Mikyung Kim Wolf

Joan L. Herman

Lyle F. Bachman

Alison L. Bailey

Noelle Griffin

RECOMMENDATIONS FOR ASSESSING ENGLISH LANGUAGE LEARNERS: ENGLISH LANGUAGE PROFICIENCY MEASURES AND ACCOMMODATION USES

RECOMMENDATIONS REPORT (PART 3 OF 3)

JULY, 2008

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

Graduate School of Education & Information Studies

UCLA | University of California, Los Angeles

**Recommendations for Assessing English Language Learners:
English Language Proficiency Measures and Accommodation Uses**

CRESST Report 737

Mikyung Kim Wolf, Joan L. Herman, Lyle F. Bachman,
Alison L. Bailey, & Noelle Griffin
CRESST/University of California, Los Angeles

July 2008

RECOMMENDATIONS FOR ASSESSING ENGLISH LANGUAGE LEARNERS:

ENGLISH LANGUAGE PROFICIENCY MEASURES

AND ACCOMMODATION USES[1]

Mikyung Kim Wolf, Joan L. Herman, Lyle F. Bachman,
Alison L. Bailey, & Noelle Griffin
CRESST/University of California, Los Angeles

Abstract

The No Child Left Behind Act of 2001 (NCLB, 2002) has had a great impact on states' policies in assessing English language learner (ELL) students. The legislation requires states to develop or adopt sound assessments in order to validly measure the ELL students' English language proficiency, as well as content knowledge and skills. While states have moved rapidly to meet these requirements, they face challenges to validate their current assessment and accountability systems for ELL students, partly due to the lack of resources. Considering the significant role of assessment in guiding decisions about organizations and individuals, validity is a paramount concern. In light of this, we reviewed the current literature and policy regarding ELL assessment in order to inform practitioners of the key issues to consider in their validation process. Drawn from our review of literature and practice, we developed a set of guidelines and recommendations for practitioners to use as a resource to improve their ELL assessment systems. The present report is the last component of the series, providing recommendations for state policy and practice in assessing ELL students. It also discusses areas for future research and development.

Introduction and Background

English language learners (ELLs) are the fastest growing subgroup in the nation. Over a 10-year period between the 1994–1995 and 2004–2005 school years, the enrollment of ELL students grew over 60%, while the total K–12 growth was just over 2% (Office of English Language Acquisition [OELA], n.d.). The increased rate is more astounding for some states. For instance, North Carolina and Nevada have reported their ELL population growth rate as 500% and 200% respectively for the past 10-year period (Batlova, Fix, & Murray, 2005, as cited in Short & Fitzsimmons, 2007). Not only is the size of the ELL population is growing, but the diversity of these students is becoming more extensive. Over 400 different languages are reported among these students; schooling experience is varied depending on the students'

---

immigrant countries. The sizable ELL population typically fails to meet the proficient level in academic standards, and the academic gap between this group and the non-ELL population is considerable. The U.S. Government Accountability Office (U.S. GAO, 2006) reported that ELL students' math proficiency level averaged 20% lower than that of the overall population in 2003–2004 across 48 states. States face challenges in educating and assessing this large and varied subgroup of the U.S. student population.

The No Child Left Behind Act of 2001 (NCLB, 2002) has had a great impact on states' policies on these ELL students. The legislation makes clear that states, districts, schools, and teachers must hold the same high standards for ELL students as for all other students, and that the states should be accountable for assuring that all students, including ELL students, meet high expectations. Under NCLB, states must annually assess the progress of ELL students' English language proficiency (ELP); they also must include these students in annual assessments in content areas such as reading (or English language arts), mathematics, and science; and include their performance in the determination of each school's Adequate Yearly Progress (AYP) reporting. Although the NCLB legislation has made a significant contribution to raising awareness about the need to improve ELL students' learning and academic performance, it has also generated challenges for states to establish a valid accountability system for ELL students.

In order to meet NCLB requirements, states have moved ahead rapidly to develop needed assessments to appropriately measure their students' language proficiency, and content knowledge and skills. They have developed or adopted new measures of ELP. Many states have refined their accommodation policies for academic content assessments in order to adequately measure content knowledge and skills without these being affected by students' lack of language proficiency. However, the rush to meet NCLB assessment requirements has left states without the expertise, time, or resources to systematically document or address fundamental, underlying validity issues that are raised by the use of ELP assessments and accommodations (U.S. GAO, 2006). The U.S. GAO report of 2007 (U.S. GAO, 2007) revealed that among 38 states in which the U.S. Department of Education (U.S. DOE) conducted peer reviews, 25 lacked evidence on the reliability and validity of their ELL assessment results. The peer reviewers commented that many states were not taking appropriate steps to ensure the sound assessment of ELL students. The report implied that states would need comprehensive guidance and additional support to build a valid assessment system for ELL students.

