

Assessment Accommodations for English Language Learners

Request for OMB Approval Supporting Statement A

December 24, 2007

Submitted to:

U.S. Department of Education
Institute of Education Sciences
555 New Jersey Ave., NW, Rm. 308
Washington, DC 20208
(202) 208-7078

Project Officer:

Rafael Valdivieso, Ph.D.
(202) 208-0662

Submitted by:

Regional Educational Laboratory West
(REL West at WestEd)
730 Harrison Street
San Francisco, CA 94107
(415) 615-3000

Project Director:

Stanley Rabinowitz, Ph.D.
(415) 615-3154

TABLE OF CONTENTS
Supporting Statement A: Study Justification

Introduction.....3

Overview: Study Scope and Sequence.....3

Research Questions.....4

A-1. Circumstances That Make Data Collection Necessary.....5

A-2. How, by Whom, and for What Purpose Information is to be Used.....6

A-3. Use of Information Technology.....10

A-4. Efforts to Identify and Avoid Duplication.....10

A-5. Sensitivity to Burden on Small Entities.....11

A-6. Consequences to Federal Program or Policies if Data are not Collected.....11

A-7. Special Circumstances.....11

A-8. Solicitation of Public Comment and Outside Consultation.....11

A-9. Payment or Gift to Participants.....13

A-10. Assurances of Confidentiality.....14

A-11. Sensitive Questions.....15

A-12. Estimates of Hourly Burden to Participants.....15

A-13. Estimate of Total Annual Cost Burden to Participants or Record-Keepers.....16

A-14. Estimates of Annualized Cost to the Federal Government.....16

A-15. Program Changes or Adjustments.....17

A-16. Tabulation, Analysis, and Publication of Results.....17

A-17. Displaying the Expiration Date for OMB Approval.....22

A-18. Exception to the Certification Statement.....22

References.....23

- Appendix A Student Language Background Survey
- Appendix B Parent Letter and Consent Form
- Appendix C Memorandum of Understanding to Districts
- Appendix D Letter of Introduction to Schools
- Appendix E Federal Register Notices
- Appendix F Pledge of Nondisclosure
- Appendix G Items Representative of Those Appearing on Final Test

SUPPORTING STATEMENT A:
STUDY JUSTIFICATION

Introduction

This document presents Supporting Statements A and B for a research study on Assessment Accommodations for English Language Learners (ELLs). Specifically, we are seeking OMB approval for four data collection activities related to this study (see Section A-2 for details on each):

- 1) Item Tryouts¹
- 2) Operational Test Administration²
- 3) Student Language Background Survey
- 4) Student Achievement Data from schools/districts

Overview: Study Scope and Sequence

This study will examine the effect of one test accommodation and its impact on the validity of assessments for ELLs. Specifically, we will investigate the ways in which linguistic modification affects students' ability to access math content during testing. Linguistic modification is a theory-based process in which the language in test items, directions, and/or response options are modified in ways that clarify and simplify the text without simplifying or significantly altering the construct tested (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005). To facilitate comprehension, linguistic modification reduces the construct-irrelevant language demands (e.g., semantic and syntactic complexity) of text through strategies such as reduced sentence length and complexity, use of common or familiar words, and use of concrete language (Abedi, Lord, & Plummer, 1997; Sireci, Li, & Scarpati, 2002).

Increased access, via linguistic modification, is believed to minimize the effects of construct-irrelevant language demands on ELLs. In this way, the accommodation facilitates ELLs' ability to demonstrate their content-related/construct-relevant knowledge and skills, without simplification of the content or significant alteration of the construct tested. By comparing the effects of linguistic modification on ELL's test performance with its effects on the performance of English language proficient general education students without disabilities (non-ELL/non-Students with Disabilities, or non-ELL/non-SDs), this study aims to increase understanding of the effects of a test accommodation that holds promise as a means of decreasing the achievement gap between ELL and non-ELL/non-SD students.

Because instrumentation is central to this study as a means for operationalizing and measuring the effects of linguistic modification on student access to test content, our initial step will focus on ensuring that the two instruments (one with linguistically modified items and one with original items) are sufficiently valid for the two large-scale data collection efforts that will follow: a) a pilot test of the modified and original items to provide additional support for the validity of the instruments; and b) an experimental study in which non-ELL/non-SD and ELL students (with both low and high reading

¹ Per instructions from IES, while items administered on achievement tests do not require OMB approval, items representative of those appearing on the final test are included in Appendix H.

² Per instructions from IES, while tests of math achievement developed and administered as part of this study do not require OMB approval, items representative of those appearing on the final test are included in Appendix H.

abilities) are randomly assigned to take either the original or modified versions of the test. Planned data analyses will systematically examine the relationship between linguistic modification and access to test content for two different student populations (ELL and non-ELL/non-SD³) as well as the effects of linguistic modification on test performance.

Research Questions

The following research questions guide this study:

RQ 1: Does the use of linguistically modified items differentially affect the technical adequacy (validity and reliability) of assessments of mathematics achievement for ELL students and non-ELL/non-SD students with both low and high reading abilities?

RQ 2: Is the difference between the mean scores of the original and modified tests for ELL students comparable to the difference between the mean scores of the original and modified tests for non-ELL/non-SD students (pooled, low and high reading abilities combined)? Is the difference between mean scores on the modified and original test greater for Non-ELL/Non-SD students who have high reading ability as compared with those who have low reading ability?

RQ 3: When comparing ELL and non-ELL/non-SD students of similar math achievement levels, do the probabilities of the students answering individual items correctly differ on the test with modified items as compared to the test with original items?

RQ 4: Are the underlying dimensions measured by the original and modified test items the same for the ELL and non-ELL/non-SD (pooled) student groups? For the ELL and non-ELL/non-SD student groups? Do the correlations (1) among latent factors (e.g., mathematics achievement, verbal ability) and (2) between latent factors and test items differ for the ELL and non-ELL/non-SD (pooled) student groups?

RQ 5: For the non-ELL/non-SD population, are the correlations with a standardized test of mathematics achievement comparable for the linguistically modified and original test forms?

³ We also will examine whether performance differences emerge between non-ELL/non-SD students with low reading abilities and non-ELL/non-SD students with high reading ability. If linguistic modification reduces the language burden of the test, as anticipated, the score difference across test forms (modified and original) will be greater for low-ability readers than for high ability readers. If a difference emerges across forms and it does not vary by reading ability of non-ELL/non-SD students, this suggests that the modification may have changed the mathematics content assessed as well as the language burden.

