Book 3 of 3

Practical Guidelines for the Education of English Language Learners

RESEARCH-BASED RECOMMENDATIONS FOR THE USE OF ACCOMMODATIONS IN LARGE-SCALE ASSESSMENTS

CENTER ON INSTRUCTION

Practical Guidelines for the Education of English Language Learners

RESEARCH-BASED RECOMMENDATIONS FOR THE USE OF ACCOMMODATIONS IN LARGE-SCALE ASSESSMENTS

David J. Francis, Mabel Rivera
Center on Instruction English Language Learners Strand
Texas Institute for Measurement, Evaluation, and Statistics
University of Houston

Nonie Lesaux, Michael Kieffer
Harvard Graduate School of Education

Hector Rivera
Center on Instruction English Language Learners Strand
Texas Institute for Measurement, Evaluation, and Statistics
University of Houston

This is Book 3 in the series Practical Guidelines for the Education of English Language Learners:
Book 1: Research-based Recommendations for Instruction and Academic Interventions
Book 2: Research-based Recommendations for Serving Adolescent Newcomers
Book 3: Research-based Recommendations for the Use of Accommodations in Large-scale Assessments

2006

CENTER ON INSTRUCTION

This publication was created by the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston for the Center on Instruction.

The Center on Instruction is operated by RMC Research Corporation in partnership with the Florida Center for Reading Research at Florida State University; RG Research Group; the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston; and the Vaughn Gross Center for Reading and Language Arts at the University of Texas at Austin.

The contents of this book were developed under cooperative agreement S283B050034 with the U.S. Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Editorial, design, and production services provided by Elizabeth Goldman, Lisa Noonis, Robert Kozman, and C. Ralph Adler of RMC Research Corporation.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). Practical Guidelines for the Education of English Language Learners: Research-Based Recommendations for the Use of Accommodations in Large-Scale Assessments. (Under cooperative agreement grant S283B050034 for U.S. Department of Education). Portsmouth, NH: RMC Research Corporation, Center on Instruction. Available online at http://www.centeroninstruction.org/files/ELL3-Assessments.pdf

2006

To download a copy of this document, visit www.centeroninstruction.org.

TABLE OF CONTENTS

1 FOREWORD

3 OVERVIEW

- *3 Who Are English Language Learners?*
- *5 Second Language Literacy Acquisition*
- *7 Academic Language as Key to Academic Success*
- *9 Importance of Including ELLs in Large-scale Assessments*
- *11 Content Knowledge and Language Proficiency*

13 ACCOMMODATIONS AND REVIEW OF STATE POLICIES

- *13 Conceptual Framework*
- *13 Use of Accommodations*
- *14 Selecting Appropriate Accommodations*
- *16 State Policies and Practices on Accommodations for ELLs*

19 EFFECTIVE ACCOMMODATIONS FOR ELLS: RESULTS OF A META-ANALYSIS

- *19 Studies Included in Meta-Analysis*
- *21 Accommodations Used in the Selected Studies*
- *22 Methods for Meta-Analysis*
- *23 Results of Meta-Analysis*
- *28 Conclusions and Recommendations*

35 REFERENCES

41 APPENDIX A: LITERATURE SEARCH STRATEGY

43 APPENDIX B: STUDIES EXCLUDED FROM META-ANALYSIS

45 APPENDIX C: OVERVIEW OF META-ANALYSIS METHODS

49 APPENDIX D: DESCRIPTIVE INFORMATION AND EFFECT SIZE CALCULATIONS FOR 11 STUDIES USED IN META-ANALYSIS

51 APPENDIX E: FORREST PLOT OF EFFECT SIZES AND 95% CONFIDENCE INTERVALS FROM RANDOM EFFECTS MODEL

53 ENDNOTES

FOREWORD

The fundamental principles underlying the *No Child Left Behind (NCLB) Act of 2001* focus on high standards of learning and instruction with the goal of increasing academic achievement—reading and math in particular—within all identified subgroups in the K-12 population. One of these subgroups is the growing population of English Language Learners (ELLs). NCLB has increased awareness of the academic needs and achievement of ELLs as schools, districts, and states are held accountable for teaching English and content knowledge to this special and heterogeneous group of learners. However, ELLs present a unique set of challenges to educators because of the central role played by academic language proficiency in the acquisition and assessment of content-area knowledge. Educators have raised multiple questions about effective practices and programs to support the academic achievement of all ELLs, including questions about classroom instruction and targeted interventions in reading and math, the special needs of adolescent newcomers, and the inclusion of ELLs in large-scale assessments. This document focuses explicitly on this last issue and in particular on research-based recommendations on the use of accommodations to increase the valid participation of ELLs in large-scale assessments.

This document is organized into three sections. The first section provides an overview with important background information on the inclusion of ELLs in large-scale assessments and the role of language in content-area assessments. This background information lays the groundwork for understanding and selecting the types of accommodations that are likely to benefit ELLs. In the second section, we provide background information on accommodations, including the complementary concepts of *effectiveness* and *validity*, as they relate to proposed accommodations. We also review relevant research on state policies regarding accommodations for ELLs. In the final section, we provide descriptions of the most common accommodations that have been studied in the empirical research and conduct a quantitative synthesis (i.e., *meta-analysis*) of this research in order to determine those accommodations that are currently known to be most effective. Also, in this final section, we offer recommendations and conclusions for the use of accommodations in order to increase the valid participation of ELLs in state assessments.

Several bodies of research were consulted in developing this report. To provide sufficient background and context for the recommendations, relevant knowledge from developmental research on aspects of cognition, language, and reading known to play an important role in all students' success in assessments of academic achievement were consulted. However, the primary source of information was the research literature on accommodations for ELLs in large-scale assessments, including studies of the National Assessment of Educational Progress (NAEP) and, to a lesser extent state accountability assessments, because of their reduced prevalence. This literature provided evidence from randomized controlled studies using accommodations with ELLs and non-ELLs, quasi-experimental studies, and *post-hoc* analyses of data from a variety of studies that examined the effects of single or multiple accommodation strategies. We also drew heavily on previous reviews of the literature by Sireci, Li, and Scarpati (2003) and by Abedi, Hofstetter, and Lord (2004). In addition, we examined recent research by Rivera and Collum (2006) and reports of the National Research Council reviewing the underlying foundations of assessment accommodations, and state policies and practices with respect to the assessment of ELLs. The third section of the report provides a *meta-analysis* of the empirical research on accommodations. We provide a more detailed description of the search methods and statistical analysis techniques used to complete the *meta-analysis* in that section of the report.

OVERVIEW

Who Are English Language Learners?

The U.S. Department of Education defines ELLs as national-origin-minority students who are limited-English-proficient. The ELL term is often preferred over limited-English-proficient (LEP) since it highlights accomplishments rather than deficits. As a group, ELLs represent one of the fastest-growing groups among the school-aged population in this nation. Estimates place the ELL population at over 9.9 million students, with roughly 5.5 million students classified as Limited English Proficient by virtue of their participation in Title III assessments of English language proficiency. The ELL school-aged population has grown by more than 169 percent from 1979 to 2003, and speaks over 400 different languages, with Spanish being the most common (i.e., spoken by 70 percent of ELLs). By 2015, it is projected that 30 percent of the school-aged population in the U.S. will be ELLs. The largest and fastest-growing populations of ELLs in the U.S. consist of students who immigrated before kindergarten and U.S.-born children of immigrants[1].

This is an especially important statistic in the context of a report, such as this one, about effective accommodations to increase the valid participation of ELLs in large-scale assessments. In fact, many ELLs with academic challenges have been enrolled in U.S. schools since kindergarten, and by the upper elementary years do not have a formal designation to receive support services for language development. Instead, they are learners who have been identified as having sufficient English proficiency for participation in mainstream classrooms without specialized support. These ELLs typically have good conversational English skills, but many lack much of the academic language that is central to text and school success. For example, in several studies with minority learners in the elementary and middle school years—whether formally designated LEP or not—these students' vocabulary levels are often between the 20th and 30th percentiles[2]. Such low vocabulary levels are insufficient to support effective reading comprehension and writing, and in turn have a negative impact on overall academic success.

Contrary to its rapid development in size, the ELL population has met with limited academic success in U.S. schools[3]. When compared to their native English-speaking peers in all grades and content areas, the subgroup of ELLs

with a formal ELL or LEP designation lags behind. For example, on a national assessment of reading comprehension in 2005, only 7 percent of fourth grade ELLs with a formal designation scored at or above the proficient level, compared with 32 percent of native English speakers. Only 4 percent of eighth grade ELLs scored at or above the proficient level, compared with 30 percent of native English speakers. Similarly, while only 36 percent of all fourth graders score at or above the proficient level on a national assessment of mathematics, within the ELL population only 11 percent score at or above the proficient level[4]. Although learning disabilities are present in all groups, regardless of age, race, language background, and socioeconomic status, estimates of their prevalence range from only 5 to 15 percent of the population. Thus it is of concern that many ELLs are failing in school even though they do not have a learning disability[5].

Statistics on the performance of ELLs are generally based on the performance of students designated as Limited English Proficient (LEP) within state accountability systems. This designation is unlike others, such as gender or ethnicity, insofar as students' membership in the group of LEP students is dynamic and meant to be temporary. When ELLs have gained the proficiency in the English language needed to participate in grade-level classes, they lose their LEP designation, are required to participate in the mainstream classroom without specialized language support, and are no longer included in percent proficient calculations for the LEP subpopulation of a school, district, or state. Because language proficiency plays a significant role in student achievement, this reporting practice will tend to underestimate the achievement of the LEP group insofar as those students with the highest language proficiency are removed from the group as they become proficient in English.

Under NCLB, students can be counted within the LEP category for up to two years after becoming proficient in English, thus allowing more proficient students to contribute to the percent proficient for accountability purposes. This reporting practice mitigates the problem of underestimation somewhat. However, states' results are generally not reported separately for current and former LEP students. Rather, the former LEP students are simply included in the LEP category for up to two years after reaching the level of being considered proficient in English. Failure to distinguish between former and current LEP students when disaggregating accountability data makes it difficult to accurately evaluate the performance of schools in educating ELLs and to

accurately describe the academic achievement of ELLs. Recent efforts to examine the performance of former LEP students have shown that some ELLs do quite well in public schools[6]. On the other hand, many ELLs who are no longer formally designated (ELL, LEP) continue to struggle with academic text and language; these learners are a growing concern for students, parents, educators, administrators, and policymakers.

One of the significant benefits of the *No Child Left Behind Act* (NCLB) has been an increase in awareness of the academic needs and achievement of ELLs as a distinct student population. Under NCLB, schools are accountable for teaching English and content knowledge to these learners. As an identified subgroup, ELLs are participating in large-scale state assessments at higher levels than in the past. However, participation of ELLs remains an issue and concern for students, parents, school administrators, and government officials. Historically, these learners have had lower rates of participation, compared to native English speakers and non-minority students[7]. Whereas student participation in assessment is a direct target of the law, meeting the law's goals in this regard raises significant challenges to states and schools. It is not enough for students to participate in state assessments. Students' participation must lead to valid inferences about their achievement, and about the effectiveness of schools in educating this diverse group of students.

Second Language Literacy Acquisition

Unlike their native English-speaking peers, ELLs—particularly young children—are charged with the task of acquiring a second language while simultaneously developing their first and while developing the content-related knowledge and skills that define state standards. Many related factors significantly influence the performance of ELLs in the classroom including educational history, cultural and social background, length of exposure to the English language, and access to appropriate and effective instruction to support second language development.

Second language development relies very heavily on the availability of input from teachers, books, and peers that is both comprehensible and appropriate—especially in the classroom—and for some learners the process is facilitated by development of the first language. For example, a student who possesses a concept in his first language needs only to learn the label for the concept in his second language, whereas the student who lacks the concept in both languages must learn the concept and the label. Therefore, the success of

"learning" a concept in a new language depends on previous experiences and on instruction to facilitate and support acquisition in the second language, with careful attention to the conceptual knowledge that ELLs possess and need.

Acquiring reading skills in a second language is similar to the process used to acquire reading skills in the first language. For those ELLs who are literate in their first language—with exposure to appropriate and sophisticated instruction—much of their native language reading skills can be applied to their reading in the second language. However, several factors affect this process of applying of first language literacy skills in the acquisition of literacy skills in a second language. These include the individual's reading proficiency in her first language and the degree of overlap between the oral and written characteristics of the second language (i.e., English) and the ELL's native language. Similarities between languages that affect this process of learning to read in a second language include the conventions for writing (e.g., are both languages alphabetic, does writing progress from left to right in both languages, do they share orthographic elements, are they based on the same script?), commonalities in the sounds of the two languages and in the orthographic conventions for representing similar and different sounds, as well as the degree of overlap between languages in semantic elements or cognates. Cognates are words that have similar meanings and are written in similar ways in two different languages, often because of shared origins in another language (e.g., words that are similar in English and Spanish because of their shared origins in Latin). These factors affect the degree of similarity between languages, which in turn influences the degree to which students are able to apply native language reading skills in the first language to reading acquisition in English[8]. Whether ELLs have full proficiency or only beginning proficiency in oral language and reading development in their native language, developing these skills in a second language is not a trivial task. While simultaneously developing conversational ability and basic reading skills, these learners must quickly begin to develop oral and written academic language skills for the development of academic knowledge and success in content area classrooms.

