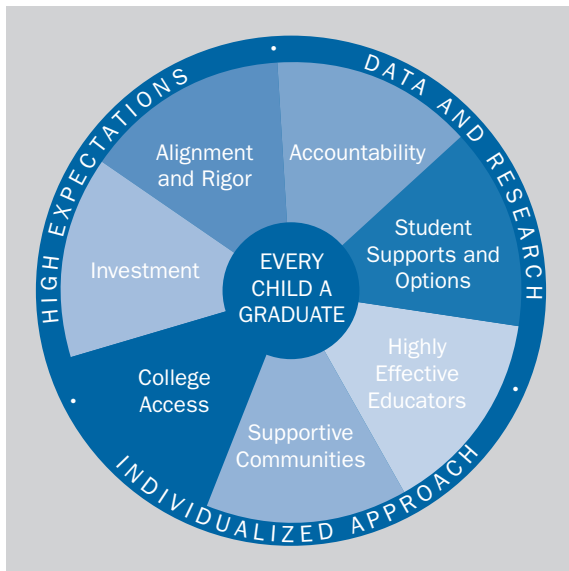
About the Alliance for Excellent Education

The mission of the Alliance for Excellent Education is to promote high school transformation to make it possible for every child to graduate prepared for postsecondary learning and success in life.

The Alliance for Excellent Education is a national policy and advocacy organization, based in Washington, DC, working to improve national and federal policy so that all students can achieve at high academic levels and graduate high school ready for college, careers, and citizenship in the twenty-first century.

The Alliance has developed a “Framework for Action to Improve Secondary Schools” that informs a set of federal policy recommendations based on the growing consensus of researchers, practitioners, and advocates about the challenges and solutions for improving secondary student learning.

The framework, shown graphically here, encompasses seven policy areas that represent key leverage points in ensuring a comprehensive, systematic approach to improving secondary education. The framework also captures



three guiding principles that apply to all of the policy areas. Although the appropriate federal role varies from one issue area to another, they are all critically important to reducing dropouts and increasing college and career readiness.

About the Editor

Lyndsay M. Pinkus is director of strategic initiatives at the Alliance for Excellent Education. Since joining the Alliance in January 2002, she has served in a variety of research, coordination, and advocacy roles, where her work has included managing policy and grant work on a range of issues including graduation rates, data, secondary school accountability, and secondary school improvement, and authoring a number of publications for the Alliance. Prior to rejoining the staff in January 2006, Ms. Pinkus served as a legislative associate at Washington Partners, LLC, providing government relations and policy research and analysis for a variety of clients, including the Alliance. She is a graduate of the School of Public Affairs at American University as a presidential scholar; the Public Affairs and Advocacy Institute at the Center for Congressional and Presidential Studies; and the Institute for Educational Leadership's Education Policy Fellowship program.

Acknowledgments

The Alliance for Excellent Education is greatly appreciative of the authors for sharing their time and expertise in writing the following chapters, as well as of the multiple Alliance staff members and advisors whose dedication contributed significantly to this volume.

The Alliance for Excellent Education is also grateful to Carnegie Corporation of New York for the financial support that made this publication possible.

The views expressed in this volume are those of the authors and do not necessarily represent those of the Alliance for Excellent Education or the funder.

CHAPTER

Assessing High School English Language Learners

Jamal Abedi

University of California at Davis

As the number of English language learners (ELLs) in the American school system grows, issues regarding their education need to be given more attention. ELL enrollment has grown 57 percent since 1995, while the rate for all students has been at less than 4 percent. Currently, there are 5.1 million ELL students, forming more than 10 percent of the country's student population.¹ Because of the rapid growth of this group, we need to accurately determine which ELL students require English language services, and then work to support all of their academic needs.

English language learners often differ with respect to their sociocultural background, parents' level of education, ethnic background, family characteristics, and level of native and English language fluency.² But they all share a common need for help with English language proficiency.

The federally mandated inclusion of ELLs in state assessment systems necessitates an examination of how assessments of both English language

proficiency and content knowledge affect these students' academic lives. It is imperative to identify and examine factors that affect the academic performance of ELL students in order to provide ways to effectively deal with the assessment and instructional issues facing this population.