A-1. Circumstances That Make Data Collection Necessary

States currently are trying to determine if their assessment practices for ELL populations are consistent with the expectations of No Child Left Behind Act of 2001 (NCLB). Particularly problematic is the issue of *access*. Although appropriate access to test content is an issue for all students, it is a fundamental concern for ELLs because it affects the accuracy of measures of their academic performance and the validity and comparability of their test scores with those of their English language proficient counterparts.

In response to increasing concerns about fair access to test content, state Departments of Education have adopted a variety of policies regarding the testing of ELLs in grades K-12 that include provisions for certain accommodations (Bielinski, Sheinker, & Ysseldyke, 2003; Rabinowitz, Ananada, & Bell, 2004; Rivera & Collum, 2004; Thurlow & Bolt, 2001). Currently, an *accommodation*⁴ is defined as a change in testing conditions implemented to increase accessibility of test content to a specific student population. Such changes are deemed fair and reasonable when standardized administration conditions do not provide an equal opportunity for all students to demonstrate what they know and can do (Abedi & Lord, 2001; Butler & Stevens, 2001; Holmes & Duron, 2002; National Research Council, 2004), and it is assumed that with or without the accommodation, the same construct is being assessed. Theoretically, an accommodation should not affect the validity or reliability of test results for the non-ELL/non-SD population who have adequate access to the test content (Baker, 2001). Rather, an accommodation is intended to minimize or remove the effects on test performance of construct-irrelevant factors that may contribute to the under-representation of student achievement in the content area—in the case of ELLs, both theory and research suggest that one source of construct irrelevance is language.⁵

This study is necessary because so little empirical data are available about the effectiveness of test accommodations in providing students with appropriate access to tested content. Compounding this lack of scientifically-based research is the acknowledgement that some of the frequently-used accommodations may not be relevant for ELL students, unless the student has the type of disability for which the accommodation is appropriate (Tindal & Ketterlin-Geller, 2004). As a result, policies on allowable accommodations for ELLs remain inconsistent across states (Goh, 2004; National Research Council, 2004; Rivera & Collum, 2004; Thurlow, Wiley, & Bielinsky, 2002).

At a minimum, states are proceeding under the assumption that they must provide evidence that the assessments for their special student populations are comparable and *at least* as valid for these students as other tests within their comprehensive assessment program. However, we suggest that the evidence may support a finding that the accommodated assessments are sufficiently *more* valid for the special student

⁴ This study draws a distinction between an accommodation and a modification. For the purposes of this study, a modification is defined as an adjustment to the test itself, its conditions, or standards for assessment that increase the accessibility of the test content for a specific student population in a manner that may be fair and reasonable, but significantly alters the construct assessed. Examples of test modifications include allowing students with specific disabilities to use calculators on mathematics computation items (when the general education students cannot) or allowing the reading comprehension portions of a test to be read aloud to ELLs.

⁵ Though other construct-irrelevant sources for ELLs may include SES and cultural biases, this study will focus on language as a source of construct irrelevance for ELLs.

populations. Current definitions of validity related to the assessment of non-ELL/non-SD students may not be sufficient when applied to the assessment of ELLs. The impact of population characteristic on the constructs assessed need to be systematically considered.

For these reasons, there is a need for methodical and rigorous investigation of the effects of test accommodations on test validity and student performance. Such investigations should build upon research-based findings that include consideration of population-relevant characteristics (Abedi, 1999, 2001, 2004; Abedi & Lord, 2001; Abedi, Courtney, & Leon, 2003; Abedi, Hofstetter, & Lord, 2004; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005). Especially given the high-stakes nature of assessments administered to students with special needs under NCLB, an empirical basis is needed for ensuring that test content is maximally accessible to these students, that the test is equitable, and that results are valid and reliable for local, state, and federal accountability needs. Considerations of individual need and the technical qualities of assessment—the need to balance test equivalence and validity—are inherent components of decision-making in high-stakes testing of students with special needs. Findings from this study may advance current understanding of technically sound assessment practices by presenting empirical evidence about the ways in which increasing ELL access to tested content may yield *more* valid measures of what students know and can do.

To meet this need for scientifically-based research, we propose that data about the effects of linguistic modification be systematically collected as described in Table 1 below (see Section A-2 for details on each strategy).

Table 1. Timeline for Data Collection Activities*

	<i>Item Tryouts</i>	<i>District Recruitment</i>	<i>Student Samples Selected</i>	<i>Test and Language Background Survey Administered</i>	<i>Data Collected from School Records</i>
March 2007					
April					
May					
June					
July					
August					
September					
October		X			
November		X			
December	X		X		
January 2008				X	X
February				X	X

*Final timeline is dependent upon OMB approval.

A-2. How, by Whom, and for What Purpose Information is to be Used

These findings are intended to be used by test developers, test consumers, and the research community. To be trustworthy, and therefore useful, we must first demonstrate that increased access to test content through accommodation is not due to a significant change in, or simplification of, the *construct* being assessed. This need is the impetus for data collection activities (i.e., item tryouts, test administration, student language background survey, school records) that will focus on comparing the effects of the accommodation on validity of findings and on test performance in both the non-ELL/non-SD and ELL populations.

Purpose of Data Collection Activities

The assessment of students who are ELLs⁶ at the elementary and secondary levels is regulated by federal laws designed to protect the rights of students who are learning English.⁷ Legally, test use resulting in disparities in student performance based on limited language proficiency may be challenged as prejudicial. In all cases, the educational system bears responsibility for providing students who are ELLs with the opportunity to learn the content and skills being tested. When high-stakes consequences are associated with test results, as they are under NCLB, stakeholders need assurance that test content is accessible to all students and that the test results are valid.

The actual effectiveness of current practices for making high-stakes tests accessible, equitable, and valid for ELLs is unclear (Butler & Stevens, 2001; Castellon-Wellington, 2000; Holmes & Duron, 2002; Rivera & Stansfield, 2001). Although scores for ELLs tested with accommodations are expected to be comparable to those for English language proficient students tested without accommodations, many questions have emerged about the validity of such assumptions (Abedi, Hofstetter, & Lord, 2004; NRC, 2002, 2004; Rivera & Collum, 2004). The validity of interpretations drawn from standardized test scores is of primary concern to stakeholders seeking fair and accurate information about the achievement of students who are ELLs and the subsequent uses of these results for local, state, and federal accountability purposes.