Language plays an integral role in all academic learning. Consequently, any test of academic achievement is also, to some degree, a test of language ability. Thus, ELLs are likely to be disadvantaged when taking tests in a language in which they are not fully proficient. Test scores are used to judge students' ability to perform grade-level work in content areas. However, these scores

may, in fact, reflect ELLs' language abilities and not necessarily their competence in the content area (i.e., conceptual understanding and key facts), which may be otherwise evident on different types of assessments and under regular classroom conditions. There is reason for concern about the validity and reliability of test scores if test performance reflects individual differences in abilities that are related to, but distinct from, those that are the target of assessment.

In order to obtain valid and reliable test scores for all students, these sources of variance in test scores that are systematic, but irrelevant to the measurement of the ability of interest, must be controlled. This control can be achieved either through test design or through changes to standard testing conditions. Accommodations are one set of tools that can be used for these purposes. States and districts use accommodations to increase the participation rates and the validity of test scores for subgroups of students by controlling or eliminating sources of variability in students' test performance that are irrelevant to the ability being assessed.

This document reviews the current research-based[a] literature on the use of accommodations to support ELLs' participation in large-scale assessments. Large-scale assessments rely on the use of standard conditions in the planning, collecting, analyzing, and reporting of student data. However, even under uniform conditions, they cannot be guaranteed to yield valid and reliable results for all students, particularly those populations with unique needs. Consequently, states and districts have adopted policies and procedures for modifying tests and testing conditions for particular subgroups of students, one of which is ELLs, in order to increase the validity and reliability of inferences based on their test scores from large-scale assessments. For ELLs participating in large-scale assessments, there are many different accommodations currently in use in schools across the nation. However, state, district, and school administrators responsible for assessment pose multiple questions about effective practice in this regard, and they require guidance in selecting appropriate accommodations for ELLs. This report serves as a tool to aid administrators and practitioners who seek to make informed decisions on supporting ELLs' valid participation in large-scale assessments.

Academic Language as Key to Academic Success

Mastery of academic language is arguably the single most important determinant of academic success for individual students. While other factors—

[a] In this section of the report, the term research-based reflects a commitment to providing recommendations on the basis of direct evidence from research conducted with ELLs, evidence from research conducted with mixed samples of ELLs and native English speakers, as well as evidence from studies of state policies and practices with respect to assessment of ELLs.

such as motivation, persistence, and quantitative skills—play important roles in the learning process, it is not possible to overstate the role that language plays in determining students' success with academic content. Unfortunately, ELLs often lack the academic language necessary for success in school. This lack of proficiency in academic language affects ELLs' ability to comprehend and analyze texts in middle and high school, limits their ability to write and express themselves effectively, and can hinder their acquisition of academic content in all academic areas, including mathematics. Given the linguistic basis of developing knowledge in academic content areas, ELLs face specific challenges to acquiring content-area knowledge. As a result, their academic language and, therefore, their academic achievement, lag behind that of their native English-speaking peers. It is important to distinguish academic from conversational language skills, as many of the ELLs who struggle academically have well-developed conversational English skills. To be successful academically, students need to develop the specialized language of academic discourse that is distinct from conversational language. An example of the distinction between conversational and academic language may help to explicate this point:

> *When a student walks up to a newspaper stand and purchases a newspaper, he utilizes his conversational language skills to converse with the clerk and make the purchase. In contrast, other skills altogether are used to read and understand the front-page article, as well as to discuss the pros and cons of the proposed policy change that the article describes. The student might use still other skills to compare the writer's opinion to his, and to the opinion of the store clerk. The oral and written language required to engage in the latter "conversation" will involve more advanced and specialized vocabulary, more complex sentence structures, and more complex discourse structures than that required for the former.*

Many skills and factors are wrapped up in the notion of academic language. These include but are not limited to: vocabulary knowledge, including the multiple meanings of many English words, the ability to handle increasing word complexity and length over time, and understanding complex sentence structures and the corresponding syntax of the English language. A particular source of ELLs' reading difficulties relates to their limitations in academic

vocabulary—the words necessary to learn and talk about academic subjects. This academic vocabulary is central to text and plays an especially prominent role in the upper elementary, middle, and high school years as students read to learn about concepts, ideas, and facts in content-area classrooms such as math, science, and social studies. In doing so, ELLs encounter many words that are not part of everyday classroom conversation. These types of words (e.g., words like *analyze*, *therefore*, and *sustain*) are more likely to be encountered in print than in oral language, and are key to comprehension and acquisition of knowledge[9].

The need for well-developed academic language skills runs well beyond the academic skills necessary for success from kindergarten through twelfth grade. In fact, many learners—especially learners from minority backgrounds—who graduate from high school and enroll in post-secondary education often need additional support and remediation to succeed in their post-secondary classrooms. Incidentally, more freshmen entering degree-granting post-secondary institutions take remedial writing courses than remedial reading courses[10]. This highlights the importance of academic English as it relates to oral language, reading skills, and writing.

There is little disagreement among researchers and educators about the importance of the development of academic language for student achievement, or that limitations in this development are the root of most ELLs' academic difficulties. Similarly, there is little disagreement on the limited attention afforded to its development in most K-12 reading/language arts and content-area curricula. For these reasons, a basic premise that organizes this report is the need to attend to the role of academic language and to support its development in all educational endeavors. This is the case whether administering large-scale assessments to ELLs, or planning appropriate and effective instructional approaches, interventions, or specialized programs to meet their needs.

Importance of Including ELLs in Large-scale Assessments

Standardized, standards-based assessments play a prominent role in current approaches to education and school accountability. Various types of assessments are needed to monitor the effectiveness of instruction and, where necessary, to serve as indicators of the need for school improvement. Under NCLB, participation rates in state accountability assessments are vital indicators of

school performance. Historically, ELLs (and other special populations) were often excluded from large-scale assessments[11]. Limited English proficiency was perceived as preventing students from understanding questions or obtaining valid test results under standard test administration procedures. However, such exclusions serve to distort states' actual levels of performance, if students who do not participate in state accountability assessments, either through forced, voluntary, or school-encouraged exclusion, are less likely to score in the proficient range in comparison to students who participate in assessments. Exclusion of large numbers of students from participation in standards-based tests can result in substantial distortion of the percentage of students achieving proficiency. Perhaps more important, differences in exclusion rates across groups of learners, states, and/or districts can significantly obscure differences among them in the percentage of proficient students.

The stakes of large-scale assessments for individual students range from "low" for national assessments such as the *National Assessment of Educational Performance (NAEP)* to "high" for some state-mandated assessments that must be passed in order to be promoted to the next grade level or obtain a high school diploma. In fact, by 2008, 28 states in the U.S. will require that students pass a state-administered test for high school graduation[12]. For schools, districts, and states, the stakes of state-mandated assessments are high. They must ensure that all students participate in school accountability assessments and that increasing numbers of students from all designated subgroups score in the proficient range. Failure to reach adequate yearly progress targets can lead to increasing levels of sanctions for schools, districts, and states. In some states, significant incentives for teachers and administrators are linked to successful school performance. Whether linked to rewards or punishments, there is no question that the consequences can be significant for schools and districts.

NCLB recognizes the importance of high participation rates in order to obtain accurate information about proficiency rates for subgroups of students. For that reason, NCLB sets targets for participation rates in all student subgroups. However, if tests are not appropriately designed and students are not tested under appropriate conditions, language proficiency may unfairly and negatively influence the performance of ELLs. For example, literature on the assessment of students with limited English proficiency has demonstrated a substantial link between students' language proficiency and their performance

in content-area tests, a relationship which holds to a lesser degree for non-ELLs. In short, while participation of ELLs in state assessments is important, the goal is to accurately assess their proficiency with grade-level content-area material. To accomplish this goal requires tests that are designed and administered with ELLs in mind.

Content Knowledge and Language Proficiency

Researchers and practitioners are not surprised to discover that assessments of content-area knowledge and skills (e.g., science vocabulary, the ability to read and understand science or social studies texts, to understand and solve applied problems in mathematics) are also tests of language proficiency. Although there may be substantial differences between ELLs and their peers regarding content knowledge, research shows that estimates of the size of this knowledge gap is significantly affected by the language demands of the assessment. For the last decade, Jamal Abedi has led a program of research that has focused on large-scale testing and accommodations for ELLs. One of the principal findings of this extensive research is that assessments which have more linguistically challenging content yield the largest performance gaps between ELLs and native English speakers[13].

This finding is not unexpected. However, because language and knowledge are so inextricable, it is often difficult for practitioners to see the distinction between them. The most common examples used to make the distinction between language and knowledge typically draw on math word problems, where it is somewhat easy to imagine that students could know and understand the application of specific mathematical principles needed to solve the problem, but fail to grasp the essence of the problem due to the language demands inherent in presenting the problem on the assessment.

While it is somewhat easy to see this distinction in the solution of mathematics problems, it can be more difficult to distinguish language from content knowledge in other areas. Consider this example: An engineer who is a recent immigrant from Russia wants to be admitted into a course of study to become licensed as an engineer in the United States. The entrance exam requires that applicants solve a common problem encountered in their everyday professional lives; of course, the problem and its solution must be addressed in English. Although the Russian engineer speaks some English, it is much inferior to her Russian. As a result, it is likely that she will score more poorly on the

test than an engineer with comparable professional knowledge and expertise who is also a native speaker of English. While the Russian engineer might also be expected to get less out of the course of study than the native English speaker with comparable knowledge, due to her more limited English she may in fact have more professional knowledge and get more out of the course than native English speakers who score at her level. How entrance exam performance might relate to subsequent performance in the course of study gets at the heart of the question of the validity of test scores. For the scores to have equal validity in predicting performance in the course, we should expect the same outcomes for native English speakers with the same score as the Russian speaker. However, it is quite possible that the Russian speaker might gain more from the course than native speakers with the same score for at least two reasons. First, she is likely to make gains in English and develop her technical language through her time in the country and the course of study. Second, she has superior professional knowledge on which to build. This example can be extended to represent the use of end-of-course exams in algebra to determine if students should be admitted to a course of study in geometry or trigonometry, or instead offered remedial instruction in algebra. The challenge is to design exams and testing situations that limit the contribution to test scores of individual differences in abilities that are not the target of assessment.

ACCOMMODATIONS AND REVIEW OF STATE POLICIES

Conceptual Framework

Assessments are given annually to large numbers of students in public schools for many purposes. The most common and most public purpose for these large-scale assessments today is school and student accountability. These assessments are generally high stakes, insofar as significant consequences are often attached to the performance of individual students (e.g., promotion to the next grade, graduation), as well as to the performance of groups of students (e.g., school accountability). The high-stakes nature of these assessments places a premium on assessment results that are valid and reliable for all students. At the same time, participation of all students in school accountability assessments is essential to ensuring that all students receive the same high-quality public education. When students are held out of the accountability system, there is the risk that they will also be ignored during instruction or held to lower performance expectations. In this light, NCLB has specific guidelines on participation rates for all students in state accountability assessments, guidelines which place considerable emphasis on the valid participation of ELLs and other designated populations (e.g., students with disabilities, ethnic minorities) in these assessments.

Use of Accommodations

When faced with a large-scale test in English, an ELL must direct more cognitive resources to processing the language of the test compared to a student who is fully proficient in English. Therefore, the ELL will have fewer resources available to attend to the content being tested. One way to facilitate the valid participation of ELLs in large-scale assessments is to provide them with appropriate *accommodations* to the testing conditions. The term *accommodation* encompasses alterations to standard test administration procedures including, but not limited to, how the assessment is presented to the student, how the student is allowed to respond, any equipment or materials to be used, the extent of time allowed to complete the test, and changes to the environment in which the student takes the test[14]. For example, students might be given extra time to complete the assessment, or might be provided a glossary that defines key terms.

An appropriate accommodation will focus on factors that affect the test scores of students who receive the accommodation, but which are not themselves the target of assessment. At the same time, these factors should not affect the performance of students who do not receive the accommodation. If all students were provided with the accommodation, only the test performance of those who need the accommodation (i.e., in this case, ELLs) would be affected by it, and the skill of interest would still be assessed. In essence, the accommodation must address the needs of the student without invalidating the test score as a reflection of the construct being assessed. In light of these factors, it is quite clear that appropriate accommodations for ELLs will provide either direct or indirect linguistic support[15] in order to minimize the cognitive effort that ELLs need to expend to process the non-construct related language of the test and to maximize the cognitive effort available for accessing the meaning of test items and passages.

