Mikyung Kim Wolf
Jenny Kao
Noelle Griffin
Joan L. Herman
Patina L. Bachman
Sandy M. Chang
Tim Farnsworth

ISSUES IN ASSESSING ENGLISH LANGUAGE LEARNERS: ENGLISH LANGUAGE PROFICIENCY MEASURES AND ACCOMMODATION USES

PRACTICE REVIEW
(PART 2 OF 3)

JANUARY, 2008

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

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

Issues in Assessing English Language Learners:
English Language Proficiency Measures
and Accommodation Uses

Practice Review

CRESST Report 732

Mikyung Kim Wolf, Jenny Kao, Noelle Griffin, Joan L. Herman,
Patina L. Bachman, Sandy M. Chang, & Tim Farnsworth
CRESST/University of California, Los Angeles

January 2008

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

ISSUES IN ASSESSING ENGLISH LANGUAGE LEARNERS:

ENGLISH LANGUAGE PROFICIENCY MEASURES

AND ACCOMMODATION USES—PRACTICE REVIEW[1]

Mikyung Kim Wolf, Jenny Kao, Noelle Griffin, Joan L. Herman,
Patina L. Bachman, Sandy M. Chang, & Tim Farnsworth
CRESST/University of California, Los Angeles

Abstract

The No Child Left Behind (NCLB) 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 an assessment in guiding decisions about organizations and individuals, it is of paramount importance to establish a valid assessment system. 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. We have compiled a series of three reports. The present report is the second component of the series, providing a comprehensive picture of states' current policies related to ELL assessment. The areas reviewed include the procedures of ELL identification and redesignation, the characteristics of English language proficiency assessments, including validity information, and the use of accommodations in the assessment of content knowledge.

Introduction

Specific mandates regarding English language learners (ELLs) under the No Child Left Behind (NCLB) Act of 2001 (NCLB, 2002) have greatly impacted states' policies in assessing ELL students. States must provide an annual academic assessment of English language proficiency and assure the monitoring of students' English language development.

States also must provide valid and reasonable accommodations to accurately measure ELL students' academic achievement on state content standards in reading (or English language arts), mathematics, and science (by school year 2007–2008). ELL students 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 increased by more than 60%, while the total K–12 growth was just over 2% (Office of English Language Acquisition [OELA], n.d.). ELL students are a heterogeneous group with over 400 different home languages reported. Their cultural and schooling experience can vary depending on the students' immigrant countries. Some states have much higher numbers of ELL students than others, with the largest populations living in California, Texas, Florida, New York, Illinois, and Arizona (Kindler, 2002; OELA, n.d.). (See Appendix A for information on the number and percentage of students receiving ELL services and the most common languages and percentage spoken by state.) Assessing this fast-growing, heterogeneous group has been challenging to states. Understanding how states follow the legislation and what policies they implement is critical because it reveals not only trends and patterns across states but also common issues that states may face. This information is useful for both researchers and practitioners in their efforts to improve the quality and practice of assessing ELL students.

The goal of this report is to provide a comprehensive picture of states' current policies in order to better understand issues surrounding ELL assessment in practice. Moreover, the review of the practice and policy in this report aims to identify key issues to consider in validating the intended use of ELL assessments and thus to improve ELL testing practice. The present report is the second in a series of three reports regarding ELL assessment validation and accountability issues compiled by a team of researchers at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). The first component of the series includes a synthesis of literature in the areas of validity theory, the assessment of English language proficiency (ELP),[2] and the effects of accommodations on testing ELL students' content knowledge and skills. We refer to the report of this component research as the *Literature Review*—CRESST Tech. Rep. No. 731 (Wolf, Kao, et al., 2008). The present report describes the second component, referred to as the *Practice Review*. In this report, we analyze the commonalities and variations across states' ELL policies and report validation issues, linking research and practice. Integrating these two components, the third component,

---

[2]NCLB mandates that states should measure ELL students' English language proficiency (ELP) and annually assess their English language development (ELD). The terms ELP and ELD are interchangeably used in practice for assessment and standards regarding ELL students' English language. In this report, we use the term ELP assessment to refer to assessing ELL students' language proficiency since Title III of NCLB also uses the term "English language proficiency."

referred to as the *Recommendations*—CRESST Tech. Rep. No. 737 (Wolf, Herman et al., 2008), presents a set of practical guidelines for policymakers and practitioners in ELL assessment practice.

There has been growing attention to state policies and practices concerning ELL assessment. Recently, Rabinowitz and Sato (2006) reviewed the types of ELP assessments that each state used for the school years 2004–2005 and 2005–2006. As discussed in our *Literature Review*—CRESST Tech. Rep. No. 731 (Wolf, Kao, et al., 2008), states have adopted or developed new ELP assessments to comply with NCLB requirements. The new ELP assessments attempt to measure academic English language as well as social language to better predict ELL students' readiness for English language school settings. Given the recent implementation of new ELP assessments, many important validity questions on the use of these assessments remain unanswered. In their review, Rabinowitz and Sato found that relatively little technical information from the test developers or states was available to indicate the technical quality of these new assessments. The researchers proposed a comprehensive set of validation criteria to evaluate the technical adequacy of an ELL assessment system. Abedi, Nambiar, and Porter (forthcoming) also examined the available data on the validity and reliability of current ELP assessments. Whereas their research primarily focuses on ELP assessment and includes test developers' (e.g., four consortia[3]) internal reviews of their own assessments, our research, as external reviewers, deals with a broader range of ELL assessment issues. In terms of policies on the use of accommodations, Koenig and Bachman (2004) reviewed states' policies and practices for the inclusion and provision of accommodations on the National Assessment of Education Progress (NAEP) and other standardized tests administered by states. The researchers pointed out that states' policies and procedures on accommodations varied widely, partly due to their different assessment systems. Rivera, Collum, Willner, and Sia's (2006) study of states' policies on the use of accommodations in all 50 states and the District of Columbia came to a similar conclusion. Although their review was comprehensive and detailed, it was based mainly on assessment practices in the school years 1999–2000 and 2000–2001. Considering the major changes that states have made in order to be in compliance with NCLB requirements, it is necessary to revisit this area to understand the current policies and issues.

The review reported here provides an updated and comprehensive view of the landscape of states' ELL assessment practice for the school year 2006–2007. In particular, it

---

[3] The four consortia are the World-Class Instructional Design and Assessment (WIDA) Consortium, the State Collaborative on Assessment and Student Standards (SCASS), the Mountain West Assessment Consortium (MWAC), and the Pennsylvania Enhanced Assessment Grant (PA EAG).

addresses each state's definition and system for identifying and designating ELL students, the nature of states' ELP assessments, and the states' policies and practices with regard to the use of accommodations in state standards-based assessments for Adequate Yearly Progress (AYP) determinations.

Information on state policies in each of these areas was gathered from of each state's department of education Web site (all 50 states and the District of Columbia) and the Web sites of the four consortia and the test development companies involved in the development of ELP assessments. Our review summary of ELL policies was sent to states for their additional feedback and input. We received feedback from 37 states. It should be noted that many states are still in the process of updating or revising their policies about ELL students, and changes may not be reflected in the present report. More detailed information on the policies from 49 states[4] and the District of Columbia[5] are summarized in the appendices. Occasionally, we were unable to find state information related to certain topics from the public Web sites. Throughout the report, we note the total number of states on which the analysis is based.

While reviewing and analyzing the commonalities and variations among states' policies, we will discuss the key issues and concerns in assessing ELL students throughout the four sections of this report. The first section analyzes policies on identifying and reclassifying ELL students. In this section, the different definitions of ELL and various criteria and processes used to identify ELL students, the ways in which states determine the level of language proficiency for ELLs, and the redesignation criteria used to exit students from their ELL status are discussed as key issues. The second section reviews the current ELP assessments that each state employs, focusing on the characteristics and use of each ELP assessment identified by states. The characteristics of each of the four consortia's assessments are particularly important, considering that the majority of states collaborated with a consortium. The areas examined include intended purposes, constructs and language domains addressed, grade bands, item format, administration, scoring, and reporting practice. Technical qualities, including validity and reliability information, are also discussed. The third section reports policies regarding the use of accommodations. Related to this area, criteria for including ELL students in state assessments for AYP testing and criteria for identifying the need for accommodations and for assigning accommodations are examined as major issues. In addition, various accommodation types used by states are reported. The final

---

[4] New Hampshire is excluded in this report. At the time of review, New Hampshire was in process of updating its Web site on information regarding ELL students' policies and requested not to be included.
[5] While the District of Columbia is not an official State, henceforth, for the purposes of reporting, instances where we refer to "the states reviewed" or "the 50 states" refers to the 49 states and the District of Columbia.

section provides recommendations for practitioners and researchers to improve the quality and practice of ELL assessment.

Policies on Identifying and Redesignating ELL Students

Koenig and Bachman (2004) noted that different terms used to indicate ELL students may imply different understandings of and attitudes toward these students. Prior to reviewing each state's policies on identifying ELL students, we first examined the terms that each state officially used.

Defining ELL and Identifying ELL Students

Limited English Proficient (LEP) and ELL are the most prevalent terms used by states to describe students who have not achieved full fluency in English. Many states use both terms interchangeably in state literature, documents, and official state definitions. Some states explicitly mention that the term ELL is preferred due to its connotation of developing English proficiency rather than being deficient in English. Since NCLB uses the term LEP in its documentation, most states have aligned their terms and state definitions of ELL with those provided by NCLB. While a few states use only the federal definition, many states have attempted to establish their own definitions for use as a supplement or alternative to the federal definition. The terms that each state uses are presented in Appendix B.