Additionally, scores of accommodated and unaccommodated tests generally are compared across student groups; i.e., levels of performance are interpreted as having comparable meaning for non-ELLs assessed without accommodations and for ELLs assessed with accommodations. We suggest that the validity and reliability of such comparisons, however, needs to be examined empirically. Recommendations in the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) specify that results from accommodated tests should be interpreted cautiously; because some accommodations make the administration "nonstandard," interpretation of results is complex. In some situations, test scores from accommodated administrations may *not* have the same meaning as those derived under standardized conditions (Camara, 1998). In the absence of empirical evidence to support assumptions about the effects of accommodation on the psychometric qualities of the test, claims about the validity of test results for special student populations may be inappropriate (AERA et al., 1999; Goh, 2004).

Because of the lack of evidence about the actual effectiveness of accommodations, the appropriateness of inferences about student test performance may be compromised when assessment conditions are changed to increase access to ELLs (Hafner, 2001; Tindal & Fuchs, 2000). To address lingering questions about one test accommodation, this study will investigate the effectiveness of linguistic modification as a means for increasing access to test content for ELLs. Because access may pose a

⁶ This study follows federal guidelines (Office of Elementary and Secondary Education, 2000) in using the term English language learner (ELL) to define students who are "national origin minority students who cannot speak, read, write, or comprehend English well enough to participate meaningfully in and benefit from the schools' regular education program." No Child Left Behind legislations (including Title III) refers to this population as limited English proficient (LEP).

⁷ These laws currently include the equal protection and due process clauses in the Fourteenth Amendment, the Civil Rights Act of 1964, the Bilingual Education Act of 1968, Section 504 of the Rehabilitation Act of 1973, the Equal Educational Opportunity Acts of 1974 and 2000, Improving America's Schools Act of 1994, and Titles I, III, and VII of the No Child Left Behind Act.

potential threat to validity of inferences from test results, this study will investigate the degree to which research-based changes that increase access to tested content for ELLs yield a *more valid and reliable* measure of what ELLs know and can do so that their scores can be more meaningfully compared with those for non-ELL/non-SDs. These investigations will gather empirical evidence to support assessment practices that should increase access for ELLs and yield *more valid* measures of what these students know and can do.

Study Design and Data Collection Strategies

The full design of the study is a 2-by-2-by-3, fully crossed design. The factors are test form (original or modified), grade level (seventh or eighth), and student population (ELL, low reading ability non-ELL/non-SD, and high reading ability non-ELL/non-SD). For this study, grade level serves as a blocking factor. Additional details about sampling strategies and power analyses are described in Supporting Statement B-2.

Table 2 below provides a summary of the data collection activities for this study. The table is followed by a detailed description of each of the data collection activities (item tryouts, operational test administration, student language background surveys, and data collected from student records or school/district databases).

Table 2. Data Collection Summary Table

Data Collection Method	How Data are Collected	Who Collects Data	Intended Use of Data	Associated Research Questions
Item Tryouts	Modified and original items (approximately 25-30) that measure math achievement are administered to 100 middle-school students.	Researchers, with assistance from teachers and school staff and cooperation of participating students.	To refine instruments and verify effectiveness of modification strategies.	RQ 1
Operational Test Administration	One of two final versions of 30-item math achievement test (accommodated or non-accommodated) is administered to 3,600 middle-school students under experimental conditions.	Researchers, with assistance from teachers and school staff and cooperation from participating students.	To examine effects of linguistically modified items on ELLs and non-ELLs/non-SDs.	RQs 1, 2, 3, 4, 5
Student Language Background Survey	A8-item language background survey (Appendix A) available in English and Spanish is administered to all students before they take the math test (in test booklet).	Researchers, with cooperation of participating students.	To provide additional types of information about language background of student participants that may help us identify factors that affect test performance.	Contextual Data for all RQs
Student-level	Districts submit student	Researchers, with	To provide	RQs 1, 2, 4, 5

Data from School Records or District Database	archived data for all participating students.	assistance from district or school data management staff.	additional types of information about student participants that may help us identify/control factors that affect test performance.	
---	---	---	--	--

Item Tryouts (Pilot Test). Researchers administer approximately 25-30 linguistically modified and original items (selected from the initial pool of 40-50) that measure math achievement to a convenience sample of 50 middle-school ELLs and 50 middle-school non-ELL/non-SDs (pooled high and low reading ability). Performance data collected from these tryouts will be used to ensure that the items are accessible to and appropriate for the range of students included in this study. The item-level statistics produced will include p-values, standard deviations, and omission rates. The small samples for this pilot test are justifiable because 1) the released NAEP and state items already have undergone extensive statistical analysis; 2) we are seeking to minimize the testing burden by not collecting unnecessary information.

Students also are administered the language background survey (Appendix A). For both the item and survey administrations, students are reminded that participation is voluntary but appreciated and that they may refuse to answer any item or question. Parents/guardians of sampled students receive a letter about the study (Appendix B) and are asked to sign and return the attached active consent form.

Operational Test Administration. Approximately 50 schools recruited from 15 districts across one state (California) are asked to submit a list of eligible 7th and 8th grade students (see Section B-1 for details). From these lists, 1200 ELLs and 2400 non-ELL/non-SD students are selected and randomly assigned either to the treatment (test with linguistically modified items) or control group (original test with non-modified items).

Random assignment is done at the student level. Students are randomly assigned to a condition within each school, grade, and English proficiency group. Test forms are randomly ordered (by WestEd research team) several days prior to the testing day. After grouping participating students by grade and English learner status, test proctors (arranged by WestEd) randomly distribute test forms (accommodated or un-accommodated forms) to students in classrooms. The language background survey (Appendix A) is included with each test form. All procedures are conducted under the direct supervision of trained senior researchers.

Students are reminded that participation is voluntary but appreciated and that they may refuse to answer any item. Parents/guardians of sampled students receive a letter (Appendix B) available in English and Spanish that introduces the study and provides details about how, when, and why testing will occur. In that letter, parents are asked to sign and return the attached active consent form. To the extent possible, school staff will be informed about the study (see Appendix D) so they may encourage student participation.

Student Language Background Survey. Prior to test administration, all participating students are asked to complete a eight-item survey⁸ about their language background (Appendix A). The survey will be available both in English and Spanish. The questions included on the survey are intended to gather information about factors known to covary with test performance, such as languages spoken at home (Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Plummer, 1995). Students are reminded that they may refuse to answer any question. For the parent's review, a copy of the survey is included with the letter/consent form sent to parents of all eligible students (Appendix B). The letters explain that students will complete the survey when they are administered the math tests. Data collected from the language background surveys will be used to provide a context for the findings from the statistical analyses.