With the purpose of helping states deal with the challenges of developing appropriate ELL assessments and using results to support sound policies and practices, a team of

researchers at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) conducted an extensive review of literature and practices regarding ELL assessment. As a result, we are producing a series of three separate but interrelated reports. The first report in conjunction with this research contains a synthesis of research literature related to ELL assessment and accommodation issues, and we will hereafter refer to it as the *Literature Review*—CRESST Tech. Rep. No. 731 (Wolf et al., 2008a). The *Literature Review* report also reviews validity theory and reports on the latest research related to ELP assessments. A second companion report, referred to as the *Practice Review*—CRESST Tech. Rep. No. 732 (Wolf et al., 2008b), entails a review of ELL assessment and ELL policies for the 2006–2007 school year across 50 states, and summarizes the results and implications of our study of ELL assessment practices. Specifically, in that report we examined state policies in identifying and redesignating ELL students, using ELP assessments and available validity evidence on the use of ELP assessments, and policies and practices on assessing ELL students' attainment of content standards, including use of testing accommodations (see Wolf et al., 2008b for details). The present report, *Recommendations*, is based on the findings of the prior two reports, and highlights recommendations for state policy and practice as well as for future research and development. For further information regarding the research literature and the states' policies and practices, please refer to the two companion reports (CRESST Tech. Rep. No. 731 & 732).

The present report is comprised of two sections. The first section highlights a series of recommendations drawn from our research and practice reviews. In the second section, we discuss the areas that call for researchers' immediate attention to fill gaps between research and practice.

Recommendations for Assessing ELL Students

States are currently undergoing a major transitional period in establishing quality assessment and accountability systems for ELL students. Most states have updated or developed, within the past 5 years, new policies on assessing both ELL students' English language development and their content knowledge using accommodations. A vast body of literature suggests that validating assessment systems for ELL students is a complex and challenging task, given the heterogeneous characteristics of ELL students. Adding to that, our *Practice Review*—CRESST Tech. Rep. No. 732 (Wolf et al., 2008b), highlights the substantial variation in states' ELL polices. Variations were also present across local districts, raising concerns about the comparability, and validity of assessment results and uses, even within a state. Without comparability, inferences about the relative success of schools and students are suspect. While it is inevitable that some decisions are left to local

districts and schools, consistency demands the use of common guidelines for assessment decision-making about ELL students. We therefore recommend that clear guidelines and procedures be established by every state and made easily accessible to their respective local education agencies. Additionally, we recommend that states periodically monitor adherence to the guidelines, and establish and maintain a centralized database to record and track every decision made about each ELL student's identification, redesignation, and accommodation use. A centralized database containing ELL students' information (e.g., identification, ELP assessment results, redesignation, and accommodation use) facilitates the process of monitoring adherence to policy. Our recommendations for assessing ELL students are described in further detail in the following pages.

This set of recommendations is intended to support policies that can assist practitioners who are involved in assessing ELL students, determining what accommodations might be most appropriate, and making academic decisions about those students. Although we address these recommendations primarily to states' policymakers, we anticipate that the recommendations will also be informative for local district administrators, classroom teachers, and test developers. In addition, our reviews in research and practice revealed that there is a great need for further research related to assessing ELL students. Our recommendations thus identify key areas that urgently need research-based evidence to help practitioners make valid decisions.

Our recommendations are divided into three areas: (a) assessing English language proficiency and development, (b) assessing academic achievement with the use of accommodations, and (c) establishing other policies in identification and redesignation.

Recommendations for Using New ELP Assessments

The majority of states have adopted newly developed ELP assessments within the past 2 years. To assure the validated and appropriate use of these new ELP assessments, we recommend considering the following issues.

States Ought to Clearly Identify Primary Purposes for Their ELP Assessment

Based on information from available states' documents, many states were found to use their ELP test for multiple purposes (e.g., identification, placement, redesignation, diagnostics, and instruction). However, it is important to recognize that a single assessment of limited duration cannot serve all purposes, and professional standards require evidence of validity for each intended use of any assessment (The *Standards for Educational and Psychological Testing,* American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education

[NCME], 1999; hereafter will be referred to as the *Standards*). For example, measuring the progress of student English development requires comparable, vertically scaled assessments from year to year; diagnosing students' strengths and weaknesses requires detail on the source of students' problems. Clearly identified primary purposes provide an essential foundation for appropriate validation procedures. By publicly document the primary purposes of their ELP assessment, states may also help to avoid the misuse and misinterpretation of the test across districts and schools.

States Should Consider a Series of Validation Studies for Each Priority Use

States need to conduct extensive validation studies relevant to priority purposes, and thus ensure the appropriate use for each intended purpose. For example, under NCLB, states are using ELP assessments for the purposes of determining levels of English proficiency and measuring the progress of ELL students' English language development. Validation studies may include reviews of test design to examine the appropriateness of plans to classify and/or distinguish proficiency levels as well as to assess progress from year to year; content analysis of the extent to which test items are aligned with intended constructs, and empirical studies to examine the reliability and validity of measures of proficiency level and of progress from one year to the next. Specifically, for example, validity evidence can be provided by examining the test blueprint, alignment of content with standards, construct comparability across students, and classification consistency (Rabinowitz & Sato, 2006; see Wolf et al., 2008a for examples of validation studies and types of validity evidence).