NCLB defines LEP in detail, characterizing these students as having a native language other than English by foreign birth or ancestry, living in an environment in which a language other than English is dominant, and having a degree of difficulty with speaking, reading, writing, or understanding the English language that interferes with social interactions and academic tasks. The review of states' definitions indicated that most states considered students' native language and English language ability to perform in classroom settings as keys to defining which students are ELLs. For instance, California defined ELLs as "a K–12 student who, based on objective assessment, has not developed listening, speaking, reading, and writing proficiencies in English sufficient for participation in the regular school program."[6] Texas described a student of limited English proficiency as "a student whose primary language is other than English and whose English language skills are such that the student has difficulty performing ordinary class work in English."[7]

---

[6] California Department of Education, English Learners in California Frequently Asked Questions (8/2006) http://www.cde.ca.gov/sp/el/er/documents/elfaq.doc
[7] Texas Education Code, Title 2, Subtitle F, Chapter 29 (1995) http://tlo2.tlc.state.tx.us/statutes/docs/ED/content/htm/ed.002.00.000029.00.htm

One notable issue that occurred during our review was that although some states explicitly mentioned including redesignated ELL students as part of their ELL population, other states were not clear about this inclusion, based on the information found on their public Web sites. Abedi (2004) pointed out that one of the reasons for ELL students' low scores in testing is due to the exclusion of scores of the newly redesignated students. This was prior to the NCLB's modification allowing redesignated students to be included as part of the ELL subgroup for up to an additional 2 years. However, it is currently unclear whether some states' definitions of ELL include redesignated ELL students, and further investigation is required. This issue raises an important comparability question across states' accountability systems.

Procedures for Identifying ELL Students

Although each state varies in its exact method of identifying ELL students and their English proficiency levels, there are some similar procedures that are frequently used by the states. Most states administer some type of home language survey, followed by an ELP assessment[8] based upon the results of the survey. Additional criteria were also found in 19 states, including information from academic achievement tests, informal classroom assessments, teacher observations, checklists, and interviews with the student and a parent/guardian. The criteria that each state applies are summarized in Appendix C.

Based on the responses to the home language survey, a student may be identified as an ELL and may then be tested for English language proficiency. All states use an ELP assessment to determine a student's English proficiency level and the level of ELL services the student may require. During the review in this area, three notable things were found.

First, although there was a trend for states to implement a single ELP assessment statewide, 16 states allowed local districts to choose a language assessment from a list of approved assessments. This practice raises an issue of comparability in determining levels of proficiency from different tests. In other words, a student may be classified at a different proficiency level within a state depending on the test that the student takes. This issue also raises a validity concern in that the results of the tests are used to place an ELL student into an appropriate instructional program. It is thus inevitable that the states that allow local districts to choose a test to determine students' levels of language proficiency need to provide evidence for the comparability of the constructs and cut score settings among different tests.

---

[8] Some states have two separate assessments: (1) an ELP assessment for the purpose of initially identifying ELL students, and (2) an ELP assessment only for the purpose of measuring the annual progress of English language development. As described earlier, we use the term ELP assessment to refer to any test that measures ELLs' language proficiency, either for initial identification or annual progress.

Secondly, it was found that most states implemented one ELP test both to initially identify ELL students (and then determine proficiency levels) and to monitor the progress of students' English language development. However, some states applied two different tests to serve the different purposes. This finding implies that states need different validation procedures for their assessment systems. In particular, states that use one test to serve multiple purposes should offer extensive validity studies to validate each intended use.

Thirdly, a closer look at the levels of proficiency defined in each state indicated a considerable amount of variability across states. As presented in Appendix D, states varied in how they defined the stages or levels of language proficiency through which students were expected to develop. Differences were evident in the number of levels and descriptive terms used for levels. The number of levels of proficiency ranged from three to six. Even among the states that had the same number of levels, the descriptions of each level were varied. The following overall descriptions in reading were taken from two states as an example. Although both states included five levels and named the highest level "advanced," the descriptions indicated differences in what each state expected. Both descriptions included academic language, but its description and focus are different. While the first one includes the importance of decoding skills and strategies to find word meanings, the second one highlights the various genres and rhetorical features in academic texts. The second one also underscores the understanding of different types of vocabulary (Note that the following are summary descriptions for reading proficiency. Each state also provides detailed descriptions, broken down by grade level bands, for each standard.).

- **5. Advanced**[9]
  Apply knowledge of sound/symbol relationships and basic word-formation rules to derive meaning from written text (e.g., basic syllabication rules, regular and irregular plurals, and basic phonics). Apply knowledge of academic and social vocabulary while reading independently. Be able to use a standard dictionary to find the meanings of unfamiliar words. Interpret the meaning of unknown words by using knowledge gained from previously read text. Understand idioms, analogies, and metaphors in conversation and written text.

- **5. Advanced**[10]
  Understand and obtain meaning from a wide range of texts available to native English speakers. Read academic texts at the appropriate level. Understand a variety of literary

---

[9] Description taken from the *English-Language Development Standard for California Public Schools*, found http://www.cde.ca.gov/re/pn/fd/documents/englangdev-stnd.pdf

[10] Description taken from the *Ohio English Language Proficiency Standards for Limited English Proficient Students*, found
http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=500&ContentID=6621&Content=32522

genres. Read and comprehend grammar and rhetorical features appropriate for the grade level. Master strategies of reading comparable to native English-speaking students at their grade level. Understand vocabulary that is basic and academic and able to figure out technical vocabulary. Read and interpret texts across the curriculum.

This kind of variability across states' standards raises an issue of comparability and thus accountability across states, questioning what is held accountable for each state. A more detailed review and discussion about the states' proficiency levels is included in the section entitled Policies and Practices on the Use of ELP Assessments, found later in this report.

Procedures and Criteria for Redesignating ELL Students

The term "redesignation" indicates the process by which ELL students are determined to no longer need special language support services and/or are considered able to fully function in English-only classes. Some states use the terms reclassification or "exit of ELL status." As presented in Appendix E, a variety of sources are used by states to determine whether ELL students are ready to be redesignated. The sources included state ELP test scores, content test scores, school personnel (e.g., teacher, ELP specialist, or committee) input, parental or guardian input, student grades, portfolio (work sample) assessments, interviews, and evaluations of classroom performance.

Figure 1 shows the prevalence of the various criteria for redesignation in the states for which information was found. Almost all of those states relied on the scores from their ELP assessments as one of the criteria to determine students' readiness for the mainstream classroom. A closer look at the use of multiple criteria in state policies highlights the complexity and variation in the redesignation process across states. As shown in Figure 2 states used between one and five different criteria in various combinations to make a decision about redesignation. Twelve states relied solely on ELP test scores to make a redesignation decision. A number of states included two criteria: test scores from both the ELP assessment and content tests. And whereas some states mandated a single method or combination of several methods, other states allowed individual districts to set their own redesignation criteria and merely provided guidelines for acceptable criteria. Of the 48 states for which information was found, 18 states allowed individual districts to decide criteria.
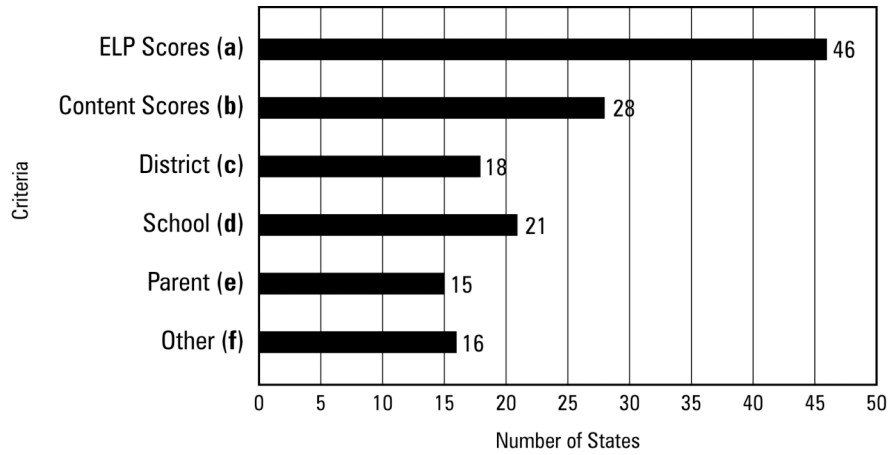
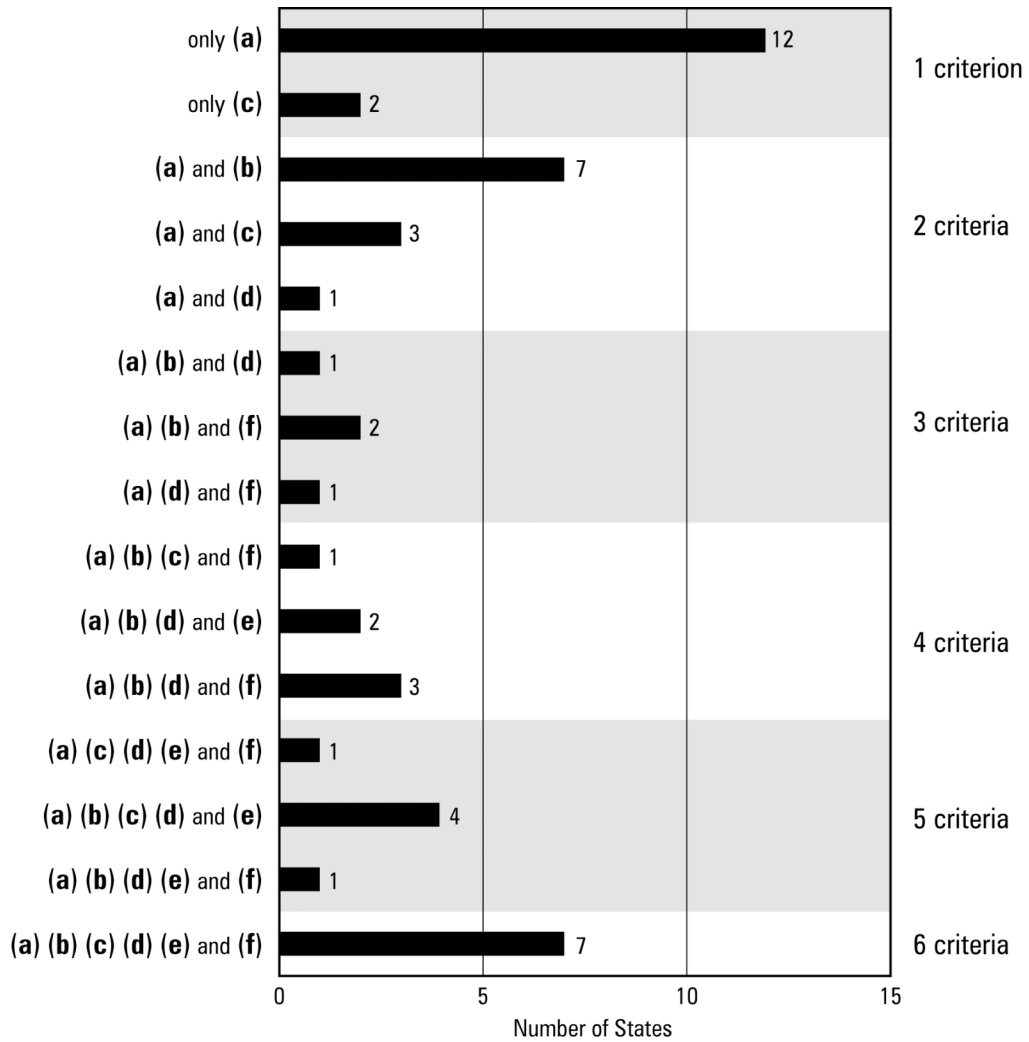*Figure 1.* Redesignation criteria used by states.