Student-level Data from Districts or Schools. Schools attended by students participating in the operational test administration are requested to provide pertinent information about these students from their permanent records or from a district database. Information requested includes student's demographic data (race/ethnicity), English language proficiency score and status (if ELL), and recent scores from standardized statewide tests of achievement in reading and mathematics. Student names will be encrypted following data entry and information linking student name to achievement data will be destroyed once test data and achievement data are matched.

A-3. Use of Information Technology

The data collection plan reflects sensitivity to issues of efficiency, accuracy, and burden. Where feasible, background information will be collected from existing data sources (e.g., school records), rather than by collecting primary data in ways that would impose additional burden on student participants. These data will be collected from participating school districts in the format most convenient for the preparers, through electronic posting to a shared secure site. Communications among the research team and school and district staff will occur through email, fax, and conference calls using technology that reduces the burden associated with paperwork and face-to-face meetings.

To maintain consistency with the standardized testing format with which students are most familiar, the math achievement tests (modified and original forms) will be administered in a group setting in a paper-based format. Trained WestEd test coordinators will administer the tests and student surveys and will explain directions and answer questions. Test items will be scored off-site using automated or Scantron technology. Once IES has approved the final report for the study, it will be made available on the world wide web for public viewing.

A-4. Efforts to Identify and Avoid Duplication

The research questions defined above call for a unique empirical study, one in which the effects of one accommodation on test validity can be compared across two populations (ELL and non-ELL/non-SD). While other researchers have examined the effects of accommodations on ELL test *performance*, a number of factors about our approach are unprecedented. First, this study incorporates an experimental design to

⁸ Adapted from Ferrara, Duncan, Freed, Velez-Paschke, McGivern, Mushlin, Mattessich, Rogers, and Westphalen, 2004

ensure that the impact of linguistic modification undergoes rigorous examination and evaluation with students randomly assigned to treatment and control conditions. Second, at the foundation of this study are two research-based, strategically designed instruments (modified and original) whose usefulness in systematically detecting performance differences related to access have been documented⁹. Third, because the actual effectiveness of current practices for making high-stakes assessments accessible remains unclear, this study will investigate the degree to which research-based changes increase access to tested content for ELLs, thereby yielding a more valid and reliable measure of what ELLs know and can do and more appropriate, meaningful comparisons with scores from non-ELL/non-SD students.

A-5. Sensitivity to Burden on Small Entities

To ensure that undue burden is not placed on the schools attended by participating students, we are soliciting only the minimum information required to meet study objectives. Research methods are intended to ensure that all data are accurate and appropriate but also efficiently collected with minimal intrusion on staff planning, administration, or instructional time. The 30-item math tests include only the number of items deemed necessary for valid inferences about student performances and the brief, eight-item student language background surveys seek only responses to critical data that are not available from district databases. To further minimize burden on school staff, test and survey administration activities may be facilitated by district and school staff, but will be conducted and coordinated by the WestEd research team.

A-6. Consequences to Federal Program or Policies if Data are not Collected

This study involves a one-time only collection of data. However, *not* collecting these data would restrict The U.S. Department of Education's ability to guide states' development of testing policies that prescribe the appropriate use of test accommodations. As noted in A-1 and A-2, NCLB requires state test administrators and policy makers to base educational decisions on scientifically-based research, and findings from this empirical study have the potential to directly influence the implementation of accommodation strategies for ELLs that may be effective in ensuring that this special population of students is tested accurately and equitably.

A-7. Special Circumstances

This study request fully complies with the regulations and seeks no consideration of special circumstances.

A-8. Solicitation of Public Comment and Outside Consultation

a. Federal Register Announcement

A notice about the study will be published in the *Federal Register* when the final OMB package is submitted. Information will be posted in a manner consistent with stated policy. A draft of the first notice is provided in Appendix E.

⁹ Sources of documentation will include expert judgment and findings from cognitive interviews and item tryouts.

b. Consultation with Experts Outside the Agency

Throughout the planning and design phases of this study, we sought technical advice from members of an technical work group (TWG) with whom we consult on a regular basis about methodological issues (experimental design, sampling frame, power estimates, data collection and analysis, and reporting strategies) and request feedback on associated products (protocols, instruments). The purpose of these consultations was to ensure that the study demonstrates appropriate technical rigor and relevance to the field so that findings will be trustworthy and useful. Educational survey experts contributed specific recommendations for improving the student language background survey so all items would be clear and unambiguous. Members of this TWG include:

- Professor Jamal Abedi, CRESST, University of California, Davis
- Dr. Lloyd Bond, Carnegie Foundation for the Advancement of Teaching
- Professor Geoffrey Borman, University of Wisconsin
- Professor Brian Flay, Oregon State University
- Professor Tom Good, University of Arizona
- Dr. Corinne Herlihy, Manpower Demonstration Research Corporation (MDRC)
- Dr. Joan Herman, CRESST, University of California, Los Angeles
- Professor Heather Hill, University of Michigan
- Dr. Roger Levine, American Institutes for Research (AIR)
- Professor Juliet Shaffer, University of California, Berkeley
- Dr. Jason Snipes, Council of the Great City Schools

We also solicited comments from current educators who work with general and special populations, from mathematics content experts, and from state-level test administrators. These experts include:

- Dr. Patrick Callahan, WestEd
- Dr. Stanley Rabinowitz, WestEd
- Dr. Charlene Rivera, Second Language Testing, Inc.
- Dr. Charles Stanfield, Second Language Testing, Inc.

Senior advisory staff with whom we will continue to consult through the course of the study are associated with the following nationally-known professional organizations and agencies:

- Assessment and Accountability Comprehensive Center (AACC)
- Council of Chief State School Officers (CCSSO)
- National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
- National Council of Teachers of Mathematics (NCTM)
- Regional Educational Laboratory-West (REL-W)

A-9. Payment or Gift to Participants For Students

According to the NCEE report on incentives in evaluation research¹⁰, incentives may be helpful in maintaining the integrity of the treatment (i.e., administered the

¹⁰ Council of Profession Associations on Federal Statistics, "Providing Incentives to Survey Respondents, Final Report," September 22, 1993.

modified test in this study) and control (i.e., administered the original test in this study) groups in experimental design studies. Incentives have been shown to be effective in improving response rates and in gaining participant cooperation (Brick, Hagedorn, Montaquila, Roth, and Chapman, 2006). When the incentive is awarded after completion of the test, it may boost completion rates (Lazear, 1997).