Given that many states are still in the process of identifying an ELP test suitable for their needs, a fundamental validation study should be concerned with examining the alignment between the constructs and content addressed by the state's ELP standards and those addressed by their ELP test. The degree of the alignment between the ELP and content-area standards (e.g., English language arts) should also be examined. NCLB legislation stipulates that the constructs of an ELP test be aligned with the state's ELP and content standards in order to measure the progress of appropriate English language development. Our review of policies and practices revealed that there were different degrees to which states incorporated the characteristics of academic English in their standards. Even though the same ELP assessment (i.e., within a consortium) may be used by several states, ELP standards in these states were not necessarily the same. In identifying the most appropriate ELP test for a state's needs and making valid inferences from the test results, states should examine the constructs addressed by their ELP tests and the match between these construct and their ELP standards.

Additionally, states should assure the alignment of the levels of proficiency defined by their standards (e.g. beginning, intermediate, advanced) and those used in the test. Many states' ELP standards included detailed descriptions on what knowledge and skills ELL students are expected to have developed at each level of proficiency. Typically these standards are divided into four language domains (i.e., listening, speaking, reading, and writing) and specific grade bands. However, for many states, there was a tendency for inconsistency between proficiency level descriptions for their standards and ELP assessments. Assuring the alignment between the two sets of proficiency levels is urgently needed in order to establish valid use of the test results.

States Should Keep a Systematic Database of the ELP Test Results to Make Validation Procedures More Effective and Efficient

As described above, validating each intended use of an ELP test is a complex process. It involves examining multiple sources of validity evidence, including the contents of test items and standards, relationships among and between other assessments, and students' background characteristics. In order to make this complex process more effective and efficient, states should develop comprehensive and systematic databases. Such databases should include individual student data at the item level, as well as for the four modality levels. The database also should contain detailed background information (e.g., native language, level of language proficiency, instructional history, mobility, socioeconomic status) to enable investigations of reliability and validity of the assessments for students of various backgrounds, as well as to examine the effects of decision rules for clarifying students' proficiency levels and readiness for redesignation. It also is recommended that the data should maintain individual student and teacher IDs so that student performance can be linked to teacher professional development needs and can guide the improvement of instructional programming improvement.

<div align="center">

Recommendations for Using Accommodations

</div>

Appropriate accommodations enable English learners to show what they know and can do on content tests administered in English (e.g. a math test) by reducing the interference of English language demands of the test. Making appropriate decisions about accommodations requires a number of inter-related issues:

- Who needs accommodations?
- What accommodations are permissible?
- Who determines accommodation provision?
- For what assessments are accommodations given?

- What types of accommodations are provided?
- What are the selection criteria for accommodations?
- How are the accommodations administered?

Because of the variability across states, districts, and schools in how these considerations are addressed, comparability of accommodated test scores is a serious problem, which in turn poses a significant validity concern. It is of paramount importance that states assure the comparability of ELL students' scores on academic assessments across districts and schools by establishing clear policies and procedures for assigning and using accommodations. Our review of practice found that many states do provide some type of accommodation guidelines. However, there was a wide range of variation in states' guidelines in terms of specificity. Some guidelines did not distinguish between ELL students and students with disabilities. A closer look at available states' guidelines also revealed that a number of guideline documents left decisions in the hands of local districts. Although a certain degree of variability among local districts may be unavoidable, states can lessen it by providing specific guidelines, criteria, and a standardized set of procedures that districts can rely on as guidance. Such guidance will help reduce the variability by making the accommodation provision process as standardized as possible across districts and schools.

One notable effort recently developed to address this variability issue in practice is the Selection Taxonomy for English Language Learner Accommodations (STELLA) designed by the University of Maryland, in collaboration with the South Carolina Department of Education. STELLA is a computerized decision-making system to help practitioners define and identify ELL students, and match these students to the appropriate accommodations (Kopriva & Carr, 2006; Zehr, 2007). This system is currently undergoing a validation process, but seems to be a potentially promising system for both facilitating decision making, as well as reducing variability.

States Should Provide a Comprehensive, ELL-Specific Accommodation Guideline Document for Local Districts

ELL-specific accommodations. States should clearly define ELL-specific accommodations and delineate procedures for determining which accommodation will be provided to any given student. While some states emphasized that decisions needed to be made on an individual, case-by-case basis, some states did not distinguish between ELL students and students with disabilities. Research suggests that direct linguistic support is crucial for ELL students; therefore, ELL-specific guidelines are necessary for teachers and test administrators to provide effective and appropriate accommodations.