Figure 2. Combinations of redesignation criteria used by states. (a) = ELP Scores; (b) = Content Scores; (c) = District; (d) = School; (e) = Parent; (f) = Other.

This variation in redesignation criteria raises the issue of comparability of ELL students' performance both across and within states, in addition to the issue of the validity and reliability of the use of the criteria. That is, states should consider and examine the validity and reliability of each criterion (e.g., portfolio assessment, school personnel input) in addition to their ELP assessments.

This variation of policies even within a single state is thoroughly addressed in Jepsen and de Alth's (2005) study of California's 2002 and 2003 redesignation data, for example. In California, each school district develops its own policy and procedures for redesignation based on the reclassification guidelines approved by the State Board of Education. The criteria include scores from the state ELP and academic achievement assessments, teacher evaluation information, and parental consultation. Jepsen and de Alth found that of the ELL students who achieved the proficient level on the ELP assessment for reclassification in 2002, only 29% were redesignated. The authors pointed out that there is a great variability in weighing redesignation criterion among districts.

Summary

Notably, states vary in their policies for identifying ELL students and their English proficiency level, and redesignating them. As mentioned, the variety of criteria used for ELL identification and redesignation decisions highlights the complexity of validating a state's accountability system for ELL students. Figure 3 displays the most common criteria and flow used in the process of identifying and redesignating ELL students. As described above, most states administer a home language survey to incoming students (or their parents) as the first step in ELL identification. If there is indication of a language other than English spoken in the home, an ELP assessment is then administered. In addition to the ELP test scores, some states also apply other criteria based on sources such as school personnel input, parental consultation, and interviews. Identified ELL students are provided with appropriate instructional services and annually reviewed for their progress in English language proficiency and redesignation decisions. The most common redesignation criteria are based on scores from an ELP assessment alone or in combination with a content test score. Some states also take student grades, a student portfolio, teacher observations, and/or interviews into consideration. Redesignated students are exited out of ELL services, but they may still be included for AYP reporting as ELL students for 2 years. Title III of NCLB requires measurable achievement goals for ELL students, referred to as "annual measurable achievement objectives" (AMAOs). States are required to set AMAOs for the percentage of ELL students achieving language proficiency, the percentage of students exiting out of ELL

services, and the percentage of ELL students meeting AYP. NCLB allows states to designate exited students as "former LEP" only for determining AMAOs for up to 2 years.
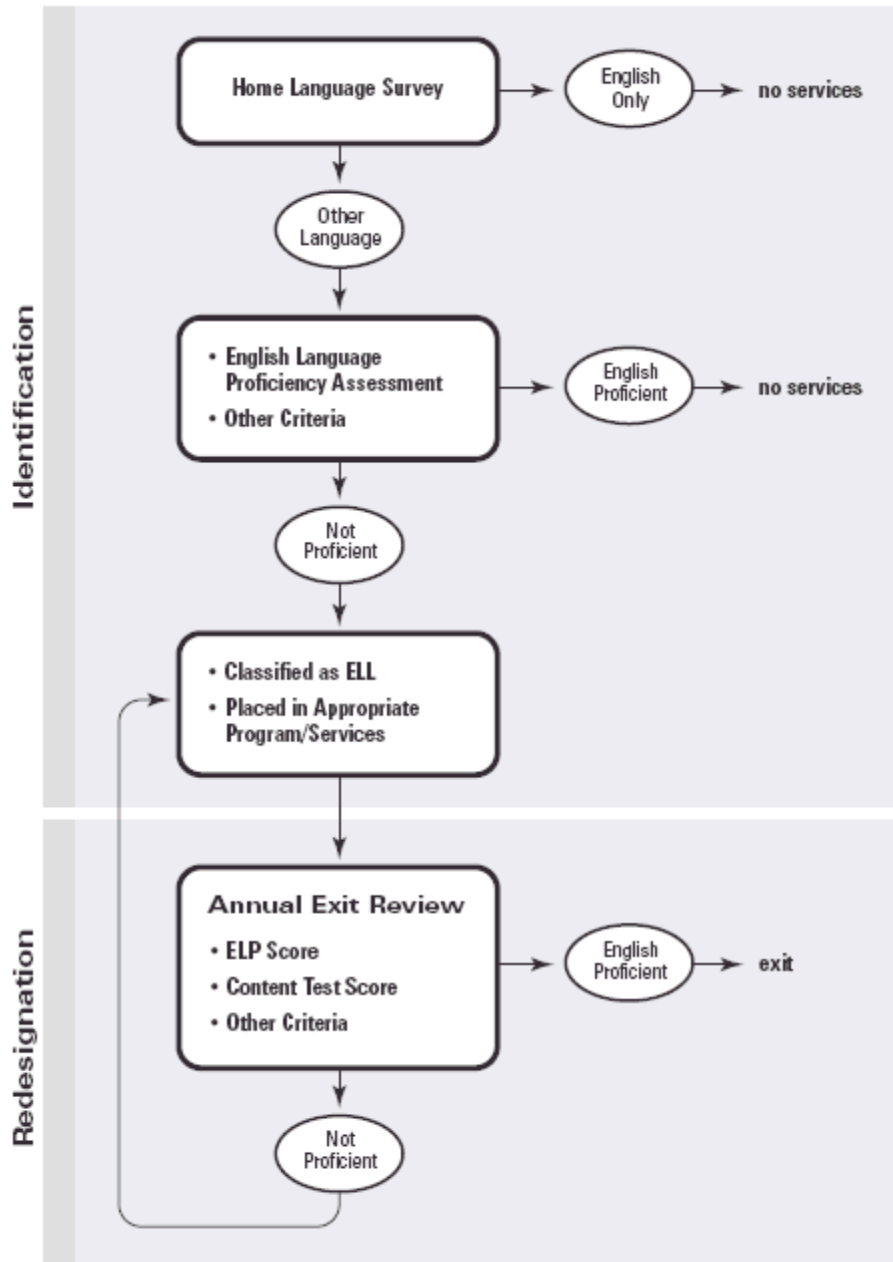

Figure 3. Common procedure for identifying and redesignating ELL students.

As discussed in our *Literature Review*—CRESST Tech. Rep. No. 731 (Wolf, Kao, et al., 2008), each criterion (e.g., home language survey, ELP assessments, content-area tests) should be validated for its intended purpose. For example, a comparison of ELL students'

performance on the ELP and content area tests can be one way to provide a piece of validity evidence to support the use of the ELP test as a redesignation criterion. (See the *Literature Review—CRESST* Tech. Rep. No. 731 [Wolf, Kao, et al., 2008], for examples of validation processes.)

Policies and Practices on the Use of ELP Assessment

As discussed earlier, all states have developed or adopted new ELP assessments since NCLB in 2002. This section summarizes the types of new ELP assessments and their characteristics. Technical qualities including reliability and validity are also discussed based on publicly available information.

Types of ELP Assessments

It was found that 43 states have used their current ELP assessment for less than 5 years. This finding signifies the states' rapid policy change in order to comply with NCLB requirements of measuring the progress of ELL students' English language proficiency in listening, speaking, reading, writing, and comprehension[11]. The states also recognize the importance of the alignment of constructs between their ELP standards and assessments. As a result, a large number of states have developed their own ELP assessments. The four consortia's tests have also been adopted by many states: *Assessing Comprehension and Communication in English State-to-State for English Language Learners* (ACCESS for ELLs®, henceforth), *Comprehensive English Language Learning Assessment* (CELLA), *English Language Development Assessment* (ELDA), and Mountain West Assessment Consortium (MWAC)[12] assessment. These tests were developed based on a set of ELP standards from the collaborating states; in the case of ACCESS for ELLs®, states agreed on a common set of standards upon which to base the new assessment. A brief summary of each consortium's assessment and collaborating states is presented in Table 1. At the time of this review, two consortia, the Pennsylvania Enhanced Assessment Group (PAEAG) and MWAC, were no longer active as consortia. Florida is the only state that implements an ELP assessment originally developed by PAEAG. MWAC states have developed their own state-

---

[11] NCLB does not define comprehension. States typically report comprehension scores as the combined scores of the listening and reading components of their ELP assessments. For this reason, this report will hereafter only review 4 domains, listening, speaking, reading, and writing.

[12] There is no specific name for the MWAC assessment. The consortium dissolved (2006) before the assessment was finalized. Some of the collaborating states used or incorporated MWAC items with other items to create their own ELP assessment. The states that selected MWAC items specific to their ELP standards each named the assessment differently (e.g., Idaho English Language Assessment; [Michigan] English Language Proficiency Assessment; Montana Comprehensive Assessment System [MontCAS] English Language Proficiency Assessment; New Mexico English Language Proficiency Assessment; Utah Academic Language Proficiency Assessment).

specific tests, in many cases using some of the items and materials developed by MWAC (Northwest Educational Regional Laboratory, 2006). Further detail about each consortium's test characteristics, including the constructs, item format, scoring, and reporting, are summarized in Appendix F.