While the issues concerning instruction and assessment for ELL students are ultimately inseparable, this chapter focuses on the assessment issues that arise for these students and offers federal policy recommendations that will help ensure successful academic careers for all ELLs.

The Role of Assessment in ELL Students' Academic Career

Traditionally, students are taught and then tested to assess what they have learned. For ELL students, however, assessment comes before instruction begins. This is due to the fact that ELL students' level of English language proficiency (ELP) must first be evaluated so ELLs can be properly placed into appropriate instructional settings when an English-only instructional environment is the preferred choice. ELL students who are in English-only instructional classes and required to take assessments in English without having mastered the necessary level of English proficiency are at risk of failure.

For ELL students, performance on ELP assessments is the main criterion for classification into the ELL category and for reclassification from ELL to RFEP (reclassified fluent English proficient). Improper classification may lead to inappropriate and inadequate instruction and may also affect accountability, such as in the reporting of Adequate Yearly Progress³ for ELLs. Misleading results of invalid ELP assessment and inaccurate classification may lead to the disproportionate placement of ELL students in special education classrooms, which can negatively affect both their academic career as a whole and the time it takes them to graduate.⁴ Furthermore, ELLs, like their peers, are subject to content-area assessments that have student-level implications, including grades, promotion, and graduation, as well as system-level implications, including AYP determinations. Thus, it is clear that assessment outcomes profoundly impact ELLs' academic performance. Proper attention must therefore be given to the ways that tests are developed, field tested, and reported for these students.

Assessment issues for ELLs are more complicated and important at the high school level than in the previous grades. Not only are assessment materials heavily impacted by linguistic factors because of the more complex language used at this level, but high-stakes decisions about students' academic performance are also made more frequently during these years. As research on the classification of ELL students reveals, there is increased pressure on schools to reclassify ELL students out of the ELL category at higher grade levels.⁵ If students are reclassified prematurely, they may not have enough language proficiency to meaningfully participate in content-area assessments. Many states, for instance, require high school students to take the state high school exit examination in order to graduate. As with many other state assessments, these tests often suffer from cultural and linguistic biases. ELL students may be unable to pass such exams, not because of a lack of content knowledge, but because of a lack of a thorough understanding of the exit examination language.

Assessment of English Language Proficiency

To make sure that ELL students are ready to take state content-area assessments in English, their level of academic English proficiency must be gauged. If they are not at a level where they can meaningfully participate in assessments conducted in English, their content-area assessment outcomes may not be valid. ELP assessments include cut scores for determining the level of English proficiency. While there are some differences between the reporting policies of ELP scores among states, Level 1 usually refers to no or very low proficiency in English and Level 5 represents high proficiency. ELL students are typically reclassified from LEP to fluent English proficient (FEP) at ELP Level 4 or above.⁶

Issues regarding ELP assessment, including those that were developed and used prior to the implementation of the No Child Left Behind Act of 2001 (NCLB) and those developed in accordance with federal regulations accompanying NCLB, are explored below.

Status of ELP assessments prior to NCLB

There were many English language proficiency assessments available for public use prior to the implementation of NCLB. These assessments

were based on one or more of at least three different schools of thought, and thus provided differing measures of proficiency.⁷ Reviews of the pre-NCLB ELP assessments found major variations in the content, structure, test administration procedures, theoretical bases, and issues related to the validity and reliability of the tests. Researchers found that the assessments also differed in their approaches to defining language proficiency, the types of tasks and specific item content, the grade-level ranges, and the specific time limits.⁸

Similarly, a review of the content and psychometric characteristics of some of the most commonly used English language proficiency tests prior to NCLB found major differences between these tests with respect to their purpose, age and language group, administration, cost, items, scoring, test design, theoretical foundation, and reliability and validity.⁹ Such discrepancies are cause for concern, as their outcomes may not be comparable and may negatively impact the authenticity of the English language proficiency assessments.¹⁰

Impact of NCLB on ELP assessments

Title III of NCLB requires states receiving Title I funding to annually assess ELL students' level of English language proficiency using reliable and valid measures in four areas: reading, writing, listening, and speaking. These assessments must be aligned with the state's ELP content standards and should measure academic English.