What should an accommodation guideline document contain? Further empirical research is needed to provide definitive answers about appropriate accommodations. However, based on the currently available research literature, which is very limited, it appears critical to consider the following areas in state accommodation guidelines for ELL students:

- Identify the background of each ELL student. Differences in students' backgrounds, such as their native language and levels of academic English proficiency, can affect their test performance. Much research suggests that the effect of an accommodation may yield different results depending on ELL students' level of language proficiency (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Short & Fitzsimmons, 2007). States should provide information about the effect of accommodations in relation to a student's background and encourage local decision makers to consider this as one of the criteria for selecting an appropriate accommodation. For instance, accommodations related to native language provision may not be effective for the students who are not proficient in their native language and who were not instructed in their native language.

- Identify decision makers. We found that states provided little guidance regarding who should be the decision makers in local districts, and that specific decision makers varied in their district roles. In some cases, states did delineate specific "teams" of individuals that met regularly to make decisions. We suggest that a decision-making team be formed to include at least a bilingual or English as a Second Language (ESL) coordinator and classroom teachers who are familiar with individual ELL students' language proficiency and needs.

- Specify allowable and prohibited accommodations. Some of the state guidelines that we reviewed only included a list of accommodations without referring to particular content areas. Only a handful of the states provided comprehensive tables specifying which accommodations were allowed on which assessment. It is important that guideline document be based on research findings (and preferably include justification) for each accommodation for why it is allowable or prohibited. For instance, a list of accommodations should be paired with corresponding research findings. This information will assist the local districts and teachers in appropriately selecting accommodations.

- Clearly define each accommodation operationally for the appropriate administration of each accommodation. During our review, we encountered difficulty in distinguishing similar accommodation terms (for example, simplifying versus clarifying, explaining, or paraphrasing directions or the language in directions). The operational definition also should specify the conditions and procedures for administration, so the accommodated testing condition can be as uniform as possible. For instance, while providing simplified directions, classroom teachers may unintentionally provide unintended help if what constitutes the "simplified directions" is not tightly specified. The provision of unintentional clues, in turn, can change the construct being measured and compromise comparability of results. In

order to avoid unintended effects, oral directions or translation may be provided by means of an audiotape recording.

During our review, we encountered states that incorporated specific features into their accommodation guidelines. Figure 1 presents portions of guidelines as examples of those we found useful.

Alabama

Alabama provides its districts with a document on assessment program policies and procedures for students of special populations which includes information specifically regarding ELL students. The document defines ELL students, identifies the state assessments they are required to participate in, and provides a brief statement about the purpose and policies for using accommodations on the state tests. Although the state allows the individual districts to make accommodation decisions, the state requires that all cases provide documentation stating need and success of accommodations in regular classroom settings and testing as well as which specific accommodations will be used on a specified assessment.

Points of Practice

- State provides district with a document or manual specifically addressing assessment and accommodation policies and procedures for ELLs or Special Populations.

- State makes a distinction between special education and ELL student accommodations.

- State specifies policies of ELL assessment, including which assessments ELL students must participate in.

- State requires documentation on the rationale for allowing accommodations, who makes the decisions for accommodation usage, and which specific accommodations will be used on a specific test.

- State provides an ELL student-specific checklist of acceptable accommodations for each state assessment as shown below.

Figure 1. Example from a state's accommodation guideline (*continues on next page*).

LEP/ELL ACCOMMODATIONS CHECKLIST   Revised May 2003
Stanford Achievement Test, Tenth Edition (Stanford 10)
Alabama Reading and Mathematics Test (ARMT)

The accommodations specified below are in keeping with what has been practiced regularly in the classroom when the student receives instruction and takes classroom tests. When completed by the LEP/ELL Committee, this checklist becomes part of the student's LEP Plan.

Name: _____ School: _____ Grade: _____ Year: _____

A. Scheduling Accommodations. Test will be administered:
   - 1. At a time of day most beneficial to student.
   - 2. In periods of one subtest followed by a break of _____ minutes.
   - 3. With flexible scheduling.
   - 4. With other accommodations needed due to the level of language proficiency.
            SDE APPROVAL ONLY.

B Setting/Administration Accommodations. Test will be administered:

   - 1. In a small group.
   - 2. With student seated in front of classroom.
   - 3. In a carrel.
   - 4. With teacher facing student.
   - 5. By student's ESL teacher.
   - 6. In ESL classroom.
   - 7. Individually.
   - 8. Using interpreter during time oral instruction is given to the student. (Interpreter may only interpret directions–interpreter may not clarify or offer interpretation of items.
   - 9. With other accommodations needed due to the level of language proficiency.
            SDE APPROVAL ONLY.

C. Format and/or Equipment Accommodations. Test will be administered with:

   - 1. Templates.
   - 2.Marker to maintain place.
   - 3. Noise buffers.
   - 4. English/native language translation dictionary (word-to-word translation/no definitions).
   - 5. English/native language electronic translator (word-to-word translation/no definitions).
   - 6. With other accommodations needed due to the level of language proficiency.
            SDE APPROVAL ONLY.