Table 1

Four Consortia's Assessments and Collaborating States

Consortium	Mountain West Assessment Consortium (MWAC)	Pennsylvania Enhanced Assessment Group (PA EAG)	State Collaborative on Assessment & Student Standards (SCASS) Consortium	World-Class Instructional Design and Assessment (WIDA) Consortium
Web site	www.measuredprogress.org/assessments/clients/MountainWest/MountainWest.html	www.fldoe.org/aala/cella.asp	www.ccsso.org/projects/ELDA/	www.wida.us/assessment/ACCESS.aspx
Test name	—	CELLA	ELDA	ACCESS for ELLs®
States originally involved	AK, CO, ID, MI, MT, NV, NM, ND, OR, UT, WY	FL, MD, MI, PA, TN	AL, CA, HI, IN, IA, KY, LA, MI, NE, NV, NJ, NY, NC, OH, OK, OR, SC, TX, VA, WV	AL, AR, DE, IL, ME, NH, RI, VT, WI
Other collaborating entities	Measured Progress	Educational Testing Service, Accountability Works	University of Maryland, Council of Chief State School Officers, American Institutes for Research	Center for Applied Linguistics
States currently using consortia test/items[a]	ID, MI, MT, NM, UT	FL	AR, IA, LA, NE, OH, SC, TN, WV	AL, DE, DC, GA, IL, KY, ME, NJ, ND, OK, PA, RI, VT, WI

Note. — Does not apply to this category. CELLA = Comprehensive English Language Learning Assessment; ELDA = English Language Development Assessment; ACCESS for ELLs® = Assessing Comprehension and Communication in English State-to-State for English Language Learners.
[a] School year 2006–2007.

The current ELP assessment that each state used for the school year 2006–2007 is listed in Appendix G-1. Appendix G-1 also presents the number of years that the test has been used and the previous ELP assessments used. Prior to NCLB, states typically allowed local districts to choose an ELP test from a list of state-approved ones. Our review reveals that

although some states still allow district level decisions, the majority of states have established a policy of recommending use of one ELP assessment to annually measure the progress of language proficiency. As seen in Appendix G-1, 20 states out of 49 states where information was found have recently implemented a new ELP assessment beginning in the school year 2005–2006, and 12 states adopted a new ELP assessment beginning in the school year 2006–2007. This recent adoption indicates that many states have endeavored to identify an ELP assessment appropriate for their ELP standards and ELL students. Appendix G-2 lists the tests, their acronyms, and information on the test developers.

Figure 4 displays the trend for the current use of ELP assessments. In this figure, commercial tests include the New IDEA Proficiency Test (New IPT), Language Assessment Scales Links K–12 (LAS Links), Maculaitis Assessment of Competencies Test of English Language Proficiency (MAC II), and the Stanford English Language Proficiency (SELP) Test. Although state-developed tests were often designed in collaboration with commercial publishers, these were distinguished from commercial tests because they were specifically developed for each state's individual needs. We also classified those tests that had augmented existing commercial tests as state-developed tests, because they modified the tests in a manner specific to states' needs. As shown in Figure 4, the trend is toward a wide use of consortia assessments and state-developed assessments. ACCESS for ELLs® was the most commonly used ELP assessment among the consortia-developed assessments.
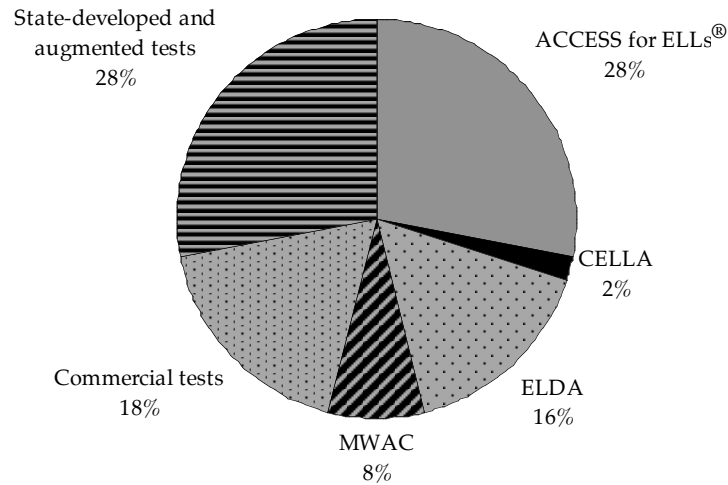
Figure 4. ELP assessments used by states for the school year 2006–2007.
Note. ACCESS for ELLs® = Assessing Comprehension and Communication in English State-to-State for English Language Learners; CELLA = Comprehensive English Language Learning Assessment; ELDA = English Language Development Assessment; MWAC = Mountain West Assessment Consortium.

Figure 5 illustrates which states fall into the following six categories in terms of their ELP assessment type: ACCESS for ELLs®, CELLA, ELDA, MWAC, commercial tests, and state-developed/augmented tests. It is notable that states with large ELL populations (e.g., California, New Mexico, New York, Texas) tended to develop their own ELP assessment. Although the consortia's assessments have been adopted by a great number of states, state-developed assessments actually have many more test takers, due to the size of the ELL population in those states.
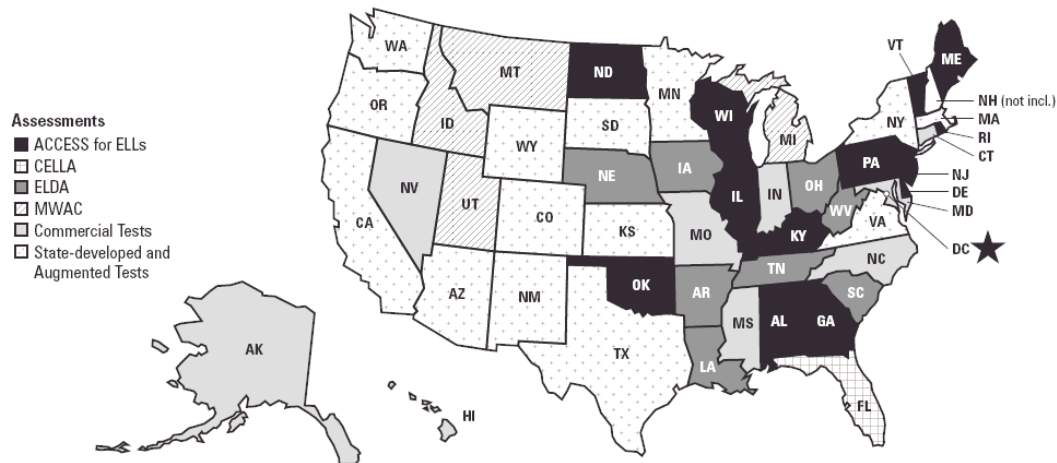
Figure 5. Types of ELP assessments used by states for the school year 2006–2007.

Test Characteristics

Purpose(s). As discussed earlier, most states reported that their ELP assessment serves multiple purposes such as identifying ELL students, determining levels of proficiency for instructional placement, redesignating ELL status, and providing diagnostic information for instruction. Similarly, each of the four consortia reported that their ELP assessment is designed to serve multiple purposes of determining ELP progress, providing diagnostic information, and redesignating ELL students. Appendix H presents the purposes of each ELP assessment that test developers or the states identified in their Web site. Almost all assessments are explicitly used for the purpose of annually assessing students' progress in English language development. Some variations were found in relation to whether the ELP test was also used for identification and/or placement. In addition, some states specified one of the purposes of the test was to provide diagnostic information or instructional planning. Considering that a separate validation argument must be made for each intended purpose, the complexity of the validation process is unavoidable.

Construct(s). In general, the constructs of the ELP assessments recently developed by consortia or states are largely based on states' ELP standards. At the same time, the constructs of the assessments take the academic content standards (i.e., English language arts, mathematics, and science standards) into consideration since NCLB stipulates that ELP standards should be aligned with academic content standards. Each assessment has multiple components in order to measure the four language domains and comply with NCLB requirements. The assessments typically have different forms for separate grade bands, K–2, 3–5, 6–8, and 9–12. All reviewed ELP assessments defined the constructs for each grade band and language domain. We also found that five states did not administer an ELP

assessment for Grades K or 1 at the time of this review. Overall descriptions of the constructs of each assessment are illustrated in Appendix I.

These newly developed ELP assessments are intended to measure both academic and social language. Our review found that in practice, the constructs tended to be defined within the context of academic content areas. This finding suggests that states and/or test developers are incorporating the NCLB stipulation described above. For example, ACCESS for ELLs® describes its construct of four language domains (listening, speaking, reading, and writing) in the context of five content areas: social and instructional, English language arts, mathematics, science, and social studies. By doing so, the assessment attempts to appropriately measure students' academic English proficiency. The review of the states' ELP standards [13] revealed that all states mentioned academic English in their standards. Commercial tests also seem to follow the trend of including academic English in their constructs (e.g., New IPT, LAS Links). Some newly developed commercial ELP tests mentioned that they used states' new ELP standards during development of the tests. Furthermore, the developers assert that the new assessments measure the language ability required for academic contexts (Ballard & Tighe, 2004; CTB/McGraw-Hill, 2006).

During our review, one noticeable finding was the varied definitions of academic English proficiency in the states' ELP standards and assessments. A closer look at the descriptions of the standards revealed little specificity about academic English. Moreover, states' definitions of academic English in their standards rarely addressed the multifaceted academic English characteristics that previous research has proposed. The variation of state standards in defining academic English reflects the lack of consensus from the research community on its definition (See the *Literature Review*—CRESST Tech. Rep. No. 731 [Wolf, Kao, et al., 2008] for various perspectives taken to define academic language). Some researchers focus on linguistic features (e.g., grammar, lexis), and others highlight cognitive aspects and contextual features (e.g., strategy use, schemata/background experiences, higher order thinking, metalinguistic awareness). The lack of a common framework to describe academic English poses a challenge to practitioners attempting to operationalize the constructs into their standards and assessments. A movement to develop a practical framework of academic English for school settings has recently arisen to encompass various aspects of academic English (Bailey, 2007; Butler, Lord & Bailey, 2004; Scarcella, 2003; Schleppegrell, 2001). However, our review of the constructs defined in state standards and

---

[13] A total of 47 states' ELP standards were available on their Web sites and reviewed.

assessments in practice revealed that the proposed framework has not yet been actively utilized.