Additionally, because this study asks middle-school students to show what they know and can do on a math test, we believe that compensation is justified. Findings from studies of the effectiveness of incentives for 8th graders participating in NAEP math testing support this decision (O'Neil, Sugrue, & Baker, 1996), suggesting that incentives may increase student effort during testing at grade 8 (O'Neil, Abedi, Lee, Miyoshi, & Mastergeorge, 2001; O'Neil, Sugrue, & Baker, 1996). The usefulness of the non-monetary compensation both for ELLs and non-ELL/non-SDs will be of interest to researchers and will be included in final report.

District/Schools

Compensation to districts and schools is justified as the demands for districts' or schools' cooperation in research at the federal and state levels and from institutions of higher education have become burdensome. We recognize that by adding an additional layer of assessments, even though the timing will not to interfere with state testing, we have increased burden on the school community. The amounts of remuneration are based on the level of burden described in the NCEE report.

To compensate for the demands and burden on students and schools, we propose the compensation structure schedule presented in Table 3.

Table 3. Schedule of Proposed Compensation

Data Collection Method	Proposed Compensation	Rationale for Compensation (Burden or Relationship to Data Quality)	Timeline for Awarding Compensation
Item Tryouts	Students: pen	Non-monetary award to promote data quality, encourage participation, and motivate pre-teens to put forth best effort	Following testing
	Schools: monetary award of \$100 ¹¹	To compensate for moderately low burden incurred in supporting administration of tests to 100 students	Following testing
Operational Test Administration	Students: pen	Non-monetary award to promote data quality, encourage participation, and motivate pre-teens to put forth best effort	Following completion of survey and testing
	Schools: monetary award of \$350 ¹²	To compensate for moderately high burden incurred in supporting administration of tests to 3,600 students and in providing student records	Following testing and release of school records
Student Language	No additional compensation	Compensation linked to Operational Test	Following completion

¹¹ Two test administrations per school, and the time needed to assist with each session is estimated to be about one hour with \$50 hourly rate (also see Table 4).

¹² Consistent with Item Tryouts, two test administrations per school, and the time needed to assist with each session is estimated to be about one hour with \$50 hourly rate (so total \$100 per school). In addition, it is estimated that about 4 hours are required for school staff to assist with student archived data collection with \$65 hourly rate (so total \$260 per school). This is how the incentive of \$350 per school was determined.

Background Survey	provided to students	Administration, above	of survey and testing
School or District Student Records	No additional compensation provided to schools	Compensation linked to Operational Test Administration, above	Following testing and release of school records

A-10. Assurances of Confidentiality

WestEd will implement long-standing procedures with proven effectiveness to protect the confidentiality of data and the rights to privacy of students and schools. These include separating identifying information from test and survey results as soon as data are filed, secure storage of paper-based data and computer files, and restricted access to data to those who have direct responsibility for sampling and data collection activities. All study procedures and protocols were designed to comply with the Department of Education's Institutional Review Board (IRB) regulations for safe and appropriate research with minimal burden to human subjects and in keeping with the principles of ethical research outlined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999).¹³ Senior members of the study team have been certified by WestEd's IRB as having received training in the importance of confidentiality and data security and all study staff have participated in the security clearance processes required for federal grant recipients.

To ensure the confidentiality of participants, WestEd obtains signed Pledges of Nondisclosure (Appendix F) from all employees, subcontractors, and consultants that may have access to these data. Once test data have been assigned a unique identification number, researchers will have access only to assigned tracking numbers for each student. No names will appear with student responses. All school- and student-level identifiable information will be kept in secure locations and unique identifiers will be removed as soon as alternate, non-traceable identifiers can be assigned. Information from participating students will be presented at the subgroup level (ELLs and non-ELL/non-SDs). Student-level data may be linked to schools but no information that identifies a student will be released and no individually identifiable information will be maintained by the study team.

Study materials will be stored in secure locations (dedicated data server) and access to data files and hard copies restricted to authorized users. WestEd will produce carefully documented archival data files for safe storage of all student-level data. Using IRB recommendations, a separate, edited version of the files will be produced, with individual and site identifiers (including small cells) removed.

All researchers and staff working on this study will comply with the Privacy Act of 1974 (P.L. 93-579, 5 USC 552a) for all individual and institutional data collected. In addition, WestEd follows the confidentiality and data protection requirements of IES (The Education Sciences Reform Act of 2002, Title I, Part E, Section 183)¹⁴. We will

¹³ The final research proposal will be submitted to the IRB for a full review in June 2007.

¹⁴ The Education Sciences Reform Act of 2002, Title I, Part E, Section 183 requires "All collection, maintenance, use, and wide dissemination of data by the Institute" to "conform with the requirements of section 552 of title 5, United States Code, the confidentiality standards of subsection (c) of this section, and sections 444 and 445 of the General Education Provision Act (20 U.S.C. 1232g, 1232h)." These citations refer to the Privacy Act, the Family Educational Rights and Privacy Act, and the Protection of Pupil Rights Amendment. In addition, for student information, "The Director shall ensure that all individually identifiable information about students, their academic achievements, their families, and information with respect to individual schools, shall remain confidential in accordance with section 552a of title 5, United States Code, the confidentiality standards of subsection (c) of this section, and sections 444 and 445

protect the confidentiality of all information collected for the study and will use it for research purposes only. No information that identifies any study participant will be released. Information from participating institutions and respondents will be presented at aggregate levels in reports. Information on respondents will be linked to their institution but not to any individually identifiable information. No individually identifiable information will be maintained by the study team. All institution-level identifiable information will be kept in secured locations and identifiers will be destroyed as soon as they are no longer required.

Finally, the following verbatim language regarding confidentiality will appear on the letters, the student language background survey, and test materials:

Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

A-11. Sensitive Questions

We are requesting OMB approval to collect math achievement data and language background information for two groups of students: ELLs and non-ELL/non-SDs. We will gather information directly from students, through the math assessments (to examine the ways in which access to test content is affected by linguistic modification) and the language background survey (see Appendix A) for the eight questions included on the surveys). No questions of a highly sensitive nature are included on the student survey, and all will be pre-tested through the cognitive interviews. Schools, parents, and students will be assured that these data will be used statistically to evaluate the effects of test accommodations, not to judge individual students' level of math achievement.

A-12. Estimates of Hourly Burden to Participants

Estimates of annual number of responses and hour/cost burden for each data collection activity are provided in Table 4. Note that each data collection activity occurs only once throughout the study.