41

Figure 1. Example from a state's accommodation guideline (*continued*).
Source: ftp://ftp.alsde.edu/documents/91/PoliciesAnd%20Procedures_SpecialPopulations_Revised0307.PDF

What are the general principles that determine allowable accommodations?
Inarguably, accommodations should be determined on the basis of individual needs. Several

states did emphasize this in their policy. However, decisions about individuals must be made in a principled and uniform way. We suggest a set of general principles for states to consider for selecting allowable accommodations in their guidelines.

- Effectiveness and validity based on supporting evidence: Accommodations should be effective in improving ELL students' performance on content tests, without providing an unfair advantage for non-ELL students who do not receive the accommodation. Previous research suggests that accommodations related to direct linguistic support are effective for ELL students. For example, linguistic modification of non-content vocabulary and sentence structure were found to be effective types of accommodations (Abedi & Lord, 2001).

- Alignment between ELL students' language proficiency and accommodations: Given that linguistic modifications are the most effective type of accommodations, adequate and timely measurement of an ELL student's language proficiency is critical as a basis for selecting appropriate accommodations. For instance, a student's poor performance on academic vocabulary on an ELP language test may suggest the need for access to a bilingual glossary with a subsequent science assessment. States should encourage teachers to systematically use the results from ELP tests, formative assessments, and teachers' observation checklists to identify appropriate accommodations that match with an ELL student's language needs. Team meetings should be conducted regularly (at least more than once annually) to re-evaluate individual student needs.

- Invariance in the construct to be measured: Accommodations should not alter the construct(s) measured by the test. The types of accommodations allowed and the way they are implemented should not compromise measured construct(s) to the maximum extent possible. The *Standards* should be applied (AERA, APA, & NCME, 1999).

- Equal accessibility to and familiarity with accommodations—see *Standards* (AERA, APA, & NCME, 1999): Allowed accommodations should be accessible to all ELL students across local districts. Accommodation procedures should also be open and transparent, and equally familiar with all test takers. If a state includes a bilingual dictionary as one type of accommodation, for example, the state should make sure that the same dictionary is allocated and available across all schools.

States Should Provide a Uniformly Structured Database for Local Districts to Use in Order to Monitor ELL Students' Accommodation Use and Keep Track of What Accommodation(s) Each Student Received

There is a great need to evaluate the effects of various types of accommodations on ELL testing in order to facilitate accommodation decision-making and uniformity in decision-making. To make this possible, states should establish a uniform data structure for local districts to utilize and make it convenient for the districts to report their data to the state. The database should include information about the student characteristics (e.g., native

language, ELP) that were used in determining the specific accommodation, the types of testing accommodations available for each student, and the types of the assessments for which the accommodations are provided. A centralized database at the state level could then be used to evaluate the effects of the accommodations as well as to monitor the use of the accommodations across local districts. Figure 2 displays one example of an accommodation recording sheet that is used for compiling such a database. This example is provided by the U.S. DOE for the National Assessment of Educational Progress (NAEP). It contains a list of allowable accommodations for ELL students and the subtests to be accommodated.

This Student	COLUMN A — Accommodations student receives on state assessment in NAEP subject	COLUMN B — Are these accommodations allowed on NAEP?			COLUMN C — If allowed on NAEP, who provides accommodation?
		Reading	Math	Writing	
Direct Linguistic Support					
Has directions read aloud/repeated in English or receives assistance to understand directions	○	Standard NAEP Practice			NAEP provides
Has directions **only** read aloud in native language	○	N	Y*	N	†Spanish/English Only
Has test materials read aloud in native language	○	N	Y*	N	†Spanish/English Only
Uses a bilingual version of the booklet	○	N	Y	N	NAEP provides (Spanish/English Only)
Uses a bilingual word-for-word dictionary without definitions	○	N	Y	Y	School provides
Has **occasional** words or phrases read aloud in English	○	N	Y	Y	NAEP provides
Has **all or most** of the test materials read aloud in English	○	N	Y	Y	NAEP provides
Has oral or written responses in native language translated into written English	○	N	N	N	NA
Indirect Linguistic Support					
Takes the test in a small group (5 or fewer)	○	Y	Y	Y	NAEP provides**
Takes the test one-on-one	○	Y	Y	Y	NAEP provides**
Receives preferential seating	○	Y	Y	Y	School provides
Has test administered by familiar person	○	Y	Y	Y	School provides
Receives extended time	○	Y	Y	Y	NAEP provides
Is given breaks during the test	○	Y	Y	Y	NAEP provides
Takes test session over several days	○	N	N	N	NA
Receives other accommodations	○				

NA = Not applicable

* Spanish only and only permissible when a Spanish/English bilingual booklet is used.

** NAEP provides staff to conduct small group or one-on-one sessions after regular sessions.

† NAEP provides written directions in the bilingual booklets for students to read. Instructions in Spanish are provided for a bilingual, school-provided interpreter to read aloud to the student, if required.

