

Assessment Accommodations for English Language Learners

Request for OMB Approval Supporting Statement B

August 8, 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 B: Data Collection Procedures and Statistical Methods

Introduction..... 3

Overview: Study Scope and Sequence 3

Research Questions..... 4

B-1. Participant Universe and Sampling Procedures..... 4

B-2. Statistical Methods for Sample Selection and Degree of Accuracy Needed..... 7

B-3. Methods to Maximize Participation Rates..... 8

B-4. Tests of Procedures and Methods 9

B-5. Individuals Consulted on Statistical Aspects of the Study Design..... 10

References..... 12

Appendix A Student Language Background Survey

Appendix B Parent Letter and Consent Form

Appendix C Memorandum of Understanding to Districts

Appendix G Items Representative of Those Appearing on Final Test

Appendix H Cognitive Interview Rubric

Note: Appendices D, E, and F are not listed as they are associated with questions in Supporting Statement A only.

SUPPORTING STATEMENT B:
DATA COLLECTION PROCEDURES AND STATISTICAL METHODS

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) Pilot Test
- 2) Operational Test Administration
- 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).

Linguistic modification is believed to minimize the effects of construct-irrelevant language demands on ELLs, 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 abilities¹) are randomly assigned to take either the original or modified versions of the

¹ 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

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-SD) 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?

B-1. Participant Universe and Sampling Procedures

Table 1 below summarizes the estimated counts for each target population (districts, schools, and students) for the pilot test and the test administration. As the student Language Background Survey is embedded in the math test, the test and survey are administered concurrently. For each student tested, achievement data will be collected from the district or school prior to or immediately following testing.

Table 1. Summary of Estimated Counts for Each Target Population

	Target Number of Districts	Target Number of Schools	Target Number of 7 th and 8 th Grade Students
--	----------------------------	--------------------------	---

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.

			<i>ELL</i>	<i>non-ELL/non-SD</i>
A. Pilot Test	3	6	50	50
B. Test Administration, Language Background Survey, and Achievement Records	15	50 (based on recruiting goals of 75-90 students per school)	1200	2400 (1200 low-ability readers + 1200 high ability readers)

A. Pilot Test

District Recruitment & School Selection

The purpose of this pilot administration is to collect information about each prospective item so WestEd analysts can 1) review performance data in selecting the final group of math items that will appear on the operational test; and 2) ensure that each language background question is clear and appropriate for the intended population and provides support information about students' language background that may be used by researchers during data analyses to provide a context for the data collected during the test administrations. Using existing relationships with districts in California who have expressed interest in participating in research studies examining the effectiveness of test accommodations, we will identify *three* districts in California that enroll high percentages of middle-school ELL students whose native language is Spanish². With written approval from district superintendents, principals from six middle-schools in these districts will be asked to support the study by communicating information about the study to their math teachers and ELL support staff. These staff will serve as the primary contacts for the identification of a convenience sample of 100 students (half ELL and half non-ELL/non-SD) to whom a set of math items (both modified and original) and the language background survey will be administered. School participation will be solicited until the target sample of students is obtained.

Student Sample

The convenience sample of 100 ELL and non-ELL/non-SD students will be volunteers recruited by each school's math teachers and ELL support staff. The ELL sample will be comprised of students whose first language is Spanish and who have moderate to high levels of English Language Proficiency. The non-ELL/non-SD sample will be comprised of general education students who are not English Language Learners.

B. Administration of Test and Language Background Survey and Collection of Achievement Data from Records

District Recruitment & School Selection

Using demographic data from the California State Department of Education (CDE), we will identify up to 15 districts in California that have high percentages of middle-school, ELL students whose native language is Spanish (per district records). As the recruitment goal is to minimize the number of districts in the study to ensure that student sampling is conducted efficiently (i.e., fewest number of schools possible to reach target numbers of students) and to limit the number of different sources of student-level administrative data that will need to be accessed, a probability sample of districts is not

² This is a convenience sample used to provide additional information about item characteristics prior to the randomized trial; it is not a probability sample and therefore is not intended to represent that population.

feasible. Each district superintendent will be asked to sign a Memorandum of Understanding (Appendix C) signifying approval to contact school principals. Following district-level granting of approval, we will review demographic data from CDE and identify a maximum of 50 target schools that show enrollments of large numbers (i.e., approximately 10% to 25%) of seventh and eighth grade students in the target ELL population. For maximum sampling efficiency, schools from each district with the largest populations of enrolled ELLs (20-25% of total school population) will be recruited for study participation first, followed by second tier schools (ELLs are 15-20% of total school population) and then third tier schools (ELLs are 10-15% of total school population), until the target sample of students is obtained.