Table 4. Estimates of Response Burden, by Data Collection Activity

A	B	C	D	E	F	G	H
---	---	---	---	---	---	---	---

of the General Education Provision Act. Subsection (d) of section 183 prohibits disclosure of individually identifiable information and makes it a felony for staff to publish or communicate individually identifiable information.

Task	Time Burden Per Form or Session (in minutes)	Annual Number of Respondents	Frequency of Data Collection	Annual Number of Responses ¹⁵	Annual Hour Burden ¹⁶	Estimated Burden Hourly Rate	Annual Cost Burden
Gaining Cooperation (from school principals) for Item Tryouts	60	2	1	2	2	\$36	\$72
Sampling/Gaining Cooperation (from school principals)	60	50	1	50	50	\$36	\$1,800
Assist with Test Administration (by school staff) for Item Tryouts	60	2 ¹⁷	1	4 ¹⁸	4	\$50	\$200
Assist with Test Administration (by school staff) for Operational Test Administration	60	50	1	100	100	\$50	\$5,000
Student Data Collection:							
Student Language Background Survey	5	3,600	1	3,600	300	\$10	\$3,000
Student Archived Data Collection (from schools) ¹⁹	240	50	1	50	200	\$65	\$13,000
GRAND TOTAL		3,754		3,806	656		\$23,072

A-13. Estimate of Total Annual Cost Burden to Participants or Record-Keepers

There are no direct costs to participants other than their time to participate in the study, as estimated above.

A-14. Estimates of Annualized Cost to the Federal Government

Total budget for the study is \$791,031. The approximate budget for each year is as follows:

January 18, 2006 – January 17, 2007: \$146,119

January 18, 2007 – January 17, 2008: \$423,343

January 18, 2008 – January 17, 2009: \$221,569

The average annual cost per year (for 3 years) is \$263,677.

A-15. Program Changes or Adjustments

The change of total 656 annual burden hours reflects new data collection.

¹⁵ E=C*D.

¹⁶ F=(B*E)/60.

¹⁷ Two schools will be selected to participate in Item Tryouts.

¹⁸ Two classrooms per school.

¹⁹ For the operational test administration only.

A-16. Tabulation, Analysis, and Publication of Results

Table 5 below summarizes the planned data analysis and reporting timeline for this study, pending OMB approval. Each analytic strategy is described in greater detail below the table.

Table 5. Timeline Estimates for Tabulation, Analysis, and Publication of Results

	<i>Student Tests and Surveys</i>	<i>School Records</i>	<i>Item-Level Analyses</i>	<i>Correlations</i>	<i>ANOVA</i>	<i>Factor Analyses and DIF</i>	<i>Drafts & Final Report</i>
May, 2007							
June							
July							
August							
September							
October							
November							
December							
January 2008							
February	X						
March	X	X	X	X			
April			X	X	X	X	
May					X	X	
June							X
July							X

As shown in Table 5, data analyses are scheduled to begin in March 2008 following the February operational administration of the tests under experimental conditions.

Item-level Descriptive Analyses. Item-level statistics will be generated from item tryouts and the operational test administration. These statistics include frequency distributions for item choices, p-values, standard deviations, point bi-serial correlations, and omission rates. Estimates of internal consistency for the original and modified forms will be computed for each population. Performance data collected from item tryouts will be used to ensure that the items are accessible to and appropriate for the range of students included in this study. These analyses are intended primarily to help answer *Research Question 1*.

Analysis of Variance. Test scores from the operational administration, disaggregated by group (ELLs or non-ELL/non-SD student group pooled) and test version (linguistically modified or original), will be summarized to provide information about how each group performs on each test version. A three-factor Analysis of Variance (ANOVA) will test mean differences in test scores for the two student populations in order to examine if ELL students are better able to demonstrate their math ability on the modified form. The three factors are Test Form (original vs. modified), Student Population (ELL vs. Non-ELL), and Grade, a blocking factor (seventh vs. eighth).

If modification provides ELL students greater access to the math content, then the score difference between original and modified forms should be greater for the ELL population than for the Non-ELL population. The interaction between student population (SP) and test form (TF) is of particular interest in this ANOVA, because it addresses this hypothesis. For example, Figure 1 shows a possible finding that would suggest an

interaction effect between SP and TF. Scores from the linguistically modified test are higher than scores from the original test for both groups, but the difference between tests is greater for the population of students who are ELLs.

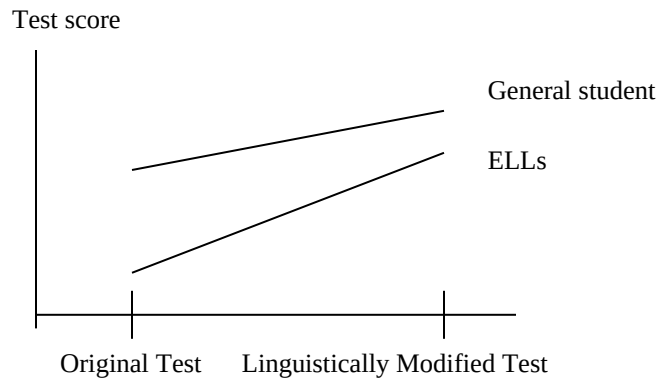


Figure 1. Hypothetical Interaction Effect

As mentioned earlier, we will also examine whether there are score differences between non-ELL/non-SDs with low reading ability and non-ELL/non-SDs with high reading ability. The expectation is that if linguistic modification has reduced the language burden of the test, the score difference in test forms will be greater for the low-ability readers than for the high-ability readers. If there is a performance difference between forms, and it does not vary by reading ability group for non-ELL students, this may be an indication that the modification has changed the mathematics as well as the language burden. These analyses are intended to help answer *Research Question 2*.

Differential Item Functioning Analysis. Using test scores from the operational administration, an analysis of differential item functioning (DIF) will be conducted to address whether the chance of a student scoring a correct response on an original item is greater in the non-ELL/non-SD population than in the EL population, even after controlling for total test score. In general, an item exhibiting DIF may indicate the multidimensionality of the item. That is, there could be another construct, other than the target achievement construct assessed by the set of items in the analysis, which is associated with group membership and is contributing to performance on the item.

DIF analysis using the Mantel-Haenszel (MH) procedure (Holland & Thayer, 1988) will be used to examine item performance between the two student groups (ELLs and the non-ELL/non-SD student population) for both the set of original items and the set of linguistically modified items. This procedure is a non-iterative contingency table method that will allow us to detect if the odds of passing the item is equal for students from each group, after controlling for student ability (as estimated by test score). Those items showing DIF will be examined closely along with any information about the item obtained from the cognitive analyses and in the confirmatory factor analysis, described below, to identify possible reasons for the differential item functioning. These analyses are intended to help answer *Research Question 3*.