Figure 2. Example of an accommodation recording sheet.
Source: National Assessment of Educational Progress (2007). *English language learner background questionnaire.* Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved May 9, 2007, from http://nces.ed.gov/nationsreportcard/pdf/bgq/ sch-sdlep/BQ07-NAEP-ELL.pdf

States Should Evaluate the Efficacy and Validity of Their Allowable Accommodations

As indicated in our *Literature Review* report (Wolf, et al., 2008a), research findings are inconclusive about the effectiveness of various types of accommodations. Limitations in research are partly due to the difficulty in obtaining relevant data and the wide variations in ELL characteristics. States should actively initiate the evaluation of the effects of their accommodations for their specific ELL populations. Keeping track of the use of accommodations and students' performance on the accommodated tests is essential in making this evaluation process feasible and effective. States should allocate resources for collecting and organizing the data, and for evaluating the effects of accommodations. One way of building the capacity to investigate the efficacy of accommodations is to collaborate with researchers and test developers.

States Should Provide Regular Professional Development on the Valid Use of Accommodations

States also should allocate resources and time to educate test administrators and teachers on what and how accommodations should be implemented. Some states may have a clear policy on accommodation use without issuing specific procedures for districts to follow. For states that leave the decision-making procedure to districts, it is crucial to hold regular professional development meetings well prior to the period of assessment for those who will make accommodation decisions, as well as for those charged with administering the accommodations. Professional development should include clear directions for standardized administration of accommodations that are established and regularly adjusted based on the most current information provided by state and research findings. Clear and specific guidance will help reduce variability that may threaten comparability of accommodated test scores across different schools and districts within a state.

<div align="center">

Recommendations for Other ELL Policies:
Identification and Redesignation

</div>

In addition to explicit guidelines regarding the use of ELP assessments and accommodations, we recommend that states provide guidelines for local districts to use in making decisions about identifying ELL students and redesignating them as fluent English proficient (FEP). The lack of consistency and clarity in states' guidelines on these issues increases difficulty in tracking ELL students' academic progress.

States Should Clearly Document Their Definition of ELL Students and the Procedures for Identifying Them

We found a lack of common definition or a lack of public documents across states on how to define and identify ELL students. It was also unclear whether states included newly redesignated ELL students in their ELL population. Procedures for identifying and determining levels of proficiency are considerably different across states owing to the different sources of information and various language assessments being used. The issue of comparability is more significant for those states that allow local districts to choose a language proficiency assessment from various tests. In making a validation argument, states should consider how their ELL students were identified and how their levels of proficiency were determined. Short and Fitzsimmons (2007) also pointed out that the lack of common criteria for identifying ELL students within and across states produces inconsistencies in research and evaluation findings (e.g., in identifying the relative success of schools and programs for ELL students). A clear definition of ELL and clear identification procedures from each state may contribute to establishing a common set of criteria across states.

States Should Establish Explicit Policies and Procedures for Redesignating ELL Students

By the same token, states should have the authority to set the criteria for reclassifying ELL students as FEP, as well as to monitor their implementation. Our review found that many states allowed the decision-making process to take place at the district level without clear guidelines. Even where states had clear guidelines, the policies of some states were such that districts and schools could use additional considerations for making redesignation decisions at the local level. While we recognize the sometimes challenging nature of state and district relationships, considering the significance of the policies for redesignated ELL students (e.g., AYP reporting, accommodation policies, and tracking the impact from state and federal educational policies over time or across districts), very serious comparability issues arise from this practice. In addition, our review revealed that it was unclear as to whether all states monitor newly redesignated ELL students and maintain a tracking system to do so. This is an important issue for states to consider in establishing a redesignation policy because an examination of these students' performance can provide evidence to support the states' redesignation criteria.

To illustrate what a clear and accessible guideline could contain in practice, we present, as an example, an excerpt from a state's redesignation policy and procedure guideline that we considered clear and useful (see Figure 3).

<u>Pennsylvania</u>

Pennsylvania has developed a standard set of ELL exit criteria to be incorporated into all state English Language Instructional programs. The state requires that ELLs reach a specified proficiency score on both the state assessment and ELP assessment, with alternatives for "special circumstances." Additionally, students must either obtain a specified grade in core subject areas or "scores on district-wide assessments that are comparable to" the specified performance level on the state assessment.

<u>Points of Practice</u>

- State provides district with a document or manual specifically addressing exit criteria for ELLs
- State specifies criteria for exiting including special circumstances
- State has set objective and unambiguous criteria for LEAs to follow

Exit Criteria for Pennsylvania's English Language Instructional Programs

for English Language Learners

The exit criteria provided below for English Language Learners (ELLs) represent valid and reliable evidence of a student's English language proficiency to exit from an English language instructional program. Every local educational agency (LEA) must include the following exit criteria in the LEA Program Plan for ELLs.

In order to meet the required State exit criteria for Pennsylvania's English language instructional programs for ELLs, LEAs must use both of the required exit criteria listed below. In addition, LEAs must ensure that students meet one of the 2 additional exit criteria provided below to exit from an English language instructional program:

Required Exit Criteria:
1. Score of Basic on the annual Pennsylvania System of School Assessment (PSSA).