At least four consortia of states developed four batteries of ELP assessments based on the NCLB Title III guidelines.¹¹ These assessments included all four modalities required by NCLB Title III and were aligned with the states' ELP content standards. They were developed for four or more grade clusters (typically K–2, 3–5, 6–8, and 9–12) and included common sets of items across adjacent grade clusters. The test developers conducted extensive pilot and field testing on large and representative samples of students. The total tests and the content and psychometric properties of the individual items were then carefully examined, and changes were made where needed.

There has clearly been improvement in the content and psychometric properties of the post-NCLB English language proficiency assessments as

compared to those created prior to NCLB implementation. However, there are still significant problems that need to be resolved before the assessments can be safely used, including the following:

- the lack of a commonly acceptable definition of English language proficiency;
- issues concerning standard settings for the newly developed ELP assessments;
- issues concerning dimensionality of scores obtained from the four modalities and reporting these scores;
- comparability of the new assessments with the pre-NCLB assessments in establishing the baseline; and
- the lack of an objective definition of the concept of academic English.¹²

Assessment of Content-Area Knowledge

NCLB requires English language learners to be assessed in the content areas in which all other students are required to be tested: reading, language arts, math, and science. Results from these standardized academic achievement tests are then used in high-stakes assessment and accountability decisions for high school students, such as classification/reclassification,¹³ promotion, and graduation.

There are a number of challenges related to assessing ELLs in the content areas. Many critics believe that the tests used for these purposes are not appropriately designed for such use with ELLs,¹⁴ and feel that there should be standardized achievement tests specifically designed to assess these students' content knowledge.¹⁵

Requiring English proficiency for participation in content-area assessments

Many state education officials and ELL assessment experts believe that ELL students should take content-area assessments in English only when they are proven proficient enough in English. In many states, students are considered proficient in English if they score at proficiency Level 4 or higher, but this is not consistent across the country.¹⁶ Defining an appropriate level

of proficiency is of paramount importance for high school students, particularly in states that administer high school exit examinations. Since these exams are mostly presented in English, ELL students with a lower level of English proficiency may not be able to fully demonstrate their content knowledge.

Impact of linguistic factors on the assessment of ELL students

The main issue with many of the achievement test items that are developed for native speakers of English is that there may be cultural and linguistic biases that affect the validity and reliability of the assessments. Research on the assessment of ELL students clearly indicates that unnecessary linguistic complexity of test items is a source of measurement error, and that construct-irrelevant variance may threaten the validity of standardized achievement tests for ELLs.¹⁷ Researchers have found that linguistically complex items largely contribute to the measurement error for ELL students, which may cause lower reliability and result in the misinterpretation and misunderstanding of test questions.¹⁸

Research has also demonstrated that unnecessary linguistic complexity of test items contributes to the performance gap between ELL and non-ELL students.¹⁹ The higher the level of linguistic complexity, the larger the performance gap between ELL and non-ELL students. In this way, language factors play an important role in the assessment of ELL students, particularly at higher grade levels.

There is a substantial performance gap between ELLs and their native English speaker peers in all content areas, and this gap widens as the level of language demand in assessments increases.²⁰ There are many linguistic features that make the comprehension of assessment materials difficult for English language learners, including unfamiliar vocabulary, complicated grammatical structures, and styles of discourse that include extra material, abstractions, and passive voice.²¹ Reports of studies on the impact of language factors on the assessment of ELLs have included a comprehensive review of linguistic features that affect performance outcomes of ELL students, along with citations to relevant research.²²

To make content-based assessments more accessible to ELL students, the concept of a linguistic modification approach has been suggested.²³ The main theme underlying this method is to reduce or eliminate unnecessary linguistic complexities in order to make assessments more reliable, more valid, and more accessible for ELLs. Under this approach, the linguistic features that make assessments more complex are identified and then revised. For example, ELL students have been shown to have difficulty with unfamiliar vocabulary, passive voice, conditional clauses, long and complex phrases, relative clauses, and long nominals. In the linguistic modification, unfamiliar or infrequent words are changed to familiar words, passive verbs are changed to active verbs, conditional clauses are replaced with separate sentences or the order of conditional and main clauses are changed, complex question phrases are changed to simple question words, relative clauses are either removed or recast, and long nominals are shortened. Below is an example of an original test item and a proposed linguistically modified version of that item. As can be seen from this example, multiple sources of linguistic complexities were involved, and multiple modifications were performed.