Factor Structure of the Tests. For each operational test form within each student population, exploratory factor analysis, using Principal Factor Analysis (PFA), will be conducted to estimate the number of constructs assessed by the test form and the underlying measurement structure (correlations) of the unobservable (latent) factor(s). The results from the PFA will serve as the foundation for a series of nested confirmatory factor analyses (CFA).

These analyses will be performed to test for differences in measurement structure across student groups (ELLs and non-ELL/non-SDs) and test type (modified or original). The factor analyses are intended to help define the effects of linguistic modification, and hence degree of access to test content, on the dimensionality of the test.

For each test, we will examine the correlation of item parcels with latent factors as well as correlations between latent factors (defined through the EFA and analysis of item content) for ELLs and the non-ELL/non-SD student groups. It is anticipated that (1) the item loadings for the non-ELL/non-SD student group on both versions will be higher than for students who are ELLs; (2) the correlations between latent factors will be higher for the non-ELL/non-SD student group than for students who are ELLs; and (3) the gap between these two groups will narrow on the linguistically modified test. These analyses are intended to help answer *Research Question 4*.

In order to further explore the factors accounting for these differences, another non-content-based (or construct-irrelevant) latent factor will be incorporated in the model. This latent factor, which may be labeled as "student verbal ability," may also affect students' performance on a math test, especially for ELLs (Abedi, Leon, & Mirocha, 2003). If this hypothesis is supported, we may expect the item parcel correlations with the linguistic latent factor to be higher for ELLs than for non-ELL/non-SDs, regardless of test version. However, we would expect these differences to be less pronounced on the modified test.

As an example, Figure 2 depicts the proposed structural model if there is an underlying general math ability (represented by F3) associated with two math content-based constructs (represented by F1 and F2) that each content-based construct has three observable items (represented by V1-V3 and V4-V6, respectively):

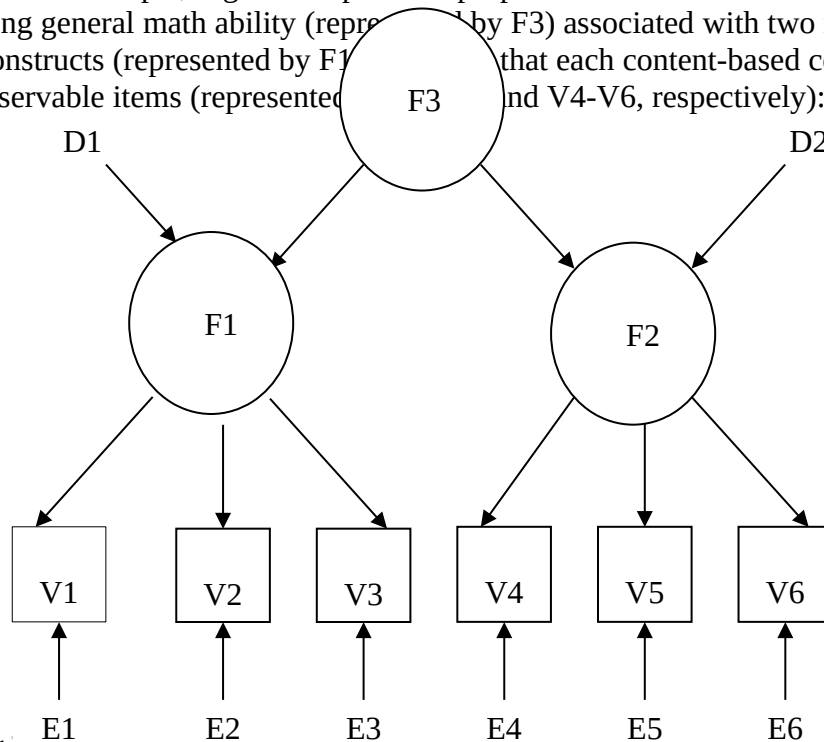


Figure 2. Structure Model of the Influence of Math Ability on Test Items

Examination of Test Correlations. We will examine the relationships of scores from the state's standardized test of mathematics achievement to both the original and linguistically modified test scores for the non-ELL/non-SD population of students. Specifically, this measure of criterion validity will be the state's standardized test that provides the basis for school accountability. We have hypothesized that linguistic modification should not alter the construct assessed, as demonstrated by the strong correlation of the study test to a standardized test of mathematics achievement. If the standardized test scores are provided on a continuous measure, then simple linear regressions will be examined to determine if the relationship with standardized test scores varies by form (linguistically modified or original) of the study test. If the standardized test provides scores on a categorical variable (e.g., proficient or not proficient) then logistic regression will be used. These analyses are intended to help answer *Research Question 5*.

Analyses of Student Language Background Survey Data. Planned analyses of student responses to the questions on the survey will provide information about the ways in which student language background characteristics influence performance on the two versions of the operational test (modified and original). Planned analyses include examining and summarizing student responses to provide rich description about the ways in which students access original and linguistically modified items. Analysts will use these descriptive data and response frequencies to provide a context for the study's quantitative findings. These analyses are intended to help answer the research questions by providing a context for the statistical analyses.

Analyses of Archived Student Data. Planned analyses of the data collected from school/district records will provide information about the relationship between 1) performance on statewide tests of achievement and English proficiency level (for ELLs); and 2) between performance on statewide tests of achievement (ELA and math) and performance on the two versions of the operational test (modified and original). Data collected on variables such as gender and race/ethnicity will be used as potential covariates in DIF analyses. These analyses are intended to provide context for the DIF analyses (*Research Question 3*) and to help answer *Research Question 5*.

Summary of Analyses. Cognitive interview data will be used to verify the fidelity with which test items assess the nature and degree of student access to tested content on tests of achievement. A synthesis of data from the item level analyses, ANOVA, factor analyses, DIF analyses, and correlational studies will be used to assess the impact of linguistic modification on 1) test performance for ELLs and non-ELL/non-SDs and on 2) the psychometric properties of the tests. Findings from school records and the language background surveys will be used to characterize student subgroups and set a context for the interpretation of the quantitative findings. The sum of data will be used to create a statistical model that describes the relationship among student characteristics, subgroup (ELL and non-ELL/non-SD) characteristics, and test performance. Table 6 below summarizes the links between research questions, data collection strategies, and planned analyses.