Principals from target schools will be sent Letters of Introduction (Appendix D) that explain the study purpose and participation guidelines. Once target schools have agreed to participate, we will conduct pre-test interviews at each school to review study guidelines and ensure accessibility of individualized achievement data³.

Student Sample

We will focus our recruiting efforts on securing the maximum number of students (approximately 70-90 students) from the minimum number of schools. To identify students for the study, we will ask participating schools to submit rosters of all students in grades 7 and 8 that meet either of the following criteria for inclusion in the study: (1) ELLs who are required to take the state's standardized assessment and who are not receiving special education services; and (2) non-ELL/non-SD students who have been identified by the district as meeting eligibility criteria (time in U.S. schools, language proficiency level) for the state assessment. We exclude students in special education from this phase of the research in order to control for effects of learning and other disabilities on test performance. We limit the study to one state so that it is possible to examine how the original and modified tests relate to student scores on one standardized achievement test. Since different states use different tests, it is important to examine this relationship across the study samples in only one state.

The initial pool of 7th and 8th grade ELL students will be narrowed to include only students whose first language is Spanish and who have moderate to high levels of English Language Proficiency⁴. We limit the population to students who have sufficient levels of proficiency in the English language to benefit from linguistic modification of test items, as students who cannot yet read English are less likely to benefit from this accommodation. From this second pool of ELLs, 1200 ELLs will be randomly selected to participate in the study. Approximately half will be seventh grade students and half will be eighth grade students.

The initial non-ELL/non-SD sample will be comprised of general education students who are not English Language Learners (non-ELL/non-SD) enrolled in grades 7 and 8. This pool will be divided into two groups, those with high reading ability and

³ These data include 1) reading and math scores from the California Standards Test, or CST, for all students, and 2) a current language proficiency score from California English Language Development Test, or CELDT, for ELLs.

⁴ Some linguistic modifications may vary depending on the native language of the student. Thus by studying only native speakers of Spanish, the study controls for this source of variability. Spanish was selected as the language for the study because 75% of students in the Western region (California, Nevada, Utah, & Arizona) who are ELL identify Spanish as their primary or secondary language.

those with low reading ability, based on state achievement test scores in reading. From this second pool, 1200 students with low reading ability and 1200 students with high reading ability will be randomly selected. Approximately half in each group will be seventh grade students and half will be eighth grade students.

B-2. Statistical Methods for Sample Selection and Degree of Accuracy Needed

Sample Size and Statistical Power

A total of 3600 students will participate in the study. This sample includes approximately 1200 ELLs and 2400 non-ELL/non-SDs. Hence, each test form (original or modified) will be administered to 1800 student-participants (600 ELLs and 1200 non-ELL/non-SDs). Collapsing across grade levels, the six test form-by-ELL group cells will have 600 students each, distributed evenly across the cells of the full 3-by-2 design.

Table 1. Full Study Design

	Original Test	Modified Test
ELL (half 7 th graders, half 8 th graders)	600 students	600 students
Non-ELL low reading ability (half 7 th graders, half 8 th graders)	600 students	600 students
Non-ELL high-reading ability (half 7 th graders, half 8 th graders)	600 students	600 students

These sample sizes were determined through analyses of the power of the different statistical tests that will be used in this study, which include ANOVAs and CFAs (Bloom, 1995). This design will provide a minimum detectable effect size (MDES)⁵ of .20⁶ for the main research question addressed by the ANOVA, which asks if the score difference between the original and modified tests differs for the ELL and non-ELL/non-SD groups.

In the 3-by-2 design above, the main research question is addressed by examining the difference between the mean scores of the original and modified tests for the ELL group and contrasting that difference with the difference of original and modified tests for the non-ELL/non-SD groups pooled. It is represented by a contrast in which each ELL group has a coefficient of +1 or - 1, and each non-ELL/non-SD group has a coefficient of +1/2 or - 1/2⁷.

The required sample size for power of .80 and MDES of .20 was estimated using the following equation:

$$n = (z_{1-\alpha/2} + z_{1-\beta})^2 (\sum c_i^2) \sigma^2 / (\sum c_i M_i)^2$$

⁵ MDES represents the level of effect that is reasonable and defensible for answering the study's research questions.

⁶ Assuming that power = .8 and alpha = .05.