Original test item:

The census showed that three hundred fifty-six thousand, ninety-seven people lived in Middletown. Written as a number, that is

- A. 350,697
- B. 356,097
- C. 356,907
- D. 356,970

Modified test item:

Janet's video game score was three hundred fifty six thousand, ninety-seven. Written as a number that is

- A. 350,697
- B. 356,097
- C. 356,907
- D. 356,970

The main issue in the implementation of the linguistic modification approach is how to decide which linguistic features are necessary and

relevant to the construct being measured and which are unnecessary and irrelevant. Researchers highly recommend that a team of specialists, including math content experts, linguists, and test item developers, decide what language in the test items is considered unnecessary and needs to be modified.²⁴

The use of accommodation in the assessment of ELL students

It is recommended that accommodations be used to provide fair content-area assessment for ELL students. Literature on the accommodations for ELL students shows that many different types of accommodations are already provided to ELL students, but some of them may not be relevant or effective. For example, some of these accommodations may alter the construct being measured; therefore their validity might be questionable, particularly for high school students.²⁵

In an analysis of seventy-three accommodations used for ELL students across the nation,²⁶ it was shown that only eleven of them (15 percent) were deemed appropriate for ELL students. Below are examples of accommodations that are used for ELL students that may not be as relevant:

- enlarged answer sheets;
- multiple breaks throughout the testing period;
- the administration of tests to individual students;
- the administration of tests in small groups; and
- the administration of tests in locations with minimal distraction.

Researchers studying accommodations used in administering the National Assessment for Educational Progress found that the use of accommodations such as one-on-one testing, small-group testing, extended time, and oral reading of directions did not help to improve the performance of ELL students or reduce the performance gap between ELLs and native English speakers in content-based assessments.²⁷

To reduce the performance gap, ELLs need help with the language of assessment and instruction. As indicated earlier, assessment and instructional materials that are developed mainly for native speakers of English may contain linguistic structures that are too complex for non-

native speakers, so accommodations that help ELLs cope with language issues would be the most effective way to increase their assessment scores. But none of the accommodations mentioned above directly address ELL language needs. Research suggests that providing language-based accommodations such as customized dictionaries or glossaries or a linguistically modified version of the assessment helps ELL students present a more valid picture of what they know and can do.²⁸

Methodological Issues Related to the Assessment of ELL Students

The fundamental principle underlying any assessment for all students is the validity of the assessments. If tests are not valid, then their outcomes can be misleading and can negatively impact students' academic life. There are many different ways to determine assessment validity for students, including content, criterion, consequential, and construct validity approaches.²⁹ Below is a brief discussion of the concerns with regard to each of those approaches.

Content validity: For an assessment to have content validity, it must include a representative sample of the universe of all possible test questions based on the relevant content standards.³⁰ Assessments may not be valid for ELL students in terms of content, since they may include language content that is unrelated to the focal content being measured.

Criterion-related validity: To be valid in terms of a criterion-related approach, the assessment outcome must be highly correlated with a valid criterion or criteria.³¹ Examining the criterion-related validity for ELL students may be technically unfit, since it is difficult to find measures that are free of linguistic and cultural biases.

Consequential validity: Assessments can be consequentially valid if the consequences of a particular use or interpretation of assessment results are important in arriving at an overall evaluative judgment of the validity of the assessment for that use or interpretation.³² Content-based assessments that are developed for native English speakers are particularly problematic for ELL students in terms of consequential validity. These assessments are used for many different purposes in the ELLs' academic career, and some of these—such as those used for classification,³³ accountability,³⁴ and

graduation—may not be relevant.³⁵ Due to technical issues involved in these assessments for ELLs, the outcomes may be misleading.