Table 6. Summary of Research Questions, Data Collected, and Analyses

Research Question	Data Collection Method	Planned Analyses
RQ 1: Does the use of linguistically modified items differentially affect the technical adequacy of assessments of mathematics achievement for ELL students and non-ELL/non-SD students with both low and high reading abilities?	Item tryouts Operational test administration Student-level data from school	Item-level analyses
RQ 2: Is the difference between the mean scores of the original and modified tests for ELL students comparable to the difference between the mean scores of the original and modified tests for non-ELL/non-SD students? Is the difference between mean scores on the modified and original test greater for non-ELL/non-SD students who have high reading ability as compared with those who have low reading	Operational test administration Student-level data from school	ANOVA

ability?		
RQ 3: When comparing ELL and non-ELL/non-SD students of similar achievement levels, do the probabilities of the students answering individual items correctly differ on the test with modified items as compared to the test with original items?	Operational test administration	DIF analyses
RQ 4: Are the underlying dimensions measured by the original and modified test items the same for the ELL and non-ELL/non-SD student groups? For the ELL and non-ELL/non-SD student groups? Do the correlations (1) among latent factors (e.g., mathematics achievement, verbal ability) and (2) between latent factors and test items differ for the ELL and non-ELL/non-SD (pooled) student groups?	Operational test administration Student-level data from school	Factor analyses
RQ 5: For the non-ELL/non-SD population, are the correlations between each test form and a standardized test of mathematics achievement comparable?	Operational test administration Student-level data from school	Correlations

Reporting of Results

Results of this study will be included in two reports: a Technical Report and a Final Report of Findings. A key objective of the Technical Report will be to provide a detailed description of the data collection and analytic methods that serve as the foundation of the study. A draft technical report will be submitted at the end of April 2008 with the final technical report due 30 days after receiving IES feedback. The Final Report will present and discuss findings, implications for the field, and study evaluation outcomes. The draft final report will be submitted in early June 2008 with the final report due following receipt of IES feedback. Following IES approval, both reports will be published through the REL network and/or made available to the Regional Comprehensive Centers. In addition, we anticipate making contributions to peer-reviewed journals and presentations at professional conferences.

A-17. Displaying the Expiration Date for OMB Approval

No waiver from displaying the expiration date is requested.

A-18. Exception to the Certification Statement

No exceptions related to the Certification Statement are requested.

References

- Abedi, J. (1999, April). *Examining the effectiveness of accommodation on math performance of English Language Learners*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada.
- Abedi, J. (2001). *Assessment and accommodations for English Language Learners: Issues and recommendations*. CRESST Policy Brief 4. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- 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. & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219-234.
- Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English Language Learners in NAEP*. 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*. Los Angeles: University of California, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C., & 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., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data*. CSE Tech. Rep. No. 603. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16-26.
- Abedi, J., Lord, C., & Plummer, J. (1995). *Language background as a variable in NAEP mathematics performance: NAEP task 3D: Language background study*. CSE Technical Report 429. Los Angeles: University of California, Center for the Study

of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Baker, C. (2001). *Foundations of Bilingual Education and Bilingualism, 3rd edition*. Philadelphia, PA: Multilingual Matters Ltd.
- Bielinski, J., Sheinker, A., Ysseldyke, J. (2003, April). *Varied opinions on how to report accommodated test scores*. NCEO Synthesis Report 49. Minneapolis: National Center on Educational Outcomes.
- Butler, F.A. & Stevens, R. (2001). Standardized assessment of the content knowledge of English Language Learners K-12: Current trends and old dilemmas. *Language Testing 2001*, 18(4), 409-427.
- Camara, W.F. (1998). *Effects of extended time on score growth for students with learning disabilities*. New York: College Board.
- Castellon-Wellington, M. (2000). *The impact of preference for accommodations: The performance of ELLs on large-scale academic achievement tests*. CRESST Technical Report 524. Los Angeles: University of California, National Center for the Study of Evaluation, Standards, and Student Testing.
- Goh, D.S. (2004). *Assessment accommodations for diverse learners*. Boston: Pearson.
- Hafner, A.L. (2001). *Evaluating the impact of test accommodations on test scores of LEP students and non-LEP students*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA, April 10-14, 2001.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (chap. 9). Mahwah, NJ: Erlbaum.
- Holmes, D. & Duron, S. (2000). *LEP students and high stakes assessment*. Washington, DC: National Clearinghouse for Bilingual Education, US Department of Education.
- Kenney, P. A. (2000). Families of items in the NAEP mathematics assessment. In N. S. Raju, J. W. Pellegrino, M. W. Bertenthal, K. J. Mitchell, & L. R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP* (pp. 5-42). Washington, DC: National Academy Press.
- National Research Council. (2002). In J. Koenig, (Eds.), *Reporting test results for*

- students with disabilities and English-language learners*. Washington, DC: National Academies.
- National Research Council. (2004). In J.Koenig & L. Bachman, (Eds.), *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Washington, DC: National Academies.
- O'Neil, H. F., Sugrue, B., & Baker, E. L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, 13, 135-157.
- Paulsen, C. A. & Levine, R. (1999). *The applicability of the cognitive laboratory method to the development of achievement test items*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Rabinowitz, S., Ananada, S., & Bell, A. (2004). *Strategies to access the core academic knowledge of English Language Learners*. San Francisco: WestEd.
- Rivera, C. & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English Language Learners*. Issue paper commissioned for the National Assessment Governing Board Conference on Increasing the Participation of SD and LEP Students in NAEP. Arlington, VA: George Washington University.
- Rivera, C. & Stansfield, C.W. (2001). *The effects of linguistic simplification of science test items on performance of Limited English Proficient and monolingual English-speaking students*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Thurlow, M. & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy*. NCEO Synthesis Report 41. Minneapolis: University of Minnesota, National Center on Outcomes.
- Thurlow, M.L., Wiley, H.I, & Bielinski, J. (2002). *Biennial Performance Reports: 2000-2001 State Assessment Data*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Tindal, G. & Fuchs, L. (2000). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center.
- Tindal, G. & Ketterlin-Geller, L. (2004). *Research on mathematics test accommodations relevant to NAEP testing*. Commissioned paper presented at the NAGB Conference on Increasing the Participation of Students with Disabilities and Limited English Proficient Students in NAEP.
- van Someren, M. W. (1994). *The think aloud method: A practical guide to modeling*

cognitive processes. San Diego, CA: Academic Press.