Selecting Appropriate Accommodations

Individual accommodations, or combinations of accommodations, should be selected on the basis of their effectiveness and the specific needs of an individual student. The fact that two separate accommodations might be effective in isolation does not imply that the two will be doubly effective, or even equally effective when used in combination. When two or more accommodations are used together, there must be a specific rationale for doing so. For example, the use of dictionaries is usually bundled with extended time, based on the rationale that use of the dictionary takes students' time away from testing. It is important to take such factors into account when examining the literature and making decisions on the likely impact of an accommodation or suite of accommodations when used in practice. In addition to consideration of their effectiveness and individual student needs, accommodations during testing must match those received during classroom instruction. For instance, ELLs vary in the language and literacy skills in their first language. One accommodation that has been studied and recommended for ELLs is bilingual dictionaries. However, bilingual dictionaries should not be expected to be effective for students who are not literate in their native language; moreover, they have been found to be ineffective when students do not have experience using them during regular class instruction. Similarly, native language adaptations of English language assessments have been found in some studies to

negatively impact student outcomes, due to mismatch between the language of assessment and the language of instruction, or a lack of native language literacy. ELLs cannot be *assumed* to be literate in their first language, nor can they be assumed to be sufficiently literate in their first language for native language assessment to serve as an effective accommodation[16].

There are several dimensions along which accommodations for use with ELLs can be evaluated. Among the most important are the dimensions of *effectiveness* and *validity*, along with the *feasibility* of implementation in terms of cost and effort. Of the three dimensions, the first two are paramount insofar as accommodations which are not effective will not lead to improved test scores for students receiving the accommodation. Thus, *effectiveness* is the extent to which the accommodation leads to improved test scores for students receiving the accommodation. However, to be *valid*, an accommodation should be differentially effective. That is, the accommodation should improve the performance of students who need the accommodation, but not improve the performance of students who do not need it. The *validity* of an accommodation is, in part, the extent to which the accommodation only affects the performance of students who need the accommodation. Accommodations which lead to improved test scores for all students may alter the construct being measured. Such accommodations are unacceptable in large-scale assessment because they alter the validity of test scores. Validity, as applied to accommodations, refers to the extent to which the accommodation preserves the nature of the construct being measured and thus allows for valid inferences about students' standing on the construct of interest when based on a test score obtained under accommodated testing conditions. Generally, accommodations are not considered valid if they lead to improved test scores for students who do not require the accommodation. Only once accommodations have been deemed effective and valid does relative cost become a factor in selecting and providing accommodations to individual students.

Finally, there is the problem that an effective accommodation for one content-area assessment, and for one student, may not be similarly effective for others. For example, simplifying the complexity of items in English (see below) may be a generally valid accommodation for math assessment, but not valid for a language arts assessment in which the ability to understand and use complex English is central to the construct being measured. Moreover, the effectiveness of an accommodation may vary according to student

characteristics (e.g., language proficiency in English, literacy in the native language, or grade level), or the instructional context (e.g., participation in native language instruction or opportunities to use an accommodation tool, such as bilingual or English language dictionaries, during regular instruction).

State Policies and Practices on Accommodations for ELLs

Educational agencies across the nation provide accommodations to ELLs as needed[17]. The criteria for selection and strategies for implementation vary by state, according to many factors, but the specific accommodations can be grouped loosely into two broad categories based on their general focus: *Modification of the Testing Conditions* (e.g., scheduling, setting, timing, use of tools such as dictionaries and overlays, etc.) and *Modification of the Test* (e.g., directions, items, and/or student response options). Rivera, Collum, Shafer Willner, and Sia (2006) provide a comprehensive table of 75 different accommodations currently in use with ELLs and a more elaborate taxonomy for classifying accommodations. However, as they note, many accommodations allowed by states are questionable for this population of students, either because they are not theoretically defensible, because they do not specifically target the language difficulties of ELLs (either directly or indirectly), or because they lack research evidence.

Although appropriate for other students, such as students with vision impairments, or with attention deficit and hyperactivity disorder, many accommodations reported to be in use by states are questionable or even inappropriate for ELLs. Some of these include testing in small groups, one-to-one testing, administering tests by specific staff, assigning students preferred seating, and allowing students to take the assessment in a separate location, such as a study carrel. While these accommodations may not lead to invalid assessment for ELLs, they are not expected to be effective in improving ELLs' performance because they neither directly nor indirectly relate to the ELLs' challenges with academic English. Some ELLs may, of course, also have a particular disability or impairment that simultaneously qualifies them for other specific accommodations unrelated to their status as an ELL. A student's status as a member of one subgroup should not preclude him from receiving accommodations appropriate for other subgroups of which the student is also a member. However, accommodations based on a disability framework are not generally responsive to the needs of ELLs, and would not be considered

generally appropriate under a theoretically sound framework for accommodations for ELLs, that is, one focused on the linguistic needs of ELLs.

Table 1 provides a partial listing of accommodations in use by states that are, at the very least, responsive to the potential needs of ELLs[18], even if not previously demonstrated to be effective or valid. Those which have been researched using experimental and quasi-experimental studies are marked with an asterisk and are discussed in detail in the next section. It is clear from the listing in Table 1 that only a handful of the theoretically defensible accommodations in use with ELLs have also been researched empirically.

Table 1. Partial Listing of Accommodations Responsive to Needs of ELLs

Accommodations of Testing Conditions	Accommodations as Test Modifications
Extended time*	Directions read in English
Breaks offered between sessions	Directions read in native language
Bilingual glossaries*	Directions translated into native language
Bilingual dictionaries*	Simplified English*
English glossaries*	Side-by-side bilingual version of the test*
English dictionaries*	Native language test*
	Dictation of answers or use of a scribe
	Test taker responds in native language

EFFECTIVE ACCOMMODATIONS FOR ELLS: RESULTS OF A META-ANALYSIS

A meta-analysis[b] of relevant research was conducted in order to address the question of which accommodations can and should be recommended for use with ELLs—those that are effective and valid, and the conditions under which they are so. A meta-analytic review is a specific approach to research synthesis that attempts to quantify the effect of an intervention and to determine if there are factors which moderate those effects. In the case of test accommodations for ELLs, likely factors that might alter the effects of accommodations are individual characteristics of students such as grade level and language proficiency, content area, and the type of accommodation (i.e., are all accommodations equally effective, or do accommodations differ in their effects for ELLs?).

Search Procedure. To be included in this review, empirical studies on accommodations for ELLs were obtained through two steps. First, we conducted a comprehensive search of online databases. Second, we examined a collection of studies previously reviewed by Sireci, Li, & Scarpati (2003) and/or by Abedi, Hofstetter, & Lord (2004). For specific search strategies, see Appendix A.

Inclusion and Exclusion Criteria. Studies included in the meta-analysis were those that employed an experimental design that allowed for the examination of the effects of individual accommodations or in some cases, two bundled accommodations. Although the initial criteria included quasi-experimental designs as well as randomized controlled trials, no studies were found with quasi-experimental designs examining individual accommodations. Hence, all studies included in the meta-analysis were true experiments. Both published studies and technical reports were included in the meta-analysis. Using these criteria, 21 studies were found. Several of these studies, however, had to be excluded from the meta-analysis for various reasons involving either reporting or methodology. In some instances, studies did not report the necessary information to quantify the effects of accommodations, or did not allow for results to be disaggregated for ELLs. For a complete list of excluded studies and a rationale for exclusion, see Appendix B.

Studies Included in Meta-Analysis

The effect of accommodations in large-scale testing for ELLs has been researched using randomized, controlled experiments. This research base is

[b] A meta-analytic review is a specific approach to research synthesis that attempts to quantify the effect of an intervention. For practical introductions to meta-analysis, see Cooper (1998) and/or Lipsey & Wilson (2001). For more extensive details on conducting meta-analytic reviews, see Cooper & Hedges (1994). For more extensive discussion of the statistical methods involved in meta-analysis, see Hedges & Olkin (1985).

large enough to merit a quantitative review/meta-analysis, but is not necessarily extensive when one considers the magnitude of the challenge facing schools and states with respect to variation in the K-12 ELL population, the variety of content areas, the possible types of accommodations, and the potential individual and contextual factors that could alter the effectiveness of any particular accommodation or bundle of accommodations.

Following application of the search rules, and the inclusion and exclusion criteria described in Appendices A and B, eleven studies remained for use in the meta-analysis. Each study used random assignment of ELLs and non-ELLs to testing conditions with and without accommodations. These eleven studies involved thirty-seven different samples of students and reported thirty-seven different tests of the effectiveness of accommodations for ELLs. Thirty-three involved either 4th (n=11) or 8th (n=22) grade students, and four involved either 5th or 6th grade students (n=2, each). Seventeen of the thirty-seven tests of the effectiveness of accommodations used a test of math as the outcome measure, nineteen used a science test, and only one used a reading test. Twenty-eight of these effects involved the NAEP assessment or particular NAEP items (n=22), or a test based on the NAEP and TIMSS (n=6) assessments, whereas nine effects were based on a state accountability assessment (eight from two studies using the Delaware state test, and one using the Minnesota state test).

Finally, together, these thirty-seven tests focused on seven different types of accommodation: Simplified English (n=15), English dictionary/glossary (n=11), bilingual dictionary/glossary (n=5), extra time (n=2), Spanish language test (n=2), dual language questions (n=1), or dual language booklet (n=1). As mentioned, some estimated effects came from studies that involved multiple accommodations in the form of extra time bundled with one of the three other accommodations: Simplified English (n=2), English dictionary (n=3), or bilingual dictionary (n=2). Thus, two effects of the thirty-seven were from studies that involved extra time without other accommodations, whereas seven effects were based on studies that involved extra time coupled with one other accommodation. One study allowed extra time to all participants, and thus is not coded as involving extra time[19]. All but two of the reported effect size estimates are based on paper and pencil tests; the remaining two used computerized assessments.

Accommodations Used in the Selected Studies

The accommodations that are theoretically justifiable for English language learners are those that address the language demands of the test and the language needs of the ELLs in some way. The accommodations may be used individually or in combination, as needed. As described above, the intention of each accommodation described below is to reduce the degree to which the test scores of ELLs represent construct-irrelevant language abilities rather than their knowledge of the content area of interest.

Simplified English. This accommodation involves linguistic changes in the vocabulary and grammar of test items to eliminate irrelevant complexity while keeping the content the same. Some of these changes may be accomplished by eliminating non-content related vocabulary, shortening sentences and using simple sentence structures where possible, using familiar or frequently used words, active instead of passive voice, and using present verb tense where possible[20].

Customized English dictionaries or glossaries. The use of customized English dictionaries or glossaries involves adding definitions or simple paraphrases for potentially unfamiliar or difficult words in test booklets (usually on the margins). Another variation on this accommodation is to provide computerized tests with built-in English glossaries. Typically, this latter variation on this accommodation involves a computer program that provides a simple and item-appropriate synonym for each difficult non-content word in a test[21].

Bilingual dictionary, glossary, or marginal glossaries. ELLs are given access to dictionaries, glossaries, and marginal glossaries with words written in English and the student's native language. Another version of this accommodation is the use of computerized tests with bilingual glossaries built in[22].

Extra time. Providing more time than usual to complete test sections is among the most frequently used accommodations. This accommodation does not involve making changes to the test itself, but to the testing conditions. Extended time is usually provided in combination with other types of accommodations. The rationale is to allow the ELL extra time to process the language of the test, or in the case of bundling extra time with another accommodation, such as an English language dictionary, to allow for the time required to use the bundled accommodation[23].

Dual language test booklets. This accommodation involves changes to the format in test booklets. The booklets include English items on one side and the

corresponding items translated into the learner's first language placed onto facing pages[24].

Native language tests. Tests are adapted to the student's primary language. Typically, these are not translated tests, but adapted to preserve the meaning of the original text. The most highly preferred method of adapting a test to another language is to use *back translation*. In back translation, the test is first translated from the original language of the test into the native language version by a proficient speaker, reader, and writer of both languages. The adapted test is then translated back into the original language by an independent, bilingually proficient individual and the two original language tests are compared for equivalence. If the two original language versions are deemed to be different, the process is repeated, focusing on correcting those areas of the test which were not successfully adapted.

Methods for Meta-Analysis [c]

To evaluate the effectiveness of accommodating assessments for ELLs, and to examine the effectiveness of the different types of accommodations, we conceptualized effectiveness as having two distinct, but related components, each reflected by an effect size. This conceptualization is especially important for educators faced by the challenge of selecting suitable accommodations that must be both effective and valid. The first component, an index of *effectiveness*, reflects the degree to which the accommodation leads to improved performance for ELLs. The second is an index of the *validity* of the accommodation, which examines the impact of the accommodation on the performance of non-ELLs, with the assumption that a valid accommodation should have, at most, a negligible effect on their performance. Larger numbers are preferred for the effectiveness index and smaller numbers are preferred for the validity index. For the sake of computing average effect sizes, we treated each study sample as the unit of analysis, for a total of thirty-seven samples.

To compute average effect sizes across the entire set of samples, and for all samples addressing specific accommodations, we averaged across different outcomes and grades involved in studies of a particular accommodation. In averaging the different effect sizes, we weighted the individual effect sizes according to their precision. The precision of the effect size estimate is determined by the estimated effect size itself and by the sample size in the two groups of students involved in the comparison. In averaging the weighted

[c] This section of the report is moderately technical. Although we have attempted to shape this section for readers with little or no experience with meta-analysis, readers who are not interested in the details on effect size measurement, computation of average effect sizes, and units of analysis can skip to the next section without loss of continuity.

effect sizes, more precise estimates are given greater weight. For a more technical and detailed description of the methods used in this meta-analysis, see Appendix C.