To illustrate the variations among descriptions of academic English in practice, some descriptors of proficiency levels in ELP assessments are presented below. The selected descriptors define proficient/advanced levels in reading across three different types of assessments: (a) a consortium-developed assessment, (b) a commercially-developed assessment, and (c) a state-developed assessment.

1. ACCESS for ELLs®:

Expanding [level 4 of 5]: Matches summaries with excerpts from genres read orally or in writing (such as mythology, science fiction, or ballads); uses organizational features of texts to compare/contrast ideas; makes inferences from text; identifies figures of speech. [Note. This descriptor is for reading in a science context.]

2. Language Assessment Scales (LAS) Links:

Advanced [level 4 of 5]: Uses knowledge of lower-frequency prefixes to determine word meaning; interprets lower-frequency idioms; recalls stated and implicit details in a variety of genres; reads for specific information in a chart, table, or diagram; determines main idea in fiction and academic texts; identifies character traits.

3. Kansas English Language Proficiency Assessment (KELPA):

Advanced (High) [level 3 of 4]: Reads standard newspaper items addressed to the general reader; reads routine reports and technical materials; gains new knowledge from materials in unfamiliar topics in areas of a general nature; interprets hypotheses; supports opinion and conjectures; able to "read between the lines;" may be unable to appreciate nuances or style.

These examples suggest that academic English constructs are described by listing tasks that occur in academic settings without specifying specific academic language features. Differences in the degree of emphasis on academic language features also imply that the current ELP assessments measure different aspects of academic English proficiency, to some extent. For example, for the reading component, one assessment appears to focus on academic language functions, such as inference and critical analysis, whereas another assessment seems to attach importance to linguistic features by including evaluation of vocabulary development and sentence-level reading ability.

Levels of ELP proficiency. A review of available technical manuals of the ELP assessments indicated that test developers primarily employed a bookmark standard setting procedure, or some modification of the bookmark method, as the primary method of setting

cut scores for each level of proficiency. In this procedure, a panel of qualified experts (e.g., teachers, curriculum specialists) reviews a booklet of items that have been ordered by item difficulty, and the panel judges the difficulty level of each item for each proficiency level. The results provide quantifiable data to determine cut scores.

The states that employ their own ELP assessments have aligned the test proficiency levels with the state ELP standards. For these states, the same terminology and descriptions are used for both ELP standards and assessments. However, the states that adopted consortia or commercial assessments have two sets of proficiency levels; one from the assessment developers and the other from the state ELP standards. Our comparison of states' ELP standards with the proficiency levels defined by their ELP assessments showed a number of mismatches (see Appendix D). Some of the states that are part of a consortium have aligned their state ELP standards with those of the consortium assessment. However, there are other states where the terminology and descriptions differ between the ELP assessment and standards. This mismatch indicates a significant problem in the interpretation of ELP test scores and standards.

Item format and administration. Multiple-choice is the most common item format for all four language domains on all ELP assessments reviewed. For the listening component, we found only one assessment that contains constructed response items. The speaking component of all tests is individually administered with orally constructed response items. For the reading component, all ELP assessments reviewed use only multiple-choice items. For the writing component, constructed response items are the most frequent, although some tests include multiple-choice items as well. The majority of tests are untimed.

Scoring and reporting. Typically, the speaking component is scored locally by the interviewer. Other components that include multiple-choice items are machine-scored. Constructed response items are scored either locally at the school level or at a central level such as by the test developer sites. In general, raw scores are converted to scale scores, and cut scores are applied to determine each student's level of proficiency. The school summary report of the assessment typically includes student counts and percentage in each proficiency level and overall proficiency level by grade. The individual student's report includes the student's scale score on each language domain and proficiency level.

During the review, it was not clear whether the reports contained appropriate information that would serve the identified purposes. Some states explicitly included two types of reports to serve two different purposes: identification and annual progress reporting. However, most states had little information about their reporting practice on their public Web

sites. Although several states identified one of the uses of their ELP assessment as being for diagnostic purposes, it was unclear whether their reports included diagnostic data, that is, detailed information about strengths and weaknesses of students' English language proficiency. The mismatch between supposed purposes and actual test reporting practices or even test design needs further investigation.

Validity Information for the Use of ELP Assessments

Validating the uses and inferences made from assessment results involves producing diverse sources of evidence for each use of and inference made based on an assessment. As discussed in the *Literature Review*—CRESST Tech. Rep. No. 731 (Wolf, Kao, et al., 2008), it is not the test itself, which is validated; rather, validation applies to specific interpretations and uses intended by a particular test. Tests themselves are not valid or invalid; validation is the accumulation of evidence by which particular interpretations or uses of test are justified. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999; hereafter referred to as the *Standards*), validity evidence can be obtained by examining the content of the assessment (usually referred to as content validity), the interrelationships between the assessment and other measures (criterion-related validity), the internal structure of the assessment (construct validity), and the test takers' response processes. Additionally, test fairness or bias is an important component of validity, particularly for ELL students, who are a culturally and linguistically diverse group. During our review, we gathered the publicly available technical manuals and examined the types of validity evidence produced regarding ELP assessments. The primary purpose of reviewing this validation information was to identify common types of evidence and methods being utilized in order to provide practitioners with guidance for their continual validation efforts in the use of new ELP assessments.

Among the ELP assessments currently used for the school year 2006–2007, we found publicly available validation studies on 13 ELP assessments. In all of these validation studies, reliability was seen as a necessary first step in the validation processes. Inter-item reliability was typically estimated using either the Cronbach's alpha internal consistency estimate or Item Response Theory (IRT) methods. Rater agreement, when applicable, and generalizability (G)–theory, were also common approaches to examining reliability. Validity evidence was typically reported using traditional terms such as content validity, construct validity, and criterion validity. The summary of the validity evidence regarding the 13 ELP assessments is presented in Appendix J-1.

In general, there was a notable lack of comprehensive validity information available beyond reliability estimates. Developers typically conducted some type of item bias review and one or two additional studies that could be classed as content, construct, or criterion-related validity studies. The most complete validation work publicly available was conducted by two consortia, WIDA and SCASS, which thoroughly documented their work in all the major areas we investigated. These consortia went further and made clear interpretive arguments explaining how the evidence collected supported the use of their assessments. The validation studies by WIDA and SCASS are summarized in Appendix J-2, which also contains other validation study examples produced by a commercial test developer and state.

In most cases, including some very large-scale tests, important work clearly remains to be conducted and made publicly available. The most noteworthy omissions from most test development work were studies involving criterion-related validity, or comparing test scores against some external standard of proficiency. To some extent, work in this area is complicated by changing conceptions of the nature of academic English proficiency and the lack of a commonly accepted metric for defining this proficiency. Developers who did study criterion-related validity typically used correlations between the new assessment and an older ELP or English language-arts assessment (which may assess rather different constructs), or alternatively used scores on the new assessment to predict ELL classification status or language proficiency level.

Some commonalities were observed in the study of bias or fairness. Most test developers included a bias review conducted by some type of expert panel, either consisting of the test developers themselves or outside experts in bias. In some cases, this qualitative item review was combined with a statistical technique using a differential item functioning (DIF) approach. For example, the Massachusetts ELP assessment developers and the Michigan assessment developers conducted DIF analyses to examine items for statistical evidence of gender and ethnic bias. Items that were identified as possibly biased were then examined by the expert panel.

One disconcerting aspect of the published studies was a relative lack of discussion about what decisions were made based on the analyses conducted. Typically, for example, reliability statistics were reported without explicit discussion of what level of reliability was considered minimally acceptable or optimal; in some cases, problematic (low) reliability findings were simply reported and not discussed. Findings of construct or criterion validity studies were often similarly not sufficiently interpreted. For example, some studies used scores on the new assessment to predict test takers' current ELL classification; quantitative

results were provided but discussion of whether or why the results supported test use were missing from the technical reports.

As pointed out, more work on test validity is needed. Specifically, evidence based on content and interrelation with other measures should be systematically provided. For instance, the degree of correspondence between the given ELP assessment content and the state's ELP and content standards is a critical piece of information in making valid inferences based on the assessment. To provide a comprehensive picture of a validation process, an example of using an ELP test with dual purposes is illustrated in Figure 6.

Validity Process

AN ELP ASSESSMENT

PURPOSE/USE 1
The ELP assessment is used to determine the level of students' English language proficiency.

PURPOSE/USE 2
The ELP assessment is used to annually measure the progress of students' English language proficiency.

CLAIM 1
The test scores adequately represent the students' knowledge and skills in the ELP standards and can be used as a basis to distinguish students' academic English language proficiency.

CLAIM 2
The test scores provide a basis to indicate the students' ELP progress.

Identifying A Specific Intended Purpose

Making a Validation Claim

Collecting Diverse Sources of Evidence

Ongoing, regular, and systematic validity investigation.

Targeted revision and improvement based on validity evidence.

SOURCES of EVIDENCE 1

1) Content-related evidence: examining the alignment between the characteristics (e.g., content coverage, cognitive demands) of test items and those of the ELP standards.

2) Criterion-related evidence: investigating the students' performance at the different levels of proficiency on the ELP test compared with performance on English language-arts tests.

3) Construct-related evidence: examining the extent to which the test items conform to measure the intended construct.

4) Evidence based on students' response process: examining whether the students are using knowledge and skills that the test is intended to measure.

5) Fairness-related evidence: examining whether a test item functions differentially for students who have the same ability level for the construct being measured.

SOURCES of EVIDENCE 2

Aside from the evidence collected for Claim 1, the following sources of validity evidence will be important to provide:

1) Content-related evidence: examining the alignment of the content between the previous and the current ELP tests.

2) Criterion-related evidence: investigating the extent to which the progress determined from the ELP test results has predictive power for students' performance on content-area tests; examining longitudinally the growth pattern of the students at the different levels of proficiency; examining the consistency of determining the levels of proficiency over the years.