Construct validity: Construct validation is the most fundamental and important approach in the validity of assessments, particularly with regard to academic assessments. Assessments have construct validity if they measure the intended construct and nothing else. Literature on the assessment of ELL students demonstrates that unnecessary linguistic complexity of content-based assessment is a source of construct-irrelevant variance and negatively impacts the validity of ELL assessments.³⁶ Language factors in content-based assessments add a new dimension and reduce internal consistency between test items within a test.

Conclusion

Assessment plays a vital role in the academic career of ELL high school students. It shapes their classification evaluations, is used for accountability purposes, and helps educators make decisions about promotion and graduation. It therefore constitutes the foundation of ELL students' educational career. Even a minor problem in test development, field testing, and scoring could cause serious consequences in ELL students' academic lives.

As the content and linguistic structure of assessments become more complex at the high school level, so do the choice and effectiveness of ELL accommodation strategies. Thus, a sound decision on which accommodations should be used in the assessment of ELL students requires a correspondingly complex set of criteria. Unfortunately, there are very few research-supported accommodations available for high school students, and even those that are supported by research are not frequently used by schools.³⁷ It is therefore imperative to carefully examine accommodation needs for these students and to provide accommodations that properly address those needs.

Federal Policy Recommendations

Federal legislation such as the No Child Left Behind Act and its predecessor, the Improving America's Schools Act of 1994, addresses the need to advance

the quality of teaching and learning for *every* child, including English language learners. These laws mandate inclusion of ELL students in large-scale state and national assessments. However, mandating inclusion of these students alone may not produce desirable outcome unless attention is paid to the quality of ELL instruction and assessments. The first and most important step in providing reliable and valid assessments for ELL students is to understand the complex nature of the ELL assessment system and to identify variables that impact such assessments. There are many issues that have a profound interactive impact on the assessment of this particular subgroup of students. These issues affect ELL classification, accountability, promotion, and graduation. Federal policymakers should consider the following recommendations:

1. Encourage state assessment divisions and test publishers to clearly examine factors that hinder the accessibility of assessments for ELL students. Examples of these factors include unnecessary linguistic complexity and cultural biases. A linguistic and cultural bias review process should be included in the test development process to address these issues.
2. Ensure that tests used for high-stakes purposes, including classification, promotion, and graduation, are carefully reviewed for any validity and accessibility concerns.
3. Support the use of multiple criteria for high-stakes decisions; a single criterion may not provide the preponderance of evidence that is needed for such decisions.
4. Support states and test publishers to provide assessments that are sensitive to ELL students' language and cultural needs. Only such assessments should be used for high-stakes decisionmaking.
5. Include formative assessment in the curriculum for all students, particularly for ELLs. The outcome of formative assessment would be extremely helpful in improving the academic performance of ELL students.

6. Support states in providing testing accommodations that are relevant for these students and are effective and valid in making assessments fair and accessible, while recognizing that many current accommodations may not be relevant for these students, or may even invalidate the assessment outcomes.

7. Encourage research on the effectiveness and validity of accommodations currently used by states.

The views expressed in this chapter are those of the author and do not necessarily represent those of the Alliance for Excellent Education.

About the Author

Jamal Abedi is a professor at the School of Education of the University of California at Davis and a research partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). His research interests include psychometrics and test and scale developments. Among his recent work are studies on the validity of assessments and accommodations and the opportunity to learn for English language learners (ELLs). He is the recipient of the 2003 Professional Service Award in recognition of “Outstanding Contribution Relating Research to Practice,” from the American Educational Research Association, and the 2008 Lifetime Achievement Award, from the California Educational Research Association. Dr. Abedi’s educational background is in psychometrics and research methodology. He holds a master’s degree and a PhD in psychometrics from Vanderbilt University.

¹ M. E. Flannery, “A New Look at America’s English Language Learners,” *NEA Today*, <http://www.nea.org/home/29160.htm> (accessed January 5, 2009).

² J. Abedi and P. Gandara, “Performance of English Language Learners as a Subgroup in Large-Scale Assessment: Interaction of Research and Policy,” *Educational Measurement: Issues and Practice* 25, no. 4 (2006): 36–46.