Results of Meta-Analysis

In Table 2 (see page 31), we present the results of the meta-analysis, including the weighted average effect sizes for each accommodation. Also included are the standard error of the average effect size, a 95% confidence interval, and a test of the hypothesis that the average effect size is zero. The results in Table 2 tell a somewhat disheartening story. Of the seven types of accommodations used, only *one* had an overall positive effect on ELL outcomes. That is, only one accommodation (viz., English language dictionaries and glossaries) produced an average effect, which is positive and statistically different from zero, while one other (Spanish language assessments) showed significant variability across the estimates of its effects. This accommodation may be effective for some, but not for all ELLs, depending on the language in which they are receiving instruction. Below we provide a more detailed discussion of the results of the meta-analysis.

Dictionaries and Glossaries (English and Bilingual). Based on eleven effects, the use of English language dictionaries (and glossaries) was the only accommodation found to have a statistically significant and positive average effect size, albeit a small one[d]. The eleven effect sizes that went into this average were based roughly equally on studies of math and science in either 4th or 8th grade. Moreover, effects were judged to be consistent across the set of eleven effects. Although there is no statistical evidence to suggest that the effect sizes are different across the collection of eleven effect sizes, studies involving this accommodation varied along several interesting and potentially important dimensions. One of these, extra time, is felt to be critical to the successful use of dictionaries as accommodations. Three of the studies of English language dictionaries and glossaries also afforded students extra time to complete the examination. A direct comparison of the three studies that used extra time plus English language dictionaries and the eight studies that did not shows a somewhat higher effect size for studies that did not involve extra time (average effect size of 0.238, s.e.=0.075) relative to accommodations that allowed extra time with the glossaries (average effect size of 0.074, s.e.=.062). A second important variation in these studies is the format of the assessment,

[d]Effect sizes did not vary significantly across the 11 effects that involved English language dictionaries or glossaries ($Q(10)=14.804$, $p<.139$). These eleven effects came from studies involving math (n=6) or science (n=5) in either 4th (n=4) or 8th (n=7) grade.

which was either a paper and pencil test with paper glossary (9 studies), or a computerized test with a computerized glossary (2 studies). Comparison of the two test formats showed a slightly higher effect size for computerized tests (average effect size of .284, s.e.=.145) relative to paper and pencil tests (average effect size of .161, s.e. = .060). Thus, although these differences are not statistically significant, the number of studies for some conditions is small. Moreover, the sample size is too small to examine possible interactions between test format and extra time in moderating the impact of English language glossaries. We should also add that in our coding of studies, Abedi, Courtney, Mirocha, Leon, & Goldberg (2005) was not coded as involving extra time because students in the standard testing condition also received extra time. Thus, from the standpoint of testing the accommodations, the time available to complete the test is consistent across study groups. However, it is also true that the effect of the glossary in this study cannot be assumed to be the same if extra time had not been allowed with the glossary. On balance, it seems reasonable to conclude at this time that English language dictionaries offer an effective accommodation for ELLs, the effects of which may be moderated by test format and the allowance of extra time. Although current evidence suggests that effects are consistent across these dimensions, more subtle conclusions may be possible with additional research.

Bilingual dictionaries and glossaries, in contrast, did not show a positive effect. Moreover, despite being based on just five estimates of effect size drawn from three studies, tests indicated that effect sizes were not consistent across the collection of effect size estimates[e]. All five effects in this collection involved 4th or 8th grade science assessment, but the two largest effects were of opposite sign, and both came from studies with 4th grade ELLs. While it is difficult to make conclusive inferences based on just two conflicting results, the findings suggest that the effect of this accommodation may be very different in different contexts or among different populations of students, and may reflect unobserved differences in instruction. It is also possible that bilingual glossaries are effective for a specific group of ELLs—those who are literate in their first language and/or who have received content-area instruction in their first language. This disparity in the collection of studies examining bilingual dictionaries and glossaries merits further study. The current pool of studies examining this accommodation is small, but the effects appear to vary despite being restricted to a relatively homogeneous set of outcomes and grades.

[e]The point estimates for the five effects ranged from -.289 to +.452. The two largest effect sizes, both of which were statistically different from 0, were of opposite sign.

Simplified English. The Simplified English accommodation has received considerable attention and been discussed favorably in the literature on accommodations. Of all the accommodations reviewed here, Simplified English has been studied most frequently. Despite the generally favorable disposition of researchers and psychometricians toward Simplified English as an accommodation, as Table 2 shows, the overall average effect size for this accommodation was not significant[f]. Moreover, the test for heterogeneity suggests that effect sizes were consistent across the collection of effects for this accommodation. In looking at the collection of individual effects, it is clear that some of the randomized studies involving this accommodation employed small sample sizes of ELLs, and as a result, effect sizes from these studies are not very precise. At the same time, the effect sizes based on the larger sample sizes tended to be very small (see Appendix D for details on all of the studies addressing each particular accommodation). On the basis of these findings, Simplified English would not be judged to be an effective accommodation to reduce performance gaps between ELLs and non-ELLs. At the same time, in reaching conclusions about the effects of Simplified English, educators must keep in mind that the pool of studies examined here for this accommodation remains small and somewhat narrowly focused in terms of grades, content areas, and type of assessment. In particular, few state tests have been involved in the research on Simplified English as an accommodation for ELLs. It is possible that results with other state tests may be different.

Still, practitioners should be realistic in their expectations for performance improvements when ELLs use Simplified English as an accommodation. In addition to the fifteen effect sizes taken from the randomized experiments, two repeated measures studies were also completed using Simplified English. In one of these studies[25], ELLs scored higher when taking a test comprising Simplified English items than when taking a test comprising standard items. While the significant difference in performance favoring Simplified English is encouraging, the improvement in performance had little practical significance[g]. In the other study[26], the overall difference between Simplified English and standard items for ELLs indicated that the accommodation had a negligible effect on students' performance. This difference, in addition to being small, was also comparable to the effects of Simplified English for non-ELLs in the sample.

[f] Moreover, the effect sizes do not differ statistically across the collection of fifteen effects, despite their ranging from -1.295 to +.649, with at least four large positive effect sizes and three large negative effect sizes.

[g] The raw mean difference in performance for ELLs was .165, or less than 2/10ths of an item on a 10 item test, and was statistically comparable to the raw mean difference of .144 between tests for non-ELLs. Even if the test were lengthened to four times its present length, the ELLs would be expected to gain less than one item from the Simplified English accommodation.

In summary, the findings supporting the effectiveness of Simplified English are weak. While it is possible that the effects of Simplified English vary according to variables such as grade level, content area, and the nature of the assessment, the evidence does not currently support this conclusion. In spite of its prevalence in the research as an accommodation for ELLs, it appears unlikely that substantial improvement in ELLs' performance will result from widespread use of Simplified English as an accommodation. Further, there is little evidence to suggest how this accommodation might be made more effective. On the positive side, there is also little evidence to suggest that Simplified English invalidates assessments, or that it can have potentially negative consequences for students. Although some researchers have cautioned that Simplified English can lead to negative performance for ELLs, there does not appear to be strong support for this assertion based on the studies reviewed here.

Spanish Versions of Assessments: The results in the top half of Table 2 show that students scored worse when Spanish language assessments were used as an accommodation. However, the test of homogeneity of effect sizes also shows that effect sizes were not consistent across the two studies, and as a result, the fixed effect mean in the top half of Table 2 should be ignored in favor of the random effects mean reported in the bottom half of Table 2. This mean is a positive .302, but is not statistically significantly different from zero. The effect sizes for this accommodation were 1.064 (s.e.=.364) and -0.376 (s.e.=.106). Both effect sizes come from the same study, but from two different samples of students. One was Hispanic students instructed in Spanish, whereas the second was Hispanic students instructed in English. Not surprisingly, the positive effect size for Spanish language accommodation occurred for students instructed in Spanish, whereas the negative effect size occurred when students instructed in English were given a Spanish language assessment. Whether similar effects would be seen in other grades or with other content areas, and whether important student characteristics (e.g., native language literacy and number of years of English instruction) might moderate these effects are questions to be addressed in future research. Despite the relatively small collection of studies involved, it stands to reason that students who have not been instructed in their first language, or who are not literate in their first language, will not have their test performance facilitated by a native language accommodation.

Extra Time and Dual Language Tests: In addition to the accommodations mentioned above, a few studies examined extra time as an accommodation. Two studies looked exclusively at extra time, while a handful of studies bundled extra time with other modifications, specifically bilingual dictionaries and glossaries (n=2), English dictionaries and glossaries (n=3), and simplified English (n=2). As mentioned above, one study also used extra time in all study conditions, including the unaccommodated condition, such that students in the accommodated and unaccommodated conditions received the same time. Finally, two studies examined the effects of dual language assessments. Dual language booklets are test booklets that contain both the traditional assessment as well as a translated or linguistically adapted test, such that the student can either answer test questions in English, or in the accommodated language, usually the child's first language.

In the collection of studies reported in the meta-analysis, extra time had a positive effect, but the effect was not statistically different from zero. In the two studies of dual language accommodations, effects were not different from zero, but they were opposite in sign, just as with Spanish language tests. These findings with regard to bilingual assessments, although inconclusive due to the small number of studies, suggest that this accommodation may operate similarly to native language assessments and only be appropriate for students who are literate and/or instructed in their native language.

Consistency in Effect Sizes: Finally, the results in Table 2 relating to tests of heterogeneity across the collection of studies shows that the effect sizes varied both within and between accommodations (see results for *Q* statistic for TOTAL WITHIN and TOTAL BETWEEN variation). These results indicate that there is substantial variability in effect sizes across the collection of studies, but that the majority of this variability (25.5 / 87.3 = 29.2%) is due to differences in average effect sizes across the seven different types of accommodation. In other words, the differences across these studies were somewhat due to the accommodations employed, although factors that vary within the group of studies on particular accommodations, such as the grade level, the content area, or the test type also potentially contribute to the variability in effect sizes.

Although the findings on the effectiveness of accommodations are not particularly strong, we must keep in mind that this is a relatively small and recent body of research. Until recently, there was only one individual, Dr. Jamal Abedi, programmatically engaged in research in this area. Researchers and

practitioners alike are deeply indebted to him for his pioneering and tireless efforts in this area, without which little, if anything, would be known about the effectiveness of accommodations for ELLs.

Conclusions and Recommendations

This document seeks to provide administrators and practitioners with research-based recommendations on the use of accommodations to increase the valid participation of ELLs in large-scale assessments. Based on the information reviewed in the three preceding sections of the document, we offer the following summary, conclusions, and recommendations.

This review highlighted the importance of academic language in the educational attainment of ELLs, and the fundamental role that language proficiency plays in assessments of all content areas. In selecting accommodations for ELLs, it is important to keep in mind that appropriate accommodations will address the linguistic needs of the student. Moreover, research on second language acquisition provides a useful framework for thinking about linguistically appropriate accommodations[27]. While it is often appropriate to bundle accommodations, in doing so there should always be an explicit rationale for combining specific accommodations. Bundling accommodations that are individually effective cannot be assumed to yield an effect that is equal to or greater than that of the individual accommodations. That is, "more" cannot be assumed to be "better."

There are many accommodations that can be considered linguistically, although not all have been tested in terms of their effectiveness or validity. Still, linguistically appropriate accommodations include changes in the testing conditions (e.g., allowing extra time, the use of dictionaries or glossaries) as well as modifications to the test itself (e.g., bilingual assessments, native language adaptations, allowing the student to respond in her native language). Regardless of the choice of accommodations, the accommodations used during testing should match those used during classroom instruction. In addition to ensuring that ELLs have had experience with accommodations in the instructional setting, one cannot assume that ELLs will perform better when tested in their first language. The choice of bilingual or native language assessments as an accommodation for ELLs must take into account the students' oral proficiency and literacy in their native language, as well as the language in which they have been instructed. Native language assessments

cannot be assumed to offer students a linguistically appropriate accommodation. Finally, in selecting accommodations, consideration must be given to both the effectiveness and the validity of the accommodation.

This review suggests that appropriate selection and differentiated use of accommodations in large-scale assessments can assist ELLs in participating in large-scale assessments without invalidating test results. And yet, none of the accommodations examined has "leveled the playing field" for ELLs. Many accommodations currently in use across the country do not directly or indirectly address the linguistic needs of ELLs. At the same time, many of the linguistically appropriate accommodations that have been studied empirically were found in this review to have little or no impact on the test performance of ELLs. There are many more linguistically appropriate modifications that have not been studied at all. Moreover, of the appropriate accommodations that have been studied, none has been widely studied in terms of the number of content areas, grade levels, test types, test formats, or student characteristics for which the accommodation has been tested. Without better access to quality instruction that works to build ELLs' academic language proficiency and content-area knowledge, we cannot expect that their test performance will substantially improve through appropriate accommodations. Research on ELLs has shown that these students, due to their deficiency in the English language skills necessary to independently read and learn from grade-level material, are regularly excluded from participation in the curriculum. Separate reports on Instruction and Interventions and on Programs for Newcomers were developed to accompany this report in an effort to provide guidance to practitioners on increasing ELLs' access to rich and challenging academic content.