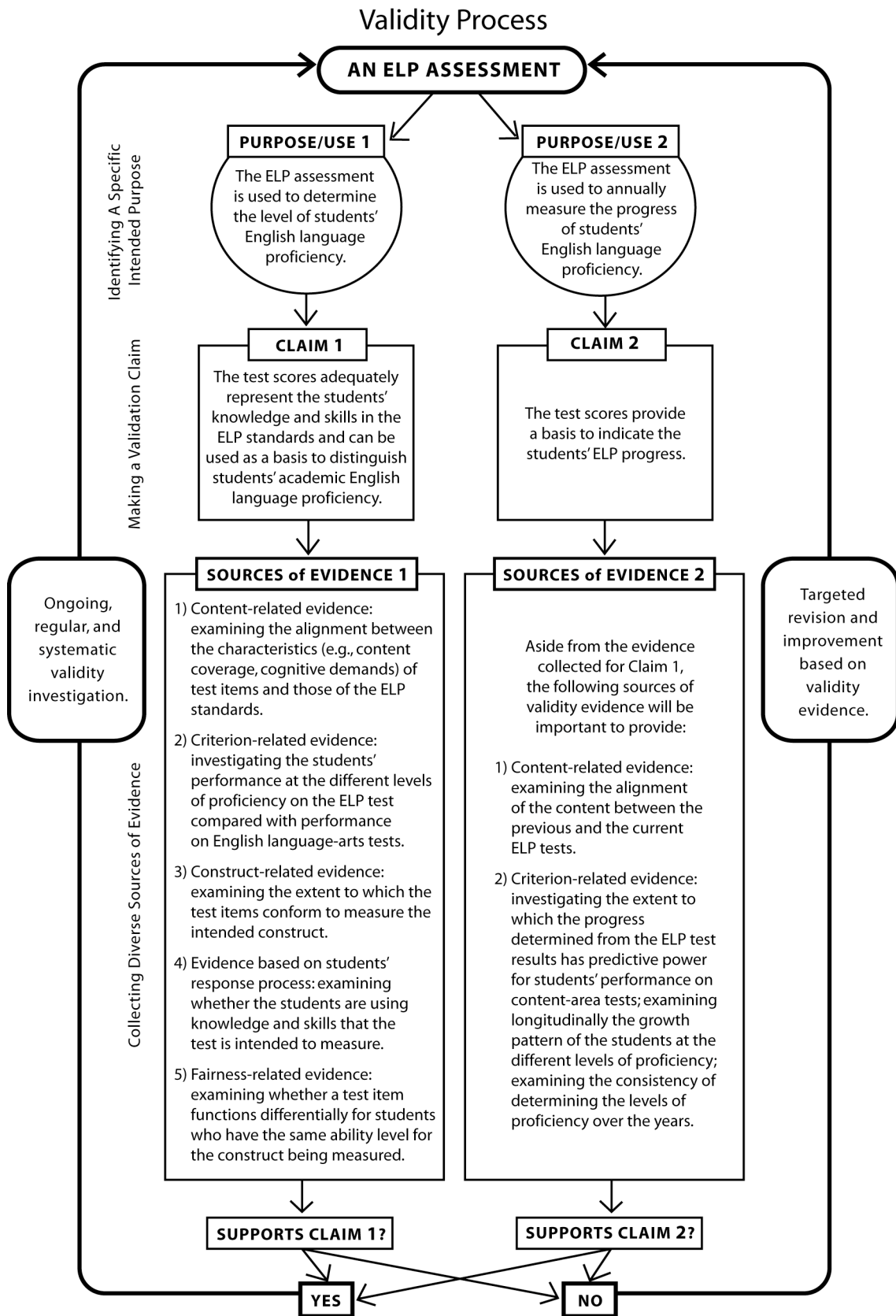**SUPPORTS CLAIM 1?**

SUPPORTS CLAIM 2?

YES

NO

Figure 6. Example of collecting validity evidence in using an ELP test.

Summary

In summary, the majority of states have recently begun to implement newly developed ELP assessments to meet the NCLB mandates. One emerging trend was the inclusion of the characteristics of academic English proficiency in the constructs. Many states have also developed their ELP standards with an emphasis on academic English. This trend is a result of the criticism that previous language proficiency assessments failed to predict ELL students' readiness for mainstream classrooms mainly due to their social language constructs. However, a closer look at the states' ELP standards and the constructs defined in the assessments reveals varying degrees of specifying the nature of academic English proficiency. The extent to which these various new ELP assessments measure the constructs of academic English poses a validity concern. Considering that the majority of states used their ELP tests as a primary source to identify ELL students, determine their level of proficiency, and redesignate their ELL status, the issues of variation in and comparability of these assessment results are notable. Another notable issue is the mismatch in proficiency levels established in the standards and those used in the ELP assessments. While states are establishing the technical adequacy of the use of their new ELP assessments, issues of comparability and alignment in the constructs also need to be addressed in their validation arguments.

Policies and Practices on Assessing ELL Students' Attainment of Content Standards

As required by law, all states now implement standardized, standards-based tests in reading/English language arts and mathematics for Grades 3 to 8, plus at least one high school grade level. Most also have now begun to assess science. Under NCLB, ELL students must be included in these standardized assessments and must be provided appropriate accommodations in order to accurately measure their content knowledge.

Our review primarily focuses on the accommodations used in states' standards-based testing for AYP calculation. Reading/English language arts and mathematics scores are used to calculate AYP. Twenty-three states also administer a high school exit examination for Grades 9, 10, or 11, which also serves as the states' high school test for NCLB purposes. Most states were found to have accommodation guideline documents for those tests used for NCLB requirements. However, the reviewed state documents often did not distinguish which specific accommodations were allowed for which specific assessments. Hence, a summary of the policies on the use of accommodation is provided at a general level in this section.

Inclusion of ELL Students in Testing

Policies on the inclusion of ELL students in testing for AYP reporting were reviewed first. The review found that all states have adopted NCLB's mandate for including all ELL students in the statewide assessments regardless of English proficiency levels. New Title I regulations do allow states to exempt recently arrived ELL students (students who have attended schools in the United States for 12 months or less) from AYP calculations (U.S. DOE, 2006, www.ed.gov/admins/lead/account/lepfactsheet.html). Recently arrived ELL students are allowed to be exempt from assessments in reading/English language arts but are required to be included in mathematics assessment (which does not have to be given in English). In accordance with this regulation, many states do follow the exemption for recently arrived ELL students, and some allowable variations pertaining to the reading/English language arts exemption were observed. Some states also delineated specific policies for the allowed exemptions. For example, one state also takes the student's English proficiency level into consideration to make an exemption ruling. In another state, school districts are allowed to make appropriate decisions on an individual basis. In other states, recently arrived ELL students are counted only for participation rates.

Decisions for Providing Accommodations

The present review gathered information on states' decision makers for determining accommodations for ELL students and what accommodations are allowed. The review also examined whether states' allowable accommodations were listed separately for ELL students from students with special needs and what criteria were used for accommodation decision making. Generally, we found variation in decision making across states.

Decisions on which accommodations ELL students were allowed to use were often part of a school- or district-level decision-making process. In the present review, we found that 23 states specified a school-level process, and 11 states specified a district-level process. Six states designated local teams or committees to make the determinations; however, it was unclear whether these local teams were at the school or district level. Furthermore, even if a state policy mandated district-level decision making, that did not preclude the district from then delegating the decision-making process to each school. At the time of this review, the remaining 10 states had no publicly available information about who was delegated to make decisions regarding accommodations for ELL students.

State policies regarding allowable accommodations for ELL students also differed. Out of the states reviewed (49 states and the District of Columbia), 41 listed specific accommodations allowable for ELL students. In other words, in lists or descriptions of

allowable accommodations, accommodations that were specifically allowed for ELL students (and not for students with disabilities) were noted as such or even listed separately. Eight states listed allowable accommodations for all students with special needs but did not distinguish which ones were specific for ELL students. In two of those cases, accommodations were allowable for all students, regardless of ELL or disability status, but their use was determined on an individual, case-by-case basis. (Other states also underscored the need for decisions to be made on an individual basis). In one state, allowable accommodations were not publicly available. States that did not distinguish accommodations between subpopulations were not necessarily states with the lowest numbers of ELL students.

Many states mentioned multiple criteria in determining which accommodations were allowed for ELL students on state academic assessments. However, specific guidelines on how to apply the criteria and how to select a particular accommodation were rarely noted. Rather, the guidelines and criteria were described at a general level. For example, 29 states mentioned that accommodations should be made on an individual basis, and 22 states said that the accommodations used during testing had to be the same accommodations as those used during classroom instruction and assessment. Other selection criteria required by some states were English proficiency level of the student (as determined by the state-approved ELP assessment) and number of years a student has been in school and designated as an ELL. Additionally, some states required documentation of allowed accommodations to be placed in the ELL student's file.

A few states also underscored prohibited accommodations, including some that were allowed in other states. For example, one state did not allow assessments to be provided in a language other than English. This state also prohibited the language in test directions to be reduced in complexity. Another state prohibited teachers from reading any words to the student from the test items in the reading test.

Some states also differentiated between standard and nonstandard accommodations, referring to accommodations that do not alter the construct, and those that do. However, information on how this was determined was not available. States varied in the amount of detail available to the public regarding allowable accommodation types, ranging from very long lists to listing a total of four accommodation strategies. Some states also specified which accommodations were allowed for which specific assessment and content area. In a couple of states, accommodation types were only listed as examples for districts or schools to follow as general guidelines.

As described, the policies on providing accommodations varied widely across and within states, largely due to the degree of specificity and clarity provided by states. We assume that policies can vary within states when states indicated allowing school districts to make decisions. Some states' policies appeared to lack information regarding designated decision makers, ELL-specific accommodations, and criteria to determine valid accommodations and prohibited accommodations.

Types of Accommodations

Specific accommodations allowed for ELL students varied among the states reviewed. It must be noted that these accommodations were listed by states as allowable (and in some cases, only as examples) and does not necessarily mean they are actively used in practice. In this review, we summarize the types of allowable accommodations under four general categories: Changes to the Timing/Scheduling, Changes to the Setting, Changes to the Presentation, and Changes to the Response. These four categories are considered traditional because they are commonly used for students with disabilities.