⁷ Statewide, in California, the proportion of non-ELL/non-SD students who meet criteria to be classified as high-ability readers is 52% for seventh grade and 51% for eighth grade. (<http://star.cde.ca.gov/star2006/>) Thus, we feel equal weighting of the two non-ELL groups in these power calculations is a reasonable assumption.

where M_i is the mean for cell i , c_i is the contrast coefficient for cell i , and σ^2 is the common population (within-cell) variance (assumed to be 1 for this computation).

$$n = 7.85 (3)(1/.20)^2$$

$$n = 589.$$

Of secondary interest is the comparison of the two non-ELL/non-SD groups. This contrast is a “difference of differences” for the two non-ELL/non-SD groups. The sample size of 600 yields a MDES of .23 (power = .8 and alpha= .05). This test is represented by a linear contrast in which the coefficients for the ELL groups are zero, and the coefficients for the four non-ELL/non-SD groups are +1 or – 1. The MDES is found from [1] (assuming $\sigma^2=1$) :

$$600 = 7.85(4)(1/x^2)$$

$$x = .23$$

We believe this design provides a reasonable balance of power and costs. It provides for a MDES of .20 for the primary research question and power of .80 to detect a .23 effect for the question of secondary interest. For the CFAs, it provides a minimum n of 600 per analysis (ELL vs. non-ELL/non-SD pooled for each of the two test forms. This is a sufficient sample size to test the fit of a model to a 30-item test (McCallum, Browne, & Sugawara, 1996).

We do not anticipate any unusual problems requiring specialized sampling procedures. All forms of data will be collected one time only.

B-3. Methods to Maximize Participation Rates

We will begin recruitment by contacting superintendents whose districts' instructional and/or support staff have expressed interest in participating in a large-scale research study on accommodations for English language learners (ELLs). During this initial call, the recruitment team will confirm the superintendent's willingness to participate in the study. The team then will ask for a referral to the district's Director of Assessment (or staff member with a comparable role) to describe the study in greater detail. If these exploratory conversations indicate interest, the recruitment team will ask for referrals to school-level staff at prospective schools to confirm participation with site principals and discuss participation opportunities with ELL instructional and support staff. It is our expectation that district- and/or school-level participation in this study will ultimately require complete review by the superintendent and school board before formal commitment can be offered.

Once oral confirmation of study participation is received, a Memorandum of Understanding (see Appendix C) will be sent to each district outlining the support they will receive for participating in the study, the roles and responsibilities of research staff and school site staff, and estimates of the time required to collect data from students and teachers. The letters will include language that assures students, parents, and teachers that participation is voluntary. A list of eligible students, i.e., those for whom individualized data (e.g., English language proficiency level, standardized test scores) are available, will be drafted in conjunction with district staff. Our research teams will request referral to

district information systems specialists for extraction of student-level data. School staff will be informed about the study so they may encourage student participation.

Prior to the week of testing, a researcher will ensure that all students who are eligible for testing have returned parental/guardian consent forms to the school. A researcher will arrange for a reminder notice to be sent home from the school with any student who has not yet returned a signed parental/guardian consent form. Up to three reminders will be sent home every other week with the student prior to the day of testing. On the day of testing, no student will be tested without a signed parental/guardian consent form on file. If on the day of testing the sample of eligible students at any school is reduced by more than 15% due to absenteeism or missing consent forms, researchers will request permission to return to the school site within one week for one make-up session. Once all data are collected from schools, students will receive a T-shirt to recognize their participation and schools will be provided with copies of the final report for use in instructional planning.

We will use a combination of sound research design, strategic recruiting of participating districts and/or schools, and administration of relatively short tests (30 items) and surveys (12 questions) to ensure high rates of participation. Data structures have been developed to track recruited districts, carefully acknowledging that once a district or school has expressed interest to be in the study, it is possible that they will drop out of the study at a later date. We will closely monitor the test administration procedures and make quick and decisive adjustments to the administration protocol if participation rates fall below key targets. Such flexibility requires high-level attention to study progress. Extensive experience in administering and managing large-scale test administration efforts enables us to anticipate when and where problems may occur and to take pre-conceived steps to address them if they occur.

B-4. Tests of Procedures and Methods

A pre-pilot test cognitive interview process will be used as a field test for test administration protocols, the language background survey questions, and 40-50 math items from both test forms (one with modified items and one with original items). Through this "think-aloud" procedure (Appendix H), a convenience sample of eight California middle-school student volunteers (four ELL and four non-ELLs/non-SDs) will provide feedback about the ways in which they access and comprehend each prospective test item and their interpretations of the language background questions. Of particular interest will be the length of time required for participants to respond to each test item. We will observe and take written notes on their strategies for answering each test item and responsiveness to test administration protocols. The rich descriptive data from these interviews will be used to refine test administration procedures, finalize item selections for the pilot test, and improve the student survey.