The accommodation that had the most substantial effect on student performance was providing ELLs with English language dictionaries. Given the underlying importance of English language proficiency on ELLs' academic success in school, this finding makes sense. However, simply providing ELLs with a dictionary when they take large-scale assessment is not effective. For any accommodation to be successful in the testing situation, students must have experience with it during regular instruction. Thus, students who have never used a dictionary during instruction cannot be expected to benefit from its use during an assessment. It is generally felt that the use of dictionaries should be accompanied with extra time to make up for time lost in use of the dictionary. However, the results in the meta-analysis do not support this

conclusion at this time. Granted, the number of studies to inform this decision is small. Nevertheless, the average effect size was somewhat smaller for studies involving dictionaries that allowed extra time for students in accommodations compared to studies involving dictionaries where the time allowed students was the same in the accommodated and unaccommodated conditions. It seems safest at this point to consider the importance of extra time to the effectiveness of English language dictionaries an open question that merits further investigation.

The alignment of curriculum, instruction, and assessment is crucial to the academic success of all students. For ELLs, this also means an understanding of their unique language learning needs and the diverse academic backgrounds they bring to the testing situation. In turn, educators must consider the student's language skills, and how they influence both the instructional needs of the student and the academic supports that will ensure his valid participation in large-scale assessments. Providing these aids during instruction and assessment will afford these students an opportunity to learn and to demonstrate their knowledge and abilities in spite of what may be their limited proficiency in English.

Table 2. Average Effect Sizes and Variance Components for Seven Accommodations Used in Randomized Experiments

Accommodation	Results for Fixed Effects Analysis									
	Number of Samples	Effect Size and 95% Confidence Interval				Test of Mean Effect = 0		Test of Heterogeneity in Effect Sizes		
		Mean Effect Size	s.e.	Lower Limit	Upper Limit	Z	p	Q	df(Q)	p(Q)
Bilingual Dictionary-Glossary	5	-.096	.065	-.223	.031	-1.479	.139	13.53	4	.009
Dual Language Booklet	1	-.177	.148	-.467	.112	-1.199	.231			
Dual Language Questions + Read Aloud in Spanish	1	.273	.195	-.109	.654	1.401	.161			
English Dictionary-Glossary	11	.146	.043	.063	.230	3.427	.001	14.804	10	.139
Extra Time	2	.209	.142	-.069	.488	1.473	.141	0.155	1	.693
Simplified English	15	.020	.043	-.064	.104	.473	.637	19.830	14	.136
Spanish Version[h]	2	-.263	.102	-.463	-.062	-2.572	.010	14.465	1	<.001
TOTAL WITHIN								62.789	30	<.001
TOTAL BETWEEN								25.540	6	<.001
OVERALL MEAN	37	.034	.025	-.016	.084	-1.342	.180	87.330	36	<.001

[h]The test for homogeneity of effect sizes indicates that effects are not consistent across the set of studies. Thus, the fixed effect mean test reported in this portion of Table 2 should be ignored in favor of the mean test reported in the second half of the table under the Random Effects model.

Table 2 (cont'd). Average Effect Sizes and Variance Components for Seven Accommodations Used in Randomized Experiments

Accommodation	Results for Random Effects Analysis									
	Number of Samples	Effect Size and 95% Confidence Interval				Test of Mean Effect = 0		Test of Heterogeneity in Effect Sizes		
		Mean Effect Size	s.e.	Lower Limit	Upper Limit	Z	p	Q	df(Q)	p(Q)
Bilingual Dictionary-Glossary	5	-.039	.131	-.285	.217	-.298	.766			
Dual Language Booklet	1	-.177	.148	-.467	.112	-1.199	.231			
Dual Language Questions + Read Aloud in Spanish	1	.273	.195	-.109	.654	1.401	.161			
English Dictionary-Glossary	11	.178	.055	.070	.287	3.232	.001			
Extra Time	2	.209	.142	-.069	.488	1.473	.141			
Simplified English	15	.018	.061	-.102	.138	0.292	.771			
Spanish Version	2	.302	.719	-1.107	1.711	.420	.674			
TOTAL WITHIN										
TOTAL BETWEEN								9.864	6	<.131
OVERALL MEAN	37	.092	.036	.021	.162	2.550	.011			

REFERENCES

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), pp. 4-14.

Abedi, J., Courtney, M, & Leon, S. (2003a). *Effectiveness and validity of accommodations for English language learners in large-scale assessments* (CSE Technical Report 608). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Courtney, M., & Leon, S. (2003b). *Research-supported accommodation for English language learners in NAEP* (CES Technical Report 586). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

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

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001, February). *NAEP math performance test accommodations: Interactions with student language background* (CSE Technical Report 536). Los Angeles, CA: 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., and Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.

Abedi, J., Lord, C., Boscardin, C. K., & Miyoshi, J. (2001, September). *The effects of accommodations on the assessment of Limited English Proficient (LEP) students in the National Assessment of Educational Progress (NAEP),* (Working Paper, Publication No. NCES 200113). Washington, DC: National Center for Education Statistics.

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance.* Los Angeles, CA: UCLA Center for the Study of Evaluation/ 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 Measurements: Issues and Practice, 19*(3), 16-26.

Abedi, J., Lord C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance* (CSE Technical Report #429). Los Angeles, CA: Center for the Study of Evaluation.

Albus, A., Bielinski, J.,Thurlow, M., and Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test (LEP Project Report 1).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved July 21, 2006 from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/LEP1.html.

Albus, A., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research, 98*(4), 245-254.

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., and Bielinski, J. (2000). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase 2* (Minnesota Report No. 31). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved July 21, 2006 from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/MnReport31.html

August, D. L., & Hakuta, K. (1997). *Improving schooling for language-minority learners.* Washington, DC: National Academies Press.

August, D.L. & Siegel, L.S. (2006). Literacy instruction for language-minority children in special education settings. In D. L. August & T. Shanahan (Eds.), *Developing Literacy in a second language: Report of the National Literacy Panel.* Mahwah, NJ: Lawrence Erlbaum Associates.

Biancarosa, G., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report from the Carnegie Corporation of New York* (2nd ed). Washington, DC: Alliance for Excellence in Education.

Brown, P. (1999). *Findings of the 1999 Plain Language Field Test* (Publication T99-013.1). University of Delaware, Delaware Education Research & Development Center.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act.* Washington, DC: The Urban Institute.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E., (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39,* 188-215.

Cooper, H. (1998). *Synthesizing Research* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Cooper, H. & Hedges, L.V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly, 34*(2): 213-238.

Dressler, C. (2006). First- and second-language literacy. In D. L. August & T. Shanahan (Eds.), *Developing Literacy in a second language: Report of the National Literacy Panel.* Mahwah, NJ: Lawrence Erlbaum Associates.

Francis, D. J.; Snow, C. E.; August, D.; Carlson, C. D.; Miller, J.; Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the Diagnostic Assessment of Reading Comprehension. *Scientific Studies of Reading, 10*(3), 301-322.

Fuhrman, S. H. (2003). Riding waves, trading horses: The twenty-year effort to reform education. In D.T. Gordon (Ed.), *A nation reformed? American education 20 years after A Nation at Risk* (pp. 7-22). Cambridge, MA: Harvard Education Press.

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

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Hedges, L.V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.

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

Johnson, E., & Monroe, B. (2004). Simplified language as an accommodation on math tests. *Assessment for Effective Intervention, 29*(3), 35-45.

Kieffer, M.J. & Lesaux, N. K. (in press). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher.*

Koenig, J. A., & Bachman, L. F. (2004). *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments.* National Research Council, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Lipsey, M.W. & Wilson, D.B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage Publications.

Lyon, G. (1995). Toward a definition of dyslexia. *Annals of Dyslexia, 451,* 3-27.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53,* 1-14.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19,* 304-330.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research Vol. 3* (pp. 269-284). New York: Longman.

National Assessment of Educational Progress. (2006). *Reading assessments.* Washington DC: U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2004). *Language minority learners and their labor market Indicators – Recent trends.* Washington, DC: U.S. Department of Education. Retrieved September 21, 2004 from http://nces.ed.gov/pubs2004/2004009.pdf

National Center for Education Statistics. (2005a). *Nation's report card for math.* Washington, DC: U.S. Department of Education, Institute of Educational Sciences.

National Center for Educational Statistics (2005b). *Nation's report card for reading.* Washington, D.C.: U.S. Department of Education, Institute of Educational Sciences.

National Center for Education Statistics. (2006). *National Assessment of Educational Progress, 2006, reading assessments.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Institute of Child Health and Human Development (2003). *National Symposium on Learning Disabilities and English Language Learners* (Symposium summary). Washington, DC: U.S. Department of Education and the National Institute of Child Health and Human Development.

National Research Council. (2004). *Keeping score for all.* Washington, DC: National Academies Press.

Pennock-Román, M. (1990). *Test validity and language background: A study of Hispanic American students at six universities.* New York: The College Board.

Pennock-Román, M. (1992). Interpreting test performance in selective admissions for Hispanic students. In K. Geisinger (Ed.), *Psychological testing of Hispanics* (pp. 99-135). Washington, DC: American Psychological Association.

Pennock-Román, M. (1993). The status of research on the Scholastic Aptitude test (SAT) and Hispanic students in post-secondary education. In B. R. Gifford (Ed.), *Policy perspectives on educational testing* (pp. 75-115). Boston: Kluwer Academic Press.

Pennock-Román, M. (2002). Relative effects of English proficiency on general admissions tests versus subject tests. *Research in Higher Education, 43*(5), 601-623.

Pennock-Román, M. (2006). Language and cultural issues in the educational measurement of Latinos. In Lourdes Diaz Soto (Ed.), *The Praeger Handbook of Latino Education.* Portsmouth, NH: Greenwood

Population Resource Center (2001). *Executive summary: A demographic profile of Hispanics in the U.S.* Washington, DC. Retrieved August 31, 2006 from the World Wide Web: http://www.prcdc.org/summaries/hispanics/hispanics.html.

Proctor, C. P., Carlo, M., August, D., & Snow, C. E. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97*(2), 246-256.

Rivera, C., Collum, E., & Shafer Willner, L. (Eds.). (2006). *State assessment policy and practice for English language learners: A national perspective.* Mahwah, NJ: Lawrence Erlbaum Associates.

Rivera, C., Collum, E., Shafner Willner, L., & Sia, J.K. (2006). An analysis of state assessment policies regarding the accommodation of English language learners. In C. Rivera and E. Collum (Eds.), *State assessment policy and practice for English language learners: A national perspective* (pp. 1-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Rivera, C., & Stansfield, C. W. (2004). The effect of linguistic simplification of science test items on score comparability. *Educational Assessment, 9* (3-4), 79-105.

Scarcella, R. (2003). *Academic English: A conceptual framework.* Los Angeles: Language Minority Research Institute.

Shepard, L., Taylor, G., & Betebenner, D. (1998). Inclusion of limited-English-proficient students in Rhode Island's grade 4 mathematics performance assessment. Los Angeles: Center for the study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Shaywitz, S.E., Fletcher, J.M., Holahan, J.M., Schneider, A.E., Marchione, K.E., Stuebing, K.K., Francis, D.J., Pugh, K.R., & Shaywitz, B.A. (1999). Persistence of dyslexia: The Connecticut longitudinal study at adolescence. *Pediatrics, 104*(6), 1351-1359.

Sireci, S., Li, S., & Scarpati, S. (2003). *The effect of test accommodation on test performance: A review of the literature* (Research Report No. 495). Amherst, MA: University of Massachusetts School of Education, Center for Educational Assessment.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academies Press.

Stahl, S. A. (1999). *Vocabulary development.* Cambridge, MA: Brookline Books.

Stahl, S. A., & Nagy, W. E. (2006). *Teaching word meanings.* Mahwah, NJ: Lawrence Erlbaum Associates.

Tabors, P., Paez, M., & Lopez, L. (2003). Dual language abilities of bilingual four-year olds: Initial findings from the Early Childhood Study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice, 1*(1), 70-91.

APPENDIX A: LITERATURE SEARCH STRATEGY

The search for studies on accommodations included a comprehensive search of online databases as well as collection of studies previously reviewed by Sireci, Li, & Scarpati (2003) and/or by Abedi, Hofstetter, & Lord (2004). The online search included a search of ERIC, PsychInfo, MLA, Education Abstracts, and Academic Search Premier using the keywords "Accommodation" and "test*" and "English language learner OR English learner OR language minority OR limited English." This search yielded 114 entries, the abstract of each of which was read to determine if it was an empirical study examining the effects of accommodations. The online database of the National Center for Research on Evaluation, Standards, and Student Testing was also searched using the keyword "accommodation" as well as an author search for "Jamal Abedi." This search produced twenty-seven entries, many of them redundant. In the online searches and collection of studies from previous reviews, published articles as well as technical reports (all of which were available online) were collected. However, several documents that were presentations at academic conferences (AERA, NCME) were not collected, due to both practical and quality concerns. The results of some of the presentations did later appear in published articles or technical reports. There were several cases in which the results of a single study were reported in multiple documents (and often cited differently in different reviews), in which case the two documents were linked together and cross-checked for complete information; the most recent document is cited.