Rivera et al., (2006) defined the four traditional categories and discussed the current trend toward a new taxonomy with only two categories specifically for ELL students. According to Rivera et al., (2006) in this new taxonomy, ELL accommodations fall under two distinct areas: direct linguistic support and indirect linguistic support. Direct linguistic support accommodations are linguistic modifications that alter the language of the test to reduce linguistic complexity while maintaining the construct of the test. Direct linguistic support accommodations are further divided into two subcategories: native language accommodations and English language accommodations (Rivera et al., 2006). English language accommodations simplify, repeat, or clarify some or all of the test language. ELL accommodations for direct linguistic support are often found under the traditional categories Changes to the Presentation and Changes to the Response.

Indirect linguistic support accommodations adjust the conditions under which ELL students take academic assessments. These accommodations account for the fact that there are greater linguistic demands placed on ELL students during testing situations, and these students need other types of supports to help them with processing the language and information on academic assessments. Some examples of indirect linguistic supports include extended time or change in location during testing. Such accommodations are not based on the actual language of tests, but they support the ELL student in negotiating the linguistic demands of tests. For example, if a state allows the use of a dictionary during testing—a direct linguistic support—then the state should also allow extended time on the test—an

indirect linguistic support—so that the ELL student will have time to look up words in the dictionary. Thus, indirect linguistic supports can go hand in hand with direct linguistic supports. Indirect linguistic supports are typically found under the traditional categories Changes to the Timing/Scheduling and Changes to the Setting. In this review, we chose to provide information in the four traditional categories to facilitate its presentation, and also because many accommodations traditionally allowed for students with disabilities are still listed as allowable for ELL students.

Changes to the timing/scheduling. As shown in Figure 7, the most frequently allowed accommodation that involved changes to timing was extended or extra time, which was mentioned by 38 states' policies. (Note that many state assessments are untimed and all students are allowed reasonable amounts of time to complete assessments.) Other types of accommodations in this category include frequent breaks (allowed by 32 states), testing over multiple or several sessions or days (allowed by 24 states), flexible scheduling (allowed by 13 states), and administering the test at a time most beneficial to the student (allowed by 13 states). See Appendix K-1 for specific accommodations used by states in this category.
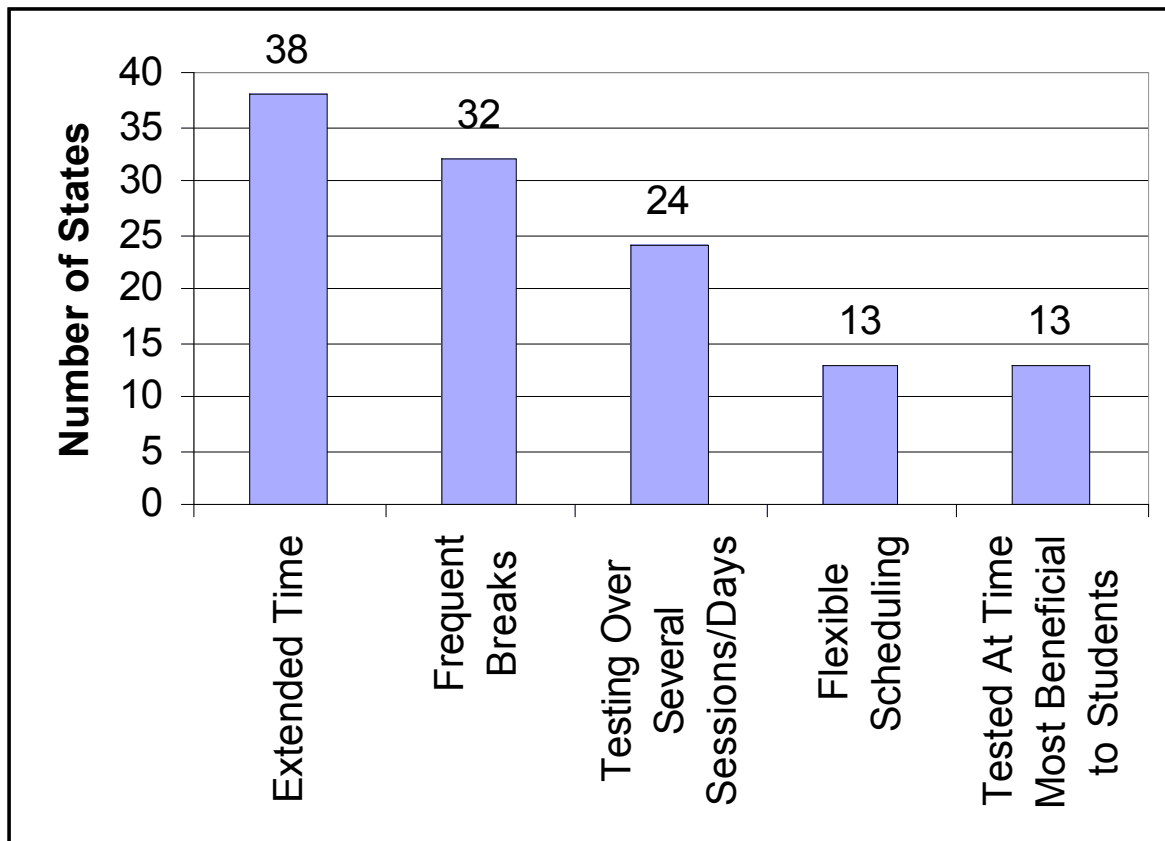

Figure 7. Number of states with accommodations involving changes to the timing or scheduling, by accommodation type.

Changes to the setting. The most common form of accommodation that involved changes to the setting was administering the test in a small group, which was allowed by 37 states (see Figure 8). Other types of changes to the setting included administering the test in a separate location, including in a study carrel, non-regular classroom, or somewhere quiet with minimal distractions (31 states); administering the test individually or one-on-one (allowed by 29 states); administering the test by other school personnel, such as ESL or bilingual teacher or person familiar with test taker (17 states); providing the test taker with preferential seating (14 states); providing special or adaptive classroom equipment, furniture, or lighting (9 states); administering the test in an ESL or bilingual classroom (6 states); and administering the test in a non-school setting, such as at home (5 states). Some states also mentioned other types of changes to the setting that can be provided to ELL students, such as the opportunity for the student to move, stand, or pace; supporting the physical position of the student; test administered with teacher facing student; and test administered with student seated in front of the classroom. See Appendix K-2 for more detail.
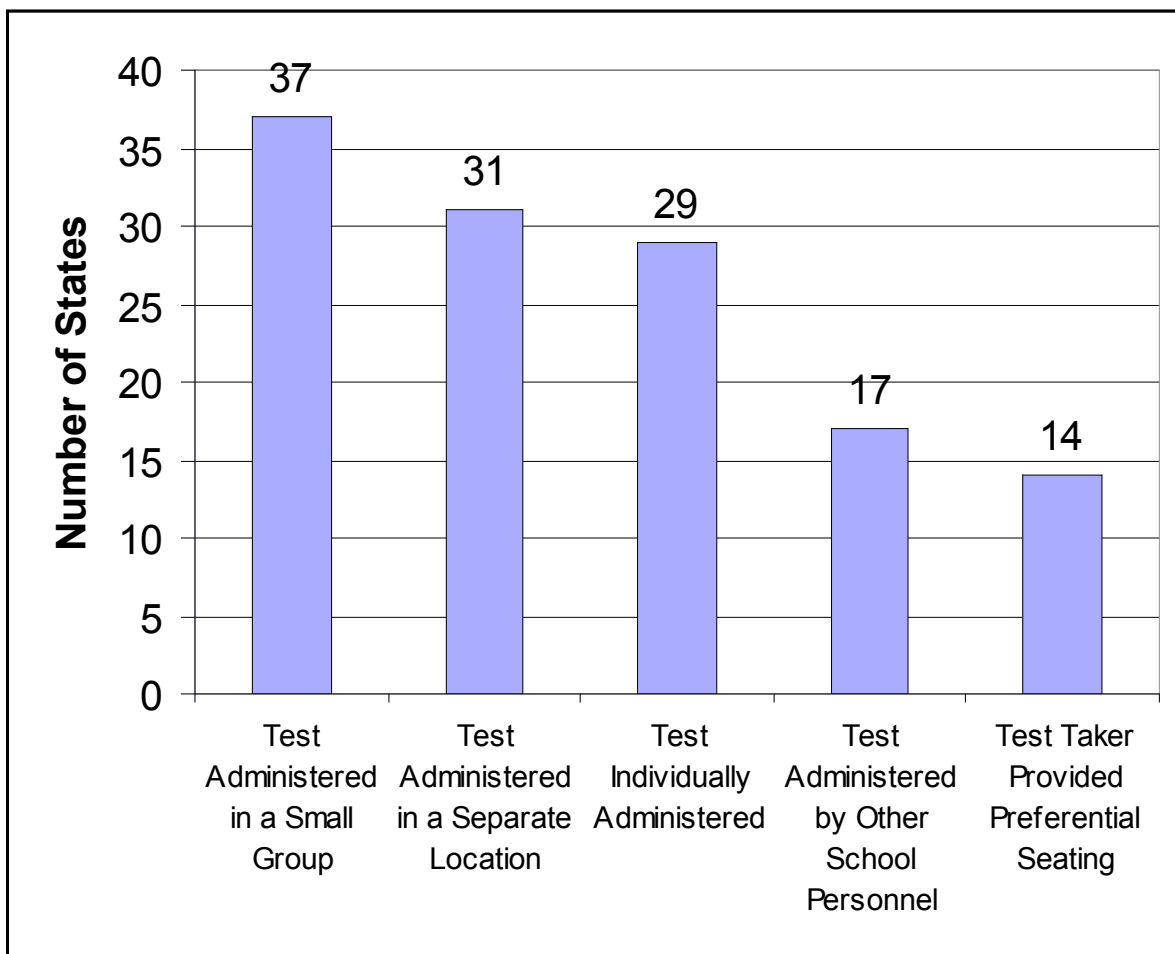

Figure 8. Number of states with accommodations involving changes to the setting, by accommodation type.

Changes to the presentation. As mentioned above, the accommodations in this category often offer direct linguistic support by providing native language translations (written and oral) or English language simplification, repetition, or clarification. Accordingly, native language accommodations include directions translated (oral or written), dual language or side-by-side bilingual test versions, translated or native language test versions, and test items read aloud in native language. English language accommodations include directions read or repeated aloud in English; test items read aloud in English; directions simplified or paraphrased, directions clarified or explained; and bilingual or English-only dictionaries, glossaries, or word lists.