In addition, during pilot testing, researchers will 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 the pilot test 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.

During the pilot test, 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 of sampled students receive a letter about the study (Appendix B) and are asked to sign and return the attached informed consent form. This data collection strategy is intended primarily to help answer *Research Question 3*: If properly implemented, can the use of linguistically modified items strengthen the technical quality (validity and reliability) of assessments of mathematics achievement for ELL students without modifying the technical adequacy for non-ELL/non-SD students?

Released NAEP and state test items are used in this study because these items have undergone thorough review by educators at the state- and national-levels for content and grade-level appropriateness and for sensitivity to bias. Assessment administrators considered these items to be exemplary and thereby fitting for release to the public. And because these items were deemed appropriate for use on operational assessments, all items have been thoroughly tested by large samples of students.

B-5. Individuals Consulted on Statistical Aspects of the Study Design

Neal Finkelstein, PhD, serves as the Project Director for this study. Dr. Finkelstein is currently the Co-director for Research Studies for the Regional Educational Laboratory-West (REL-W). As a Senior Research Scientist, he develops research and evaluation designs that study the impact of program implementation in K–12 public schools. He ensures that evaluation designs feature high standards of evidence, and oversees the implementation of randomized field trials in education settings, including site recruitment, data collection and analysis.

Prior to WestEd, Dr. Finkelstein worked on large-scale program evaluations and policy analyses encompassing K–12 and higher education, and the bridge between them. His areas of expertise include K–12 school finance, academic preparation programs for high school youth, school-to-work, and early childhood education. Each area involves the collection, management, and analysis of large quantitative data sets as well as questions of cost, cost-effectiveness, and the marginal cost of policy decisions in education at the state and federal level.

Dr. Finkelstein served as Director of Educational Outreach Research and Evaluation for the University of California Office of the President. There he implemented research and evaluation designs that studied the effectiveness of K–12 student and school academic programs initiated by the University of California on 10 campuses throughout the state. These programs emphasized the connections between K–12 education and postsecondary education opportunities for students.

Dr. Finkelstein can be reached by phone at (877) 938-3400, ext. 3171.

Andrea Lash, PhD, with 30 years of experience in social science research, brings to the project expertise in research methodology, psychometrics, and statistical analysis. Her recent experience with experimental research includes preparing, for the Institute of Education Sciences, the design of a national impact evaluation of teacher induction; conducting, with colleagues, an examination of the instructional sensitivity of NAEP test items in an NSF sponsored study, Validities of Science Inquiry Assessments; and designing methods to assess implementation fidelity in a national experiment of early mathematics instruction. As a psychometrician, Dr. Lash recently co-led a team examining the application of evidence-centered design to performance assessments. She also has conducted research into educational applications of Item Response Theory, served as psychometric advisor to research projects, and guided local educational agencies in the development of educational tests.

Dr. Lash has led the statistical analysis for large-scale program evaluations, such the evaluation of Title I accountability systems and a national evaluation of Charter Schools, using complex, multivariate methods to examine how the federal programs may influence student achievement. She received her MA in Educational Evaluation from Ohio State University, her MS in Statistics from Stanford University and her Ph.D. in Educational Psychology from Stanford University.

Dr. Lash can be reached by phone at (877) 938-3400, ext. 3103.

Chun-Wei (Kevin) Huang, PhD, serves as a Senior Data Analyst responsible for instrument design and data analysis for this study. As a Senior Research Analyst at WestEd, he works with other researchers to design and implement rigorous experimental trials within WestEd's Regional Educational Laboratory-West (REL-W). Dr. Huang ensures that the instruments used in these studies are reliable and valid and is responsible for conducting statistical analyses during all phases of the research. In addition to his work with REL-W, he provides assistance to colleagues with statistical and measurement modeling for other WestEd projects.

Prior to WestEd, Dr. Huang worked at CTB/McGraw-Hill as a Research Scientist. He was involved in several projects including two statewide testing programs. His main responsibilities as a Research Project Manager were to lead and conduct data analyses (e.g., test equating and scaling) in accordance with customers' requirements. He has taught statistics at both the undergraduate and graduate level.

Dr. Huang can be reached by phone at (877) 938-3400, ext. 3162.

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.
- Bloom, Howard S. (1995) Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5), 547-556.
- 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.