APPENDIX B: STUDIES EXCLUDED FROM META-ANALYSIS

A handful of the empirical studies that have been included in previous qualitative reviews was excluded from the meta-analysis for various reasons involving either reporting or methodology. Abedi & Hejri (2004), Castellon-Wellington (1999), and Shepard, Taylor, and Betebenner (1998) were excluded because they examined the effects of multiple accommodations, chosen individually for students. Hafner (2001) was excluded because it did not disaggregate results by ELL and non-ELL groups, making it impossible to determine the effect of the accommodation for ELLs. Lotherington-Woloszyn (1993) was excluded because it did not report means or standard deviations, and did not provide other information that could have been used to estimate the effect size for ELLs. Miller, Okum, Sinai, & Miller (1999) was excluded because it was a presentation at the National Council on Measurement in Education (NCME) conference and was not accessible. Anderson, Jenkins, & Miller (1996) was excluded because it did not compare accommodated and non-accommodated groups.

Three studies, Abedi & Lord (2001), Albus, Thurlow, Liu, & Bielinski (2005), and Johnson & Monroe (2004) were excluded from meta-analyses of effect sizes because they employed repeated measures designs, such that all ELLs and non-ELLs were tested with and without accommodations. These studies give effect size estimates within the ELLs which are not strictly comparable to the estimates from designs where different ELLs are randomly assigned to conditions of testing with and without accommodations. Two of these studies (Abedi & Lord, 2004; Johnson & Monroe, 2004) involved Simplified English, and the other (Albus et al., 2005) involved the use of an English dictionary on a reading assessment. We do consider the findings from these well-designed studies in making our recommendations, but have excluded them from meta-analytic computations of average effect sizes and variability in effect sizes, because of the critical difference in study design, problems in reporting for at least some of these studies, and the limited number of such studies addressing any particular accommodation.

Two studies, Abedi, Lord, & Hofstetter (1998) and Hofstetter (2003) involved a common sample. Hofstetter (2003) focused on the Hispanic students who participated in the Abedi, Lord, & Hofstetter (1998) study. These students comprised roughly 2/3 of the original study sample. Because both studies

reported means, standard deviations, and sample sizes for their samples, we were able to compute means, standard deviations, and sample sizes for Abedi, Lord, & Hofstetter (1998) for the non-Hispanic portion of their sample so that the statistics reported for these two studies are non-overlapping. Aggregate results reported in the text of Hofstetter (2003) were used to produce statistics for the non-Hispanic sample in Abedi, Lord, & Hofstetter (1998). However, we use the means and standard deviations reported in Table 3 on page 172 of Hofstetter (2003) as the raw statistics for the meta-analysis for this study. In Table 3, Hofstetter (2003) provides means and standard deviations for LEP students broken down by language of instruction. Because assignment in Hofstetter (2003) was random "within-classroom," this allowed us to examine the effects of Simplified English and Spanish version tests separately for Hispanic students receiving English instruction and those receiving Spanish instruction. This distinction is especially important for the Spanish version accommodation. Abedi, Lord, and Hofstetter (1998) report effects for Spanish version accommodation that is different from Hofstetter (2003) and is based on a sample size that involves 15 more subjects than Hofstetter (2003). However, since Hofstetter (2003) involves only the Hispanic students from Abedi, Lord, and Hofstetter (1998), we used the results from Hofstetter (2003) for the Spanish version accommodation and dropped results for that accommodation from Abedi, Lord, and Hofstetter (1998) since it is redundant with Hofstetter (2003). Thus, we report four effect size estimates from Hofstetter (2003): Simplified English for Hispanic LEP students receiving Spanish language instruction, Simplified English for Hispanic LEP students receiving English language instruction, Spanish language assessment for Hispanic LEP students receiving Spanish language instruction, and Spanish language assessment for Hispanic LEP students receiving English language instruction. In the Table in Appendix D, Samples 1 and 2 are Spanish-instructed students, while Samples 3 and 4 are English-instructed students. In addition, we report one effect size from Abedi, Lord, & Hofstetter (1998), namely the effect of Simplified English for non-Hispanic LEP students. Additional information on how the effect size for non-Hispanic students in Abedi, Lord, and Hofstetter (1998) was computed using information from Hofstetter (2003) is available from the authors on request.

APPENDIX C: OVERVIEW OF META-ANALYSIS METHODS

To evaluate the effectiveness of different types of accommodations for ELLs, we conducted a meta-analysis of the results from the 11 randomized experiments that met the inclusion criteria. To conduct this meta-analysis, we first had to resolve three methodological issues. First, we made a choice between two distinct but related options for the measure of effectiveness. One option was to conceptualize the effect of the accommodation in these randomized studies as the difference in the effect of the accommodation on ELLs and non-ELLs, i.e., as the degree to which the effect of the accommodation for ELLs was different from the effect of the accommodation for non-ELLs in the study. In statistical terms, an effective accommodation would produce a significant interaction between ELL status and the accommodation. A second option, which is more commonly used in the accommodation literature, is to conceptualize the effect of the accommodation for ELLs alone, i.e., the difference in test performance between ELLs taking the accommodated test and ELLs taking the test without accommodations. Then, the effect of the accommodation for non-ELLs assesses the validity of the accommodation as a second question. A valid accommodation would have no statistically significant effect on the test scores of the non-ELLs. We opted for this latter, two part conceptualization of the effect of the accommodations because it was consistent with the research literature and with a straight-forward process of finding suitable accommodations that are both effective (i.e., have an effect on the test scores of ELLs who need the accommodation) and valid (i.e., do not have an effect for non-ELLs who do not need the accommodation).

Given this conceptualization of the question of effectiveness, a second methodological issue is the choice of effect size statistics. As our measure of effect size we first computed the mean difference in performance between ELLs receiving the accommodated test and ELLs taking the test without accommodations. This difference in mean performance was then standardized using the pooled within-groups estimate of the standard deviation. This measure of effect size is the common Cohen's *d*, which is known to be biased in small samples. We then corrected this measure of effect size using a transformation of *d* recommended by Hedges (1981). The resulting effect size estimates are termed Hedges's g^{u} and were computed directly from the means

and standard deviations reported in the studies by using a programmed routine in the *Comprehensive Meta-Analysis (Version 2)* (Borenstein, 2006) software, which was also used to conduct the meta-analysis. Thus, we measure the effect of the accommodation as the mean difference between ELLs receiving the accommodation and those taking the test under standard conditions, and express this difference relative to the standard deviation, and adjust this measure to control for bias in small samples.

Appendix D provides a table with the results from each study, including the means and standard deviations for the ELLs in both testing conditions and the measure of effect size g^u, along with a measure of its standard error. Also included in Appendix D is tabular information on the grade level of the student participants, the nature of the accommodations, whether other accommodations were also used, the content area of the assessment, and the nature of the outcome measure. For all studies, positive values of g^u indicate that ELLs taking the accommodated test scored higher than ELLs taking the test without accommodations. Negative values of g^u indicate the ELLs taking the test without accommodations scored higher than ELLs taking the test with accommodations.

Two studies, Abedi, Lord, & Hofstetter (1998) and Hofstetter (2003) involved a common sample. Hofstetter (2003) focused on the Hispanic students who participated in the Abedi, Lord, & Hofstetter (1998) study. These students comprised roughly 2/3 of the original study sample. Because both studies reported means, standard deviations, and sample sizes for their samples, we were able to compute means, standard deviations, and sample sizes for Abedi, Lord, & Hofstetter (1998) for the non-Hispanic portion of their sample so that the statistics reported for these two studies are non-overlapping. Aggregate results reported in the text of Hofstetter (2003) were used to produce statistics for the non-Hispanic sample in Abedi, Lord, & Hofstetter (1998). However, we use the means and standard deviations reported in Table 3 on page 172 of Hofstetter (2003) as the raw statistics for the meta-analysis for this study.

In Table 3, Hofstetter (2003) provides means and standard deviations for LEP students broken down by language of instruction. Because assignment in Hofstetter (2003) was random "within-classroom," this allowed us to examine the effects of Simplified English and Spanish version tests separately for Hispanic students receiving English instruction and those receiving Spanish instruction. This distinction is especially important for the Spanish version

accommodation. Abedi, Lord, and Hofstetter (1998) report effects for Spanish version accommodation that is different from Hofstetter (2003) and is based on a sample size that involves 15 more subjects than Hofstetter (2003). However, since Hofstetter (2003) involves only the Hispanic students from Abedi, Lord, and Hofstetter (1998), we used the results from Hofstetter (2003) for the Spanish version accommodation and dropped results for that accommodation from Abedi, Lord, and Hofstetter (1998) since it is redundant with Hofstetter (2003). Thus, we report four effect size estimates from Hofstetter (2003): Simplified English for Hispanic LEP students receiving Spanish language instruction, Simplified English for Hispanic LEP students receiving English language instruction, Spanish language assessment for Hispanic LEP students receiving Spanish language instruction, and Spanish language assessment for Hispanic LEP students receiving English language instruction. In the Table in Appendix D, Samples 1 and 2 are Spanish-instructed students, while Samples 3 and 4 are English-instructed students. In addition, we report one effect size from Abedi, Lord, & Hofstetter (1998), namely the effect of Simplified English for non-Hispanic LEP students. Additional information on how the effect size for Abedi, Lord, and Hofstetter (1998) was computed using information from Hofstetter (2003) is available from the authors on request.

A third methodological issue is the choice of the unit of analysis. Because many of the studies examined the effects of multiple accommodations or the effects of accommodations on students at two different grade levels, we had to choose between using these samples within studies or the studies themselves as the unit of analysis. We chose to use the sample as the unit of analysis because doing so preserved the maximum amount of information in the collection of studies about different accommodations, in different grades, and for different content areas. The alternate strategy of treating the study as the unit of analysis would have required that we average across the effects of different accommodations (as well as across grades and content areas), even though the samples were independent, at least to an extent. It is worth noting that in some studies, a single control group (i.e., ELLs taking the test without accommodations) was compared to more than one treatment (i.e., accommodated ELL group), rendering some comparisons within a study dependent on one another. Because these different comparisons involving the control group addressed questions about different accommodations in our analysis, this dependence would serve to increase the correlation between

findings across different sets of accommodations. Nevertheless, on balance, we felt that this drawback was worth the added information gained by using the sample as the unit of analysis at this stage of our investigation into the effectiveness of different accommodations.