> ***SPECIAL CIRCUMSTANCES:***
> • *For students transferring from other states, out-of-state academic achievement assessment results may be considered when the academic proficiency level is comparable to Basic on the PSSA.*
> • *For students that are in a grade that is not assessed with the PSSA, LEAs must use each of the remaining criteria listed below to exit students.*

2. Score of Proficient (Bridging as per the Pennsylvania Language Proficiency Standards for English Language Learners) in the areas of Listening, Speaking, Reading and Writing on the annual state English language proficiency assessment. The Proficient (Bridging) score will be based on the total composite assessment results.

Additional Exit Criteria:
1. Final grades of C or better in core subject areas (Mathematics, Language Arts, Science and Social Studies).

2. Scores on district-wide assessments that are comparable to the Basic performance level on the PSSA.

In addition to the release of this Pennlink, Pennsylvania Department of Education (PDE) will post the required statewide limited English proficient (LEP) exit criteria on our website at www.pde.state.pa.us/esl . The PDE is committed to assisting all LEAs in Pennsylvania and will be available to answer any questions regarding the required statewide ELL exit criteria and the timeframe established for implementation. Questions may be directed to …[omitted by the authors].

Figure 3. Example of a state's redesignation policy and procedure guideline that we considered clear and useful. Source: http://www.pde.state.pa.us/esl/lib/esl/Pennlink_State_Required_Exit_Criteria.pdf

<center>**Research Agenda**</center>

More research is needed to assure evidence–based reform efforts and to improve the quality of assessment and accountability for ELL students. We suggest the following research agenda in order to bridge the gaps between current research and practice and to assist practitioners in validly assessing ELL students.

Expanding the Current Empirical Study of Constructs of Academic English

More extensive research is needed to provide a better understanding of the constructs of academic English proficiency. As reported in the *Literature Review* (Wolf, et al., 2008a), research has suggested that measuring academic English is necessary to better predict ELL students' readiness for mainstream classrooms. Although many new ELP assessments and states' ELP standards have attempted to include the features of academic English, a comprehensive, operationalized definition of academic English proficiency has yet to be developed. Research in this area needs to be expanded to examine the language demands of various subject areas at various grade levels for all four language domains. Research on literacy is abundant, but sparser in oral proficiency, for instance.

Investigating the constructs of new ELP assessments. Since many states use newly-developed ELP assessments based on their states' standards to monitor the progress of ELL students' language proficiency based on their states' standards, an examination of the alignment between the states' ELP assessments and their ELP standards will yield helpful information for states. Alignment studies that look at both language and cognitive demands will provide a fundamental piece of validity evidence to support states' appropriate use of their ELP assessments.

Another way to examine the constructs of these new ELP assessments is to compare the nature of language (language demands) between the ELP assessment and content area assessments. An investigation of the extent to which academic English proficiency is required in both ELP and content assessments will provide guidance on the use of ELP assessment results. For example, if the constructs of the ELP assessment are aligned with language demands in the content area assessments, the ELP assessment scores will be a good indicator for determining ELL students' readiness for mainstream classroom and high-stakes standardized content area assessments. An alignment study between the ELP and content area assessments would also provide important evidence for valid use of the ELP assessment results.

It also is important to examine the consistency of the constructs addressed across these new ELP assessments. Although the test developers claim that the constructs of their tests

<center>17</center>

include both academic and social language proficiency, empirical research to examine the way each test measures these constructs is needed, as well as to determine the comparability of tests across states.

Conducting empirical research on academic English language development. Although states have made great efforts to define their ELP standards, it was unclear whether their proficiency levels (e.g., beginning, intermediate, advanced) were defined based on empirical research. Empirical research on the stages of proficiency which students actually progress in developing academic ELP will provide useful information for practitioners to guide instruction, as well as being essential for establishing and validating ELP standards. In collaboration with states, for example, researchers could investigate ELP state data on ELP test results closely to uncover the stages at which ELL students actually attain specific features defined in states' standards.

Investigating Confounding Interactions between Content Knowledge and Linguistic Knowledge

To adequately assess ELL students' content knowledge and skills, it is important to reduce unnecessary linguistic complexity that may be present on a content area assessment. Investigating ELL students' performance on the test items in terms of their language demands and the depth of content knowledge processing may reveal a confounding interaction between the two. Experimental studies with test items with different degrees of language demand may also yield concrete linguistic features with which ELL students struggle. Research in this area can provide guidance on test development as well as appropriate interpretation of the test results.