³ No Child Left Behind Act of 2001, HR 1, 107th Cong., 1st sess.

⁴ J. Abedi, “Classification System for English Language Learners: Issues and Recommendations,” *Educational Measurement: Issues and Practices* 27, no. 3 (2008): 17–22.

⁵ Abedi, “Classification System for English Language Learners.”

⁶ J. Abedi, ed., *English Language Proficiency Assessment in the Nation: Current Status and Future Practice* (Davis, CA: University of California Press, 2007).

- ⁷ G. Valdes and R. A. Figueroa, *Bilingual and Testing: A Special Case of Bias* (Norwood, NJ: Ablex, 1994).
- ⁸ A. M. Zehler et al., *An Examination of Assessment of Limited English Proficient Students* (Arlington, VA: Development Associates, Special Issues Analysis Center, 1994).
- ⁹ A. Del Vecchio and M. Guerrero, *Handbook of English Language Proficiency Tests* (Washington, DC: National Clearinghouse for Bilingual Education, 1995).
- ¹⁰ For a more detailed description of this issue, see J. Abedi, "Measuring Students' Level of English Proficiency: Educational Significance and Assessment Requirements," *Educational Assessment* 13 (2008): 193–214.
- ¹¹ See Abedi, *English Language Proficiency Assessment in the Nation*.
- ¹² For a more detailed discussion of issues concerning the newly developed ELP assessments, see Abedi, "Measuring Students' Level of English Proficiency."
- ¹³ A. L. Kindler, *Survey of the States' Limited English Proficient Students and Available Educational Programs and Services, 2000–2001 Summary Report* (Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, 2002).
- ¹⁴ K. S. Mahoney and J. MacSwan, "Reexamining Identification and Reclassification of English Language Learners: A Critical Discussion of Select State Practices," *Bilingual Research Journal* 29 (2005): 31–42.
- ¹⁵ E. H. Stefanakis, *Whose Judgment Counts?: Assessing Bilingual Children, K–3* (Portsmouth, NH: Heinemann, 1998).
- ¹⁶ Abedi, "Measuring Students' Level of English Proficiency."
- ¹⁷ J. Abedi, "Language Issues in Item-Development," in *Handbook of Test Development*, ed. S. Downing and T. Haladyna, 377–98 (Mahwah, NJ: Lawrence Erlbaum Associates, 2006); R. A. Figueroa, "Psychological Testing of Linguistic-Minority Students: Knowledge Gaps and Regulations," *Exceptional Children* 56, no. 2 (1989): 145–53; R. A. Figueroa, "Best Practices in the Assessment of Bilingual Children," in *Best Practices in School Psychology*, ed. A. Thomas and J. Grimes, 93–106 (Washington, DC: National Association of School Psychologists, 1990); T. M. Haladyna and S. M. Downing, "Construct-Irrelevant Variance in High-Stakes Testing," *Educational Measurement: Issues and Practice* 23, no. 1 (2004): 17–27; S. Messick, "The Interplay of Evidence and Consequences in the Validation of Performance Assessments," *Educational Researcher* 23, no. 2 (2004): 13–23; G. Solano-Flores and E. Trumbull, "Examining Language in Context: The Need for New Research and Practice Paradigms in the Testing of English-Language Learners," *Educational Researcher* 32, no. 2 (2003): 3–13.
- ¹⁸ Solano-Flores and Trumbull, "Examining Language in Context."
- ¹⁹ Abedi, "Language Issues in Item-Development"; J. Abedi and C. Lord, "The Language Factor in Mathematics Tests," *Applied Measurement in Education* 14, no. 3 (2001): 219–34; Solano-Flores and Trumbull, "Examining Language in Context"; Valdes and Figueroa, *Bilingual and Testing*.
- ²⁰ Abedi, "Language Issues in Item-Development"; Solano-Flores and Trumbull, "Examining Language in Context."
- ²¹ Abedi, "Language Issues in Item-Development"; J. Abedi, C. Lord, and J. Plummer, *Language Background as a Variable in NAEP Mathematics Performance* (Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing, 1997).
- ²² Readers who are interested in a more detailed discussion of linguistic features are referred to Abedi, "Language Issues in Item-Development," and Abedi, Lord, and Plummer, *Language Background*.
- ²³ See, for example, Abedi, "Language Issues in Item-Development"; Abedi, Lord, and Plummer, *Language Background*.
- ²⁴ Abedi, Lord, and Plummer, *Language Background*.