APPENDIX D: DESCRIPTIVE INFORMATION AND EFFECT SIZE CALCULATIONS FOR 11 STUDIES USED IN META-ANALYSIS

Accommodation	Study name	Study Number	Sample Number	ELL–Accommodations			ELL–No Accommodations			Hedges's g^U	Std Err
				Mean	SD	N	Mean	SD	N		
Bilingual Dictionary-Glossary	Abedi, Courtney & Leon #608 (2003a)	2	1	45.62	8.19	135	48.23	9.38	268	-0.289	0.106
	Abedi, Courtney & Leon #608 (2003a)	2	2	44.58	8.38	119	45.73	9.41	199	-0.127	0.116
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	1	11.72	3.73	64	10.04	3.66	62	0.452	0.179
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	2	9.38	2.69	16	10.32	3.99	22	-0.262	0.323
	Abedi, Lord, Boscardin, & Miyoshi (2001)	5	1	8.51	4.72	70	8.36	4.4	58	0.033	0.177
Dual Language Booklet	Garcia Duncan et al. (2005)	11	1	30.65	11.74	74	32.92	13.36	119	-0.177	0.148
Dual Language Questions + Read Aloud in Spanish	Anderson et al. (2000)	8	1	17.7	7.31	53	15.85	6.09	52	0.273	0.195
English Dictionary - Glossary	Abedi, Courtney & Leon #586 (2003b)	1	1	13.81	6.043	64	12.27	5.242	80	0.273	0.168
	Abedi, Courtney & Leon #586 (2003b)	1	2	9.95	3.835	86	9.41	4.005	86	0.137	0.152
	Abedi, Courtney & Leon #586 (2003b)	1	3	14.69	5.115	35	12.27	5.242	80	0.462	0.204
	Abedi, Courtney & Leon #586 (2003b)	1	4	10.17	4.361	84	9.47	4.005	86	0.167	0.153
	Abedi, Courtney & Leon #608 (2003a)	2	3	48.37	9.75	270	48.23	9.38	268	0.015	0.086
	Abedi, Courtney & Leon #608 (2003a)	2	4	46.68	9	206	45.73	9.41	199	0.103	0.099
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	3	11.97	3.47	59	10.04	3.66	62	0.537	0.184
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	4	11.52	3.53	23	10.32	3.99	22	0.313	0.295
	Abedi, Hofstetter, Baker, & Lord (2001)	4	1	11.84	5.94	146	12.07	5.47	144	-0.040	0.117
	Abedi, Hofstetter, Baker, & Lord (2001)	4	2	13.69	6.74	29	12.07	5.47	144	0.283	0.203
	Abedi, Lord, Boscardin, & Miyoshi (2001)	5	2	10.18	5.26	55	8.36	4.4	58	0.374	0.189
Extra Time	Abedi, Courtney & Leon #586 (2003b)	1	5	13.74	6.024	35	12.27	5.242	80	0.266	0.202
	Abedi, Hofstetter, Baker, & Lord (2001)	4	4	12.93	5.99	30	12.07	5.47	144	0.154	0.200
Simplified English	Abedi, Courtney & Leon #608 (2003a)	2	5	47.36	9.48	284	48.23	9.38	268	-0.092	0.085
	Abedi, Courtney & Leon #608 (2003a)	2	6	47.63	9.53	209	45.73	9.41	199	0.200	0.099
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	5	10.55	3.37	20	10.04	3.66	62	0.141	0.255
	Abedi, Hofstetter, Baker, & Lord (2001)	4	3	12.63	5.23	124	12.07	5.47	144	0.104	0.122
	Rivera & Stansfield (2003)	6	1	4.33	2.52	18	4.67	1.91	15	-0.146	0.342
	Rivera & Stansfield (2003)	6	2	4.38	1.71	16	3.48	1.89	23	0.485	0.324
	Rivera & Stansfield (2003)	6	3	2.11	1.27	9	4	1.5	9	-1.295	0.498
	Rivera & Stansfield (2003)	6	4	2	1.79	6	3.23	2.45	13	-0.516	0.479
	Hofstetter (2003)	7	1	5.50	2.73	6	5.00	3.50	9	0.146	0.497
	Hofstetter (2003)	7	3	11.49	5.40	222	11.32	4.900	229	0.033	0.094
	Abedi, Lord, & Hofstetter (1998)	9	1	12.69	6.03	117	13.41	6.14	115	-.118	0.131
	Brown (1999)	10	1	16.5	6.58	16	16	5.02	13	0.082	0.363
	Brown (1999)	10	2	14	3.37	4	10.5	6.36	2	0.649	0.718
	Brown (1999)	10	3	29.22	12.47	18	33.55	14.47	11	-0.318	0.374
	Brown (1999)	10	4	19.25	13.87	4	9.5	2.12	2	0.647	0.718
Spanish Version	Hofstetter (2003)	7	2	8.68	3.41	63	5.00	3.50	9	1.064	.364
	Hofstetter (2003)	7	4	9.64	3.66	147	11.32	4.90	229	-0.376	.106

APPENDIX D: DESCRIPTIVE INFORMATION AND EFFECT SIZE CALCULATIONS FOR 11 STUDIES USED IN META-ANALYSIS (CONTINUED)

Accommodation	Study name	Study Number	Sample Number	Test Format	Grade	Domain	Test	Bundled with Extra Time
Bilingual Dictionary-Glossary	Abedi, Courtney & Leon #608 (2003a)	2	1	Paper&Pencil	4	Science	NAEP & TIMSS	YES
	Abedi, Courtney & Leon #608 (2003a)	2	2	Paper&Pencil	8	Science	NAEP & TIMSS	YES
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	1	Paper&Pencil	4	Science	NAEP	See Note
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	2	Paper&Pencil	8	Science	NAEP	See Note
	Abedi, Lord, Boscardin, & Miyoshi (2001)	5	1	Paper&Pencil	8	Science	NAEP	
Dual Language Booklet	Garcia Duncan et al. (2005)	11	1	Paper&Pencil	8	Math	NAEP items	
Dual Language Questions + Read Aloud in Spanish	Anderson et al. (2000)	8	1	Paper&Pencil	8	Reading	Minnesota Basic Standards Test	
English Dictionary - Glossary	Abedi, Courtney & Leon #586 (2003b)	1	1	Paper&Pencil	4	Math	NAEP	
	Abedi, Courtney & Leon #586 (2003b)	1	2	Paper&Pencil	8	Math	NAEP	
	Abedi, Courtney & Leon #586 (2003b)	1	3	Computerized	4	Math	NAEP	
	Abedi, Courtney & Leon #586 (2003b)	1	4	Computerized	8	Math	NAEP	
	Abedi, Courtney & Leon #608 (2003a)	2	3	Paper&Pencil	4	Science	NAEP & TIMSS	YES
	Abedi, Courtney & Leon #608 (2003a)	2	4	Paper&Pencil	8	Science	NAEP & TIMSS	YES
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	3	Paper&Pencil	4	Science	NAEP	See Note
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	4	Paper&Pencil	8	Science	NAEP	See Note
	Abedi, Hofstetter, Baker, & Lord (2001)	4	1	Paper&Pencil	8	Math	NAEP	
	Abedi, Hofstetter, Baker, & Lord (2001)	4	2	Paper&Pencil	8	Math	NAEP	YES
	Abedi, Lord, Boscardin, & Miyoshi (2001)	5	2	Paper&Pencil	8	Science	NAEP	
Extra Time	Abedi, Courtney & Leon #586 (2003b)	1	5	Paper&Pencil	4	Math	NAEP	YES
	Abedi, Hofstetter, Baker, & Lord (2001)	4	4	Paper&Pencil	8	Math	NAEP	YES
Simplified English	Abedi, Courtney & Leon #608 (2003a)	2	5	Paper&Pencil	4	Science	NAEP & TIMSS	YES
	Abedi, Courtney & Leon #608 (2003a)	2	6	Paper&Pencil	8	Science	NAEP & TIMSS	YES
	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3	5	Paper&Pencil	4	Science	NAEP	See Note
	Abedi, Hofstetter, Baker, & Lord (2001)	4	3	Paper&Pencil	8	Math	NAEP	
	Rivera & Stansfield (2003)	6	1	Paper&Pencil	4	Science	Delaware Science (C & E)	
	Rivera & Stansfield (2003)	6	2	Paper&Pencil	4	Science	Delaware Science (D & F)	
	Rivera & Stansfield (2003)	6	3	Paper&Pencil	6	Science	Delaware Science (C & E)	
	Rivera & Stansfield (2003)	6	4	Paper&Pencil	6	Science	Delaware Science (D & F)	
	Hofstetter (2003)	7	1	Paper&Pencil	8	Math	NAEP	
	Hofstetter (2003)	7	3	Paper&Pencil	8	Math	NAEP	
	Abedi, Lord, & Hofstetter (1998)	9	1	Paper&Pencil	8	Math	NAEP	
	Brown (1999)	10	1	Paper&Pencil	5	Science	Delaware State Test	
	Brown (1999)	10	2	Paper&Pencil	8	Science	Delaware State Test	
	Brown (1999)	10	3	Paper&Pencil	5	Math	Delaware State Test	
	Brown (1999)	10	4	Paper&Pencil	8	Math	Delaware State Test	
Spanish Version	Hofstetter (2003)	7	2	Paper&Pencil	8	Math	NAEP	
	Hofstetter (2003)	7	4	Paper&Pencil	8	Math	NAEP	

Note: We have coded Abedi, Courtney, Mirocha, Leon, & Goldberg (2005) as not involving extra time because extra time was not unique to one or more of the accommodated conditions, but rather was granted to students in all conditions, including the "unaccommodated" condition

APPENDIX E: FORREST PLOT OF EFFECT SIZES AND 95% CONFIDENCE INTERVALS FROM RANDOM EFFECTS MODEL

Group by Accommodation	Study name	Sample Number	Statistics for each study						
			Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
BilingualDict-Glossary	Abedi, Courtney & Leon #608 (2003)	1.000	-0.289	0.106	0.011	-0.497	-0.082	-2.735	0.006
BilingualDict-Glossary	Abedi, Courtney & Leon #608 (2003)	2.000	-0.127	0.116	0.013	-0.354	0.100	-1.097	0.273
BilingualDict-Glossary	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	1.000	0.452	0.179	0.032	0.100	0.803	2.519	0.012
BilingualDict-Glossary	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	2.000	-0.262	0.323	0.104	-0.896	0.371	-0.812	0.417
BilingualDict-Glossary	Abedi, Lord, Boscardin, & Miyoshi (2001)	1.000	0.033	0.177	0.031	-0.313	0.379	0.185	0.854
BilingualDict-Glossary			-0.039	0.131	0.017	-0.295	0.217	-0.298	0.766
DualLanguageBooklet	Garcia Ducan et al. (2005)	1.000	-0.177	0.148	0.022	-0.467	0.112	-1.199	0.231
DualLanguageBooklet			-0.177	0.148	0.022	-0.467	0.112	-1.199	0.231
DualLanguageQuestions+ReadAloudinSpanish	Anderson et al. (2000)	1.000	0.273	0.195	0.038	-0.109	0.654	1.401	0.161
DualLanguageQuestions+ReadAloudinSpanish			0.273	0.195	0.038	-0.109	0.654	1.401	0.161
EnglishDict-Glossary	Abedi, Courtney & Leon #586 (2003)	1.000	0.273	0.168	0.028	-0.055	0.601	1.629	0.103
EnglishDict-Glossary	Abedi, Courtney & Leon #586 (2003)	2.000	0.137	0.152	0.023	-0.161	0.435	0.902	0.367
EnglishDict-Glossary	Abedi, Courtney & Leon #586 (2003)	3.000	0.462	0.204	0.041	0.063	0.861	2.269	0.023
EnglishDict-Glossary	Abedi, Courtney & Leon #586 (2003)	4.000	0.167	0.153	0.023	-0.133	0.466	1.089	0.276
EnglishDict-Glossary	Abedi, Courtney & Leon #608 (2003)	3.000	0.015	0.086	0.007	-0.154	0.183	0.170	0.865
EnglishDict-Glossary	Abedi, Courtney & Leon #608 (2003)	4.000	0.103	0.099	0.010	-0.092	0.298	1.038	0.299
EnglishDict-Glossary	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	3.000	0.537	0.184	0.034	0.177	0.898	2.921	0.003
EnglishDict-Glossary	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	4.000	0.313	0.295	0.087	-0.264	0.891	1.063	0.288
EnglishDict-Glossary	Abedi, Hofstetter, Baker, & Lord (2001)	1.000	-0.040	0.117	0.014	-0.270	0.189	-0.343	0.732
EnglishDict-Glossary	Abedi, Hofstetter, Baker, & Lord (2001)	2.000	0.283	0.203	0.041	-0.115	0.681	1.393	0.164
EnglishDict-Glossary	Abedi, Lord, Boscardin, & Miyoshi (2001)	2.000	0.374	0.189	0.036	0.004	0.743	1.982	0.048
EnglishDict-Glossary			0.178	0.055	0.003	0.070	0.287	3.232	0.001
ExtraTime	Abedi, Courtney & Leon #586 (2003)	5.000	0.266	0.202	0.041	-0.130	0.662	1.316	0.188
ExtraTime	Abedi, Hofstetter, Baker, & Lord (2001)	4.000	0.154	0.200	0.040	-0.238	0.546	0.770	0.441
ExtraTime			0.209	0.142	0.020	-0.069	0.488	1.473	0.141
SimplifiedEnglish	Abedi, Courtney & Leon #608 (2003)	5.000	-0.092	0.085	0.007	-0.259	0.075	-1.083	0.279
SimplifiedEnglish	Abedi, Courtney & Leon #608 (2003)	6.000	0.200	0.099	0.010	0.006	0.394	2.020	0.043
SimplifiedEnglish	Abedi, Courtney, Mirocha, Leon, & Goldberg (2005)	5.000	0.141	0.255	0.065	-0.359	0.640	0.551	0.581
SimplifiedEnglish	Abedi, Hofstetter, Baker, & Lord (2001)	3.000	0.104	0.122	0.015	-0.135	0.344	0.852	0.394
SimplifiedEnglish	Abedi, Lord, & Hofstetter (1998)	1.000	-0.118	0.131	0.017	-0.375	0.139	-0.900	0.368
SimplifiedEnglish	Brown (1999)	1.000	0.082	0.363	0.132	-0.630	0.793	0.225	0.822
SimplifiedEnglish	Brown (1999)	2.000	0.649	0.718	0.515	-0.758	2.055	0.904	0.366
SimplifiedEnglish	Brown (1999)	3.000	-0.318	0.374	0.140	-1.051	0.416	-0.849	0.396
SimplifiedEnglish	Brown (1999)	4.000	0.647	0.718	0.515	-0.760	2.053	0.901	0.367
SimplifiedEnglish	Hofstetter (2003)	1.000	0.146	0.497	0.247	-0.828	1.120	0.294	0.769
SimplifiedEnglish	Hofstetter (2003)	3.000	0.033	0.094	0.009	-0.151	0.217	0.350	0.726
SimplifiedEnglish	Rivera & Stansfield (2003)	1.000	-0.146	0.342	0.117	-0.816	0.523	-0.429	0.668
SimplifiedEnglish	Rivera & Stansfield (2003)	2.000	0.485	0.324	0.105	-0.150	1.119	1.498	0.134
SimplifiedEnglish	Rivera & Stansfield (2003)	3.000	-1.295	0.498	0.248	-2.272	-0.319	-2.600	0.009
SimplifiedEnglish	Rivera & Stansfield (2003)	4.000	-0.516	0.479	0.229	-1.455	0.422	-1.078	0.281
SimplifiedEnglish			0.018	0.061	0.004	-0.102	0.138	0.292	0.771
SpanishVersion	Hofstetter (2003)	2.000	1.064	0.364	0.132	0.352	1.777	2.928	0.003
SpanishVersion	Hofstetter (2003)	4.000	-0.376	0.106	0.011	-0.585	-0.168	-3.537	0.000
SpanishVersion			0.302	0.719	0.517	-1.107	1.711	0.420	0.674
Overall			0.092	0.036	0.001	0.021	0.162	2.550	0.011

ENDNOTES

[1] For documents that outline the demographics of this population, including its size, see NCES (2004); Capps, Fix, Murray, Ost, Passel, & Herwantoro (2005); Population Resource Center (2000).