Expanding Accommodation Research into Various Accommodation Types and Content Areas

Despite numerous studies on the effects of accommodations, findings are inconsistent, and provide limited evidence to assure valid procedures for selecting and applying appropriate accommodations. The types of accommodations being examined to date are also limited. The current review found that some accommodations that many states listed as allowable had not always been investigated in previous research. An expansion of research in ELL-specific accommodation types is critical to provide empirical evidence for practitioners to use to make provision decisions. As Koenig and Bachman (2004) pointed out, more experimental studies are needed in order to determine the effectiveness of specific accommodation types. Additionally, previous accommodation research mainly has been confined to mathematics and science assessments (Francis et al., 2006). Given that a different

set of accommodations apply in a language-arts or reading assessment, research should be expanded to these content areas. From another perspective, there has been a move toward the concept of Universal Design, which has been researched in the population of students with disabilities, but can be applied toward ELL students (Thompson, Johnstone, & Thurlow, 2002). Designing assessments that would be accessible to the greatest number of students possible would reduce the need for accommodations, as well as the controversy surrounding them. Empirical research to investigate the effects of universally-designed assessments for ELL students would offer useful information for developing ELL assessments. Lastly, an examination of students' response processes may also provide valuable insight on the effects of accommodations.

Conducting Longitudinal Studies

Investigating ELL students' performance on tests over time will provide validity evidence on the use of the assessments in determining students' levels of proficiency in language and content knowledge. Examining the growth patterns of redesignated FEP students is also important to determine whether or not these students are truly proficient enough to master academic ELP and handle academic materials.

Concluding Remarks

Given the rapid increase of ELLs in the nation, and the climate of accountability, efforts to address the education of this subpopulation of students continue to be of paramount importance. Thus, every aspect of their educational experience must be thoroughly examined and improved upon whenever possible. In the case of the present review, establishing a valid assessment system is imperative for ELL students in that assessment results not only are used to make academic decisions (e.g., identification of proficiency level, placement of instructional program, inclusion in large-scale assessments), but also possess major implications for ameliorating the quality of teaching and learning. Establishing sound measures to serve these functions requires a process of continual validation. One of the most serious issues emerging from our reviews is the comparability of assessment and research results both within and across states, or more precisely, the lack thereof. Such incomparability stems partly from each state's having its own specific academic and language development standards and assessment system. However, our concern arises primarily for the lack of comparability within states. States lack common criteria and clear guidelines on how to apply those criteria in making decisions about ELL students' identification and redesignation, and about ELP assessment and accommodation uses. Our recommendations thus are intended to provide information to assist states in establishing a

set of clear and comprehensive guidelines for making ELL-related decisions. While we recognize the limited or strained resources in many states, we also encourage states to collaborate with researchers and test developers to address states' specific issues in validating their assessment systems. We anticipate that rigorous collaboration between researchers and practitioners in this research-based reform effort will ultimately benefit ELL students and help them to meet high academic expectations.

References

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

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Francis, D. J., 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.* Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved November 21, 2006, from http://www.centeroninstruction.org/files/ELL3-Assessments.pdf

Koenig, J. A., & Bachman, L. F. (2004). *Keeping score for all: The effects of inclusion and accommodation polices on large-scale educational assessments.* Washington, DC: The National Academies Press.

Kopriva, R., & Carr, T. G. (2006, June). *STELLA: A computerized system for individually assigning test accommodations to ELLs.* Paper presented at the annual meeting of the Council of Chief State School Officers, San Francisco, CA.

National Assessment of Educational Progress. (2007). *English language learner background questionnaire.* Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved May 9, 2007, from http://nces.ed.gov/nationsreportcard/pdf/bgq/sch-sdlep/BQ07-NAEP-ELL.pdf

No Child Left Behind Act of 2001, 20 U.S.C. § 6301 et seq. (2002).

Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (n.d.). *The growing numbers of limited English proficient students: 1994/95–2004/05* [poster]. Washington, DC: U.S. Department of Education. Retrieved July 6, 2007, from http://www.ncela.gwu.edu/policy/states/reports/statedata/2004LEP/GrowingLEP_0405_Nov06.pdf

Rabinowitz S., & Sato E. (2006, April). *Technical adequacy of assessments for alternate student populations: Technical review of high-stakes assessment for English language learners*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Short, D. J., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners— A report to Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education. Retrieved May 8, 2007, from http://www.all4ed.org/publications/DoubleWork/DoubleWork.pdf

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

U.S. Government Accountability Office. (2006). *No Child Left Behind Act: Assistance from education could help states better measure progress of students with limited English proficiency* (GAO-06-815). Washington, DC: Author. Retrieved July 20, 2006, from: www.gao.gov

U.S. Government Accountability Office. (2007). *No Child Left Behind Act: Assistance could help states better measure progress of students with limited English proficiency* (GAO-07-646T). Washington, DC: Author.

Wolf, M. K., Kao, J. C., Herman, J. L., Bachman, L. F., Bailey, A. L., Bachman, P. L., et al. (2008a). *Issues in assessing English language learners: English language proficiency measures and accommodation uses—Literature review* (CRESST Tech. Rep. No. 731). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Wolf, M. K., Kao, J. C., Griffin, N., Herman, J. L., Bachman, P. L., Chang, S. M., et al. (2008b). *Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review* (CRESST Tech. Rep. No. 732). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Zehr, M. A. (2007, February 6). Pilot program could help English-learners. *Education Week, 26*(22), 15–16. Retrieved August 10, 2007, from http://www.edweek.org/ew/articles/2007/02/07/22testing.h26.html