- ²⁵ J. Abedi, C. Hofstetter, and C. Lord, "Assessment Accommodations for English Language Learners: Implications for Policy-Based Empirical Research," *Review of Educational Research* 74, no. 1 (2004): 1–28; C. Rivera et al., "Study 1: An Analysis of State Assessment Policies Regarding the Accommodation of English Language Learners," in *State Assessment Policy and Practices for English Language Learners*, ed. C. Rivera and E. Collum, 1–174 (Mahwah, NJ: Lawrence Erlbaum Associates, 2006); S. G. Sireci, S. Li, and S. Scarpati, *The Effects of Test Accommodation on Test Performance: A Review of the Literature* (Amherst, MA: School of Education, University of Massachusetts, Amherst, 2003).
- ²⁶ C. Rivera, "State Assessment Policies for English Language Learners," presented at the Council of Chief State School Officers' 33rd Annual Conference on Large-Scale Assessment, June 2003, San Antonio, TX.
- ²⁷ J. Abedi and F. Hejri, "Accommodations for Students with Limited English Proficiency in the National Assessment of Educational Progress," *Applied Measurement in Education* 17, no. 4 (2004): 371–92.
- ²⁸ Abedi, Hofstetter, and Lord, "Assessment Accommodations for English Language Learners"; V. L. Kiplinger, C. A. Haug, and J. Abedi, "Measuring Math—Not Reading—on a Math Assessment: A Language Accommodations Study of English Language Learners and Other Special Populations," paper presented at the annual meeting of the American Educational Research Association, April 2000, New Orleans, LA; N. A. Maihoff, "Using Delaware Data in Making Decisions Regarding the Education of LEP Students," paper presented at the Council of Chief State School Officers 32nd Annual National Conference on Large-Scale Assessment, June 2002, Palm Desert, CA; Sireci, Li, and Scarpati, *The Effects of Test Accommodation on Test Performance*.
- ²⁹ M. J. Allen and W. M. Yen, *Introduction to Measurement Theory* (Monterey, CA: Brooks/Cole, 1979); R. L. Linn and N. E. Gronlund, *Measurement and Assessment in Teaching*, 7th ed. (Saddle River, NJ: Merrill, 1995); R. M. Thorndike, *Measurement and Evaluation in Psychology and Education* (Saddle River, NJ: Merrill, 2005).
- ³⁰ Thorndike, *Measurement and Evaluation in Psychology and Education*.
- ³¹ *Ibid.*
- ³² Linn and Gronlund, *Measurement and Assessment in Teaching*.
- ³³ Abedi, "Classification System for English Language Learners."
- ³⁴ J. Abedi, "The No Child Left Behind Act and English Language Learners: Assessment and Accountability Issues," *Educational Researcher* 33, no. 1 (2004): 4–14.
- ³⁵ S. J. Cech, "Graduation Hurdles Prove High for ELLs," *Education Week* 28, no. 17 (2009); P. A. Garcia and M. Gopel, "The Relationship to Achievement on the California High School Exit Exam for Language Minority Students," *Journal of Research and Practice* 1, no. 1 (2003): 126–40; J. Wang, D. Niemi, and H. Wang, *Predictive Validity of an English Language Arts Performance Assessment* (Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing, 2007).
- ³⁶ Abedi, "Language Issues in Item-Development"; Haladyna and Downing, "Construct-Irrelevant Variance in High-Stakes Testing"; S. Messick, "The Interplay of Evidence and Consequences in the Validation of Performance Assessments," *Educational Researcher* 23, no. 2 (2004): 13–23.
- ³⁷ Abedi, Hofstetter, and Lord, "Assessment Accommodations for English Language Learners"; Rivera et al., "Study 1"; Sireci, Li, and Scarpati, *The Effects of Test Accommodation on Test Performance*; Thorndike, *Measurement and Evaluation in Psychology and Education*.