[2] Biancarosa & Snow (2006); Kieffer & Lesaux, in press; Carlo et al. (2004); Proctor, Carlo, August & Snow (2005); Tabors, Paez, & Lopez (2003); Francis et al. (2006).

[3] August & Hakuta (1997); Biancarosa & Snow (2004); NCES (2005a, 2005b).

[4] NCES (2005a).

[5] For research on the prevalence and definition of learning disabilities in native English speakers see Lyon (1995); Lyon, Shaywitz, & Shaywitz (2003); Shaywitz et al. (1999); for a review of the research on learning disabilities in language minority learners see Lesaux (2006). For a discussion of the difficulties in and the need for increased opportunities to learn for ELLs to prevent and reduce reading difficulties see NICHD (2003); Snow, Burns, & Griffin (1998). For a review of research on literacy instruction for ELLs in special education see August & Siegel (2006).

[6] Texas reported performance on the 2002 state accountability assessment in English Reading for ELL students as a function of their scores on the Reading Proficiency Test in English (RPTE). The RPTE is designed to assess proficiency in English and is used to indicate when students are ready to take the state accountability test in English. The study found that 15.8% of students passed the English reading test if they scored at the Beginning level on the RPTE in 2002. This percent passing compared to 30.4% for Intermediates, 76.4% for students who scored Advanced in 2002, and 89.6% for students who scored Advanced in 2000. Similar results were found at each grade from 3 through 10, although some differences are noted between the early and later grades. Results can be found at http://www.tea.state.tx.us/student.assessment/reporting/results/rpteanalysis/2002/reading/statewide.html. In a study of students who first entered Grade 9 in 1996, the New York State Education Agency found that 32.6% of current ELLs graduated high school in four years, while 60.1% of former ELLs graduated high school in four years, as compared to 54.5% of students who had never been ELLs. These percentages increased to 49.5%, 76.5%, and 70.5% at seven years. Thus, while former ELLs are completing high school at rates comparable to non-ELL students, it's clear that many ELL students are still not successful. For the complete report see: http://www.regents.nysed.gov/2005Meetings/March2005/0305emscvesidd4.html. Both reports were last accessed by the authors on September 28, 2006 in preparing this report.

[7] See the introduction to Rivera, Collum, & Shafer Willner (2006) for an overview of the history of practices relating to the participation of ELLs in state assessment programs.

[8] For a review of the relationship between first and second language literacy processes see Dressler (2006).

[9] For a discussion of academic language see Scarcella (2003), and of reading vocabulary see Nagy & Anderson (1984); Nagy & Scott (2000); Stahl (1999); Stahl & Nagy (2006). Readers may also wish to consult the Academic Word List website at www.vuw.ac.nz/lals/research/awl/awlinfo.html and references on the development of the Academic Word List in Coxhead (2000).

[10] NCES (2004).

[11] See Rivera, Collum, & Shafer Willner (2006).

[12] See Fuhrman (2003).

[13] Abedi, Lord, Hofstetter, & Baker (2000); Abedi, Lord & Hofstetter (1998), Abedi, Lord & Plummer (1997); Pennock-Román, M. (1990; 1992; 2002; 2006).

[14] See Rivera, Collum, Shafer Willner, & Sia (2006) for a taxonomy of accommodations for ELLs.

[15] In the past, a disability framework has guided the choice of accommodations for ELLs. Rivera, Collum, Shafer Willner, & Sia (2006) discuss in more detail why the appropriate framework for accommodations for ELLs addresses their linguistic needs, and why the traditional disability framework is not appropriate for these students. The vestiges of the disability framework can still be seen in the policies and recommended accommodations of some states, but this framework does not address the needs of ELL students. Rivera et al. (2006) argue convincingly that a more appropriate framework for thinking about the needs of ELLs comes from research on second language acquisition. That research shows us how students process linguistic information in a second language and shows the

importance of linguistic simplification, repetition, and clarification in negotiating meaning in language exchanges. See Rivera et al. (2006), pp. 22-24.

[16] Abedi, Lord, & Hofstetter (1998); Hofstetter (2003).

[17] The information compiled in this section is taken from several sources. The most comprehensive and recent study of state policies regarding accommodations for ELLs is Rivera, Collum, and Shafer Willner's (2006) edited volume entitled *State Assessment Policy and Practice for English Language Learners: A National Perspective*. In addition to this volume, we examined the National Research Council's 2004 report entitled *Keeping Score for All* (see Koenig & Bachman [2004]). Although both of these volumes are very recent, we also searched the websites of all 50 states for available documents regarding current state policy and practice. Rivera and colleagues (2006) examined a variety of source documents from states, including state websites as well as documents and survey data solicited directly from states. However, due to the time lag involved in processing the data and getting to publication, the authors indicate that their findings reflect state policy and practice as of 2002. Although our review of state policies is much less extensive than Rivera and Collum's and draws on their excellent work and that of the National Research Council, our tabled information about state policies reflects information taken from state websites in the current year.

[18] See Rivera, Collum, Shafer Willner, & Sia (2006) for tables listing 75 accommodations in use by states, 44 of which are deemed by these authors to be minimally responsive to the needs of ELLs.

[19] See Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). Because extra time was given to students in the control conditions as well as to students in the control conditions, this study does not provide a test of the effects of extra time plus another accommodation, nor does it provide an explicit test of extra time. Rather the study estimates the effects of the studied conditions over and above any effects of extra time. Consequently, we have coded this study as not involving extra time because extra time was not unique to one or more of the accommodated conditions.

[20] *Simplified English:* According to previous reports, the results of the small body of research using Simplified English are divided regarding the validity and effectiveness of making linguistic modifications to test items. Specifically, according to authors of individual articles and previous, narrative reviews of this research, this accommodation has been reported valid and/or effective for some grades, but not for all in content area tests, such as math and science. See Abedi, J. Courtney, M, & Leon, S. (2003a), Abedi, Courtney, Mirocha, Leon, & Goldberg (2005); Abedi, Hofstetter, Baker, & Lord (2001); Abedi & Lord (2001); Abedi, J., Lord, C., & Hofstetter, C. (1998); Albus, A., Bielinski, J.,Thurlow, M., and Liu, K. (2001); Brown, P. (1999); Hofstetter (2003); Rivera & Stansfield (2004).

Abedi, J. Courtney, M, & Leon, S. (2003a). *Effectiveness and validity of accommodations for English language learners in large-scale assessments* (CSE Technical Report 608). Los Angeles: CA. National Center for Research on Evaluation, Standards, and Student Testing,.

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

Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001, February). *NAEP math performance test accommodations: Interactions with student language background* (CSE Technical Report 536). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance.* Los Angeles, CA: Center for the Study of Evaluation/ National Center for Research on Evaluation, Standards, and Student Testing.

Albus, A., Bielinski, J.,Thurlow, M., and Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Project Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved July 21, 2006 from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/LEP1.html

Brown, P. (1999). *Findings of the 1999 Plain Language Field Test* (Publication T99-013.1). University of Delaware, Delaware Education Research & Development Center.

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

Rivera, C., & Stansfield, C. W. (2004). The effect of linguistic simplification of science test items on score comparability. *Educational Assessment, 9*(3-4), 79-105.

21 *Customized English Dictionaries and Glossaries:* Authors of individual studies and of previous narrative reviews have reported that the effectiveness of the use of dictionaries or glossaries may vary across grade levels and subject matter. According to individual reports, customized English dictionaries or glossaries were found valid and/or effective depending on the grade level and content area. See Abedi, Courtney, & Leon (2003a); Abedi, J., Courtney, M., & Leon, S. (2003b); Abedi, Courtney, Mirocha, Leon, & Goldberg (2005); Abedi, Hofstetter, Baker, & Lord (2001); Abedi, Lord, Boscardin, & Miyoshi (2001); Albus, A., Thurlow, M., Liu, K., & Bielinski, J. (2005).

Abedi, J. Courtney, M, & Leon, S. (2003a). *Effectiveness and validity of accommodations for English language learners in large-scale assessments* (CSE Technical Report 608). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Courtney, M., & Leon, S. (2003b). *Research-supported accommodation for English language learners in NAEP* (CES Technical Report 586). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

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

Abedi, J., Hofstetter, C., Baker, E., and Lord, C. (2001, February). *NAEP math performance test accommodations: Interactions with student language background* (CSE Technical Report 536). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Boscardin, C. K., & Miyoshi, J. (2001, September). *The effects of accommodations on the assessment of Limited English Proficient (LEP) students in the National Assessment of Educational Progress (NAEP)* (Working Paper, Publication No. NCES 200113). Washington, DC: National Center for Education Statistics.

Albus, A., Thurlow, M., Liu, K., & Bielinski, J. (2005). Reading test performance of English-language learners using an English dictionary. *The Journal of Educational Research, 98*(4), 245-254.

22 *Bilingual Dictionary and Glossary or Marginal Glosses:* Authors of individual studies have reported bilingual dictionaries, glossaries, and marginal glosses to be effective and/or valid for some grade-level science tests. See Abedi, Courtney, & Leon (2003a); Abedi, Courtney, Mirocha, Leon, & Goldberg (2005); Abedi, Lord, Boscardin, & Miyoshi (2001).

Abedi, J. Courtney, M, & Leon, S. (2003a). *Effectiveness and validity of accommodations for English language learners in large-scale assessments* (CSE Technical Report 608). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

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

Abedi, J., Lord, C., Boscardin, C. K., & Miyoshi, J. (2001, September). *The effects of accommodations on the assessment of Limited English Proficient (LEP) students in the National Assessment of Educational Progress (NAEP)* (Working Paper, Publication No. NCES 200113). Washington, DC: National Center for Education Statistics.

23 *Extra Time:* This accommodation was reported on in two independent studies (Abedi, Courtney, & Leon, 2003b; Abedi, Hofstetter, Baker, & Lord, 2001) when used in combination with other accommodations in 4th and 7th grade math tests. Several other studies bundled extra time with other accommodations (viz., bilingual dictionaries or glossaries, English dictionaries or glossaries, and simplified English). One study included extra time in all conditions, including control conditions (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005). This study is coded as not involving extra time because all students in all conditions were afforded the same time. See Appendix D for a list of these studies.

Abedi, J., Courtney, M., & Leon, S. (2003b). *Research-supported accommodation for English language learners in NAEP* (CES Technical Report 586). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Hofstetter, C., Baker, E., and Lord, C. (2001, February). *NAEP math performance test accommodations: Interactions with student language background* (CSE Technical Report 536). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

[24] *Dual Language Test Booklets and Questions:* This accommodation has been examined in 8th grade for reading and math. Authors of the two individual studies that involved dual language test booklets or questions reached different conclusions. Anderson et al. (2000) found a positive effect that was not statistically significant, but found test scores for students on the accommodated version correlated with self-rated English proficiency, whereas scores on the unaccommodated version of the test did not. This suggests that the accommodated test scores better reflected students' English language proficiency. Garcia Duncan et al. concluded that the accommodation was detrimental to student outcomes, and that effects did not vary as a function of students' English proficiency, although the test for interaction was not statistically significant at $p < .06$ (Anderson, Liu, Swierzbin, Thurlow, & Bielinski, 2000; Garcia Duncan et al., 2005) suggesting this question merits further examination. It is also important to point out that all students involved in the randomized study conditions in Garcia Duncan et al. (2005) had at least three years of English instruction, and did not differ significantly in self-rated English language proficiency from a group of native English speakers included in the study. A group of students involved in Garcia Duncan et al. (2005) with fewer than three years of English instruction was not involved in the randomized study (i.e., were not randomly assigned to the English-only test booklet), but was only given the dual-language test booklet. Based on the study design, it is impossible to say what the effect of the accommodation would have been for students with fewer than three years in English instruction.

Anderson, M., Liu, K., Swierzbin, B., Thurlow, M., and Bielinski, J. (2000). *Bilingual accommodations for limited English proficient students on statewide reading tests: Phase 2* (Minnesota Report No. 31). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved July 21, 2006 from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/MnReport31.html.

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

[25] Abedi & Lord (2001).

[26] Johnson & Monroe (2004).

[27] Rivera et al. (2006), pp. 22-24.

CENTER ON
INSTRUCTION