As illustrated in Figure 9, the most commonly allowed accommodation was an English language accommodation of providing bilingual dictionaries, glossaries, or word lists to clarify the test language, which was allowed by 43 states. Most states that allowed the use of these accommodations emphasized that they must be "word-to-word" or "direct translation" dictionaries or glossaries that do not provide definitions or explanations. Some states also allowed the use of English dictionaries, word lists, or glossaries. However, states did not specify the type of dictionaries, word lists, or glossaries or whether they were standardized. Reading test items aloud in English to students was also a common presentation-type accommodation (allowed by 39 states). Other types of accommodations involving presentation included reading or repeating aloud test directions in English (33 states); translating directions, either oral or written (28 states); simplifying or paraphrasing directions (18 states); clarifying or explaining directions (14 states); using audio recording or similar technology to present the test (15 states); providing translated or native language tests (10 states); providing dual language or side-by-side bilingual test versions (9 states). Thirty-two states also listed other types of accommodations relating to changes in presentation, which included using place markers, highlighting key words or phrases, providing electronic translators, allowing the student to self-vocalize, using templates or color overlays, administering the test using sheltered English, providing written versions of oral directions, and using noise buffers. See Appendix K-3 for more detail.
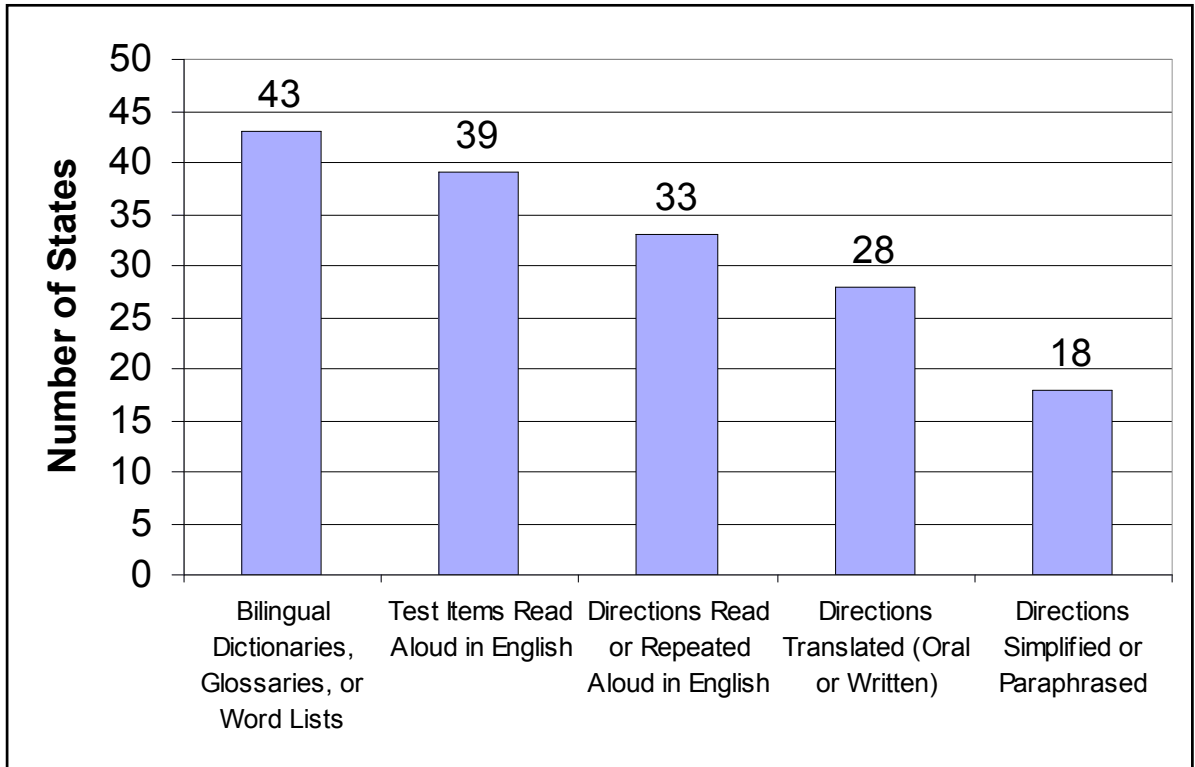
Figure 9. Number of states with accommodations involving changes to the presentation, by accommodation type.

Changes to the response. These accommodations often provide direct linguistic supports. In general, for these accommodations, an ELL student is allowed to respond in writing or orally in his or her native language and a translation is made of the response. Accommodations that involved changing ways a student responds was the least common category of the four, with just 32 states allowing accommodations that fall into this category. As seen in Figure 10, accommodations involving changes to response included having the test taker dictate to or use a scribe to respond in English (20 states); having the test taker mark answers directly into the test booklet (16 states); having the test taker use a computer, word processor, or typewriter to respond (13 states); having the test taker dictate to a scribe to respond in the test taker's native language (7 states); verifying the test taker's understanding of directions (6 states); having the test taker point to or indicate response (4 states); and providing a spelling or grammar checker (3 states). See Appendix K-4 for more detail.
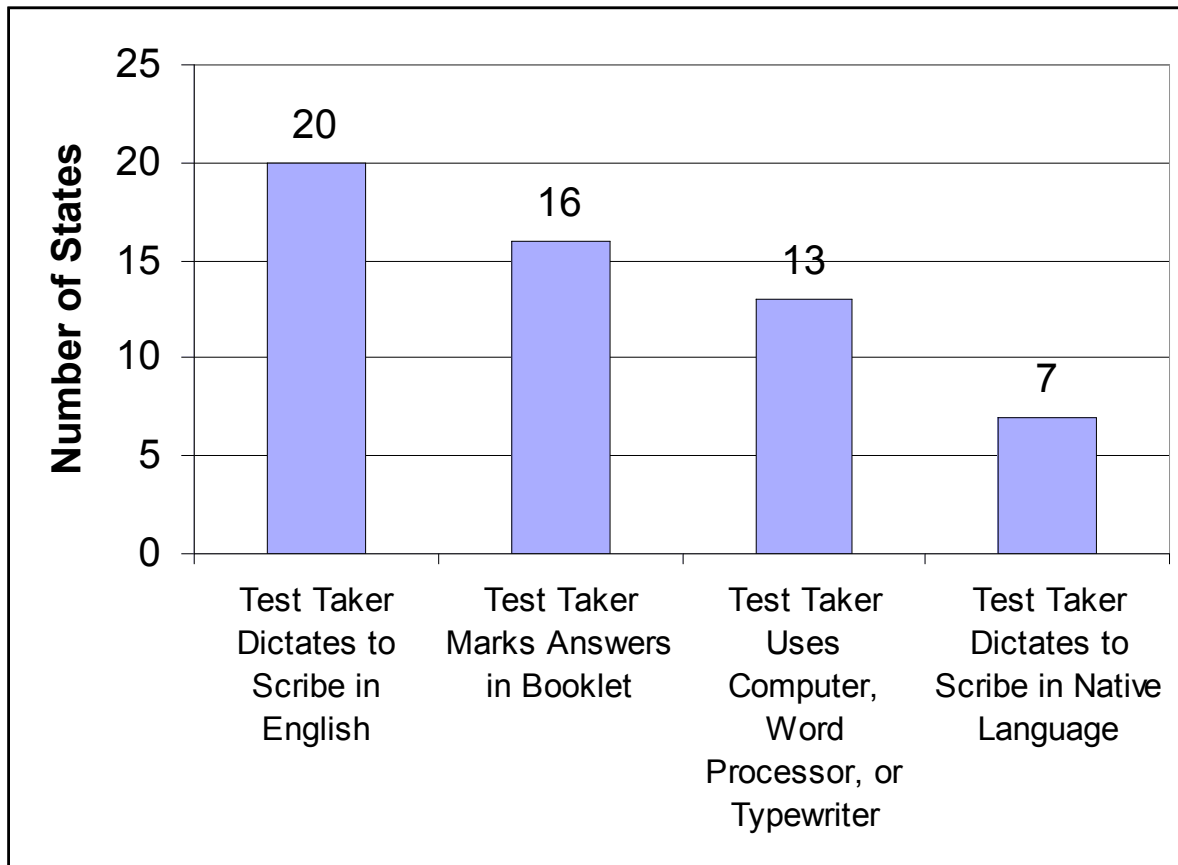
Figure 10. Number of states with accommodations involving changes to the response, by accommodation type.

Summary

In summary, the most popular accommodation that states listed as allowable was the use of bilingual dictionaries, glossaries, or word lists (43 states), which are related to a direct linguistic support accommodation. While this appears to be in contrast to previous research (e.g., see Rivera et al., 2006), which cites extra time as the most common, many state assessments are now untimed, and states may not specifically list extra time as an allowable accommodation for ELL students because all students are allowed as much time as needed (with certain restrictions). However, it is also a promising indication that more states are moving toward providing direct linguistic support as accommodations for their ELL students.

Recent studies found that the use of customized glossaries was a highly effective and valid accommodation for ELL students (see the *Literature Review*—CRESST Tech. Rep. No. 731 [Wolf, Kao, et al., 2008] for more detailed explanation). Other studies have shown that linguistic modifications that reduce the linguistic demands of the text were highly salient accommodations for ELL students, especially for math assessments (see the *Literature Review*—CRESST Tech. Rep. No. 731 [Wolf, Kao, et al., 2008]). However, only one state

openly mentioned that its math assessment for ELL students simplifies the linguistic demands of the test. Eighteen states allow test directions, not items, to be simplified or paraphrased. Although common accommodation types were observed across states (e.g., use of dictionaries, glossaries, read aloud), the exact way of implementing those accommodations was not specified based on our review. This again raises an issue of comparability, as well as validity, of accommodated test results both across and within states. It is also noteworthy that, again, some accommodations allowed in one state may be prohibited in another state (e.g., assessment in native languages), which poses a question of fairness.

As discussed earlier, the variability in procedures of selecting and implementing accommodations across and within states implies that the issue of comparability lies both in the accommodated test results and in the process of applying various accommodations.

Recommendations in Assessing ELL Students

The present review analyzed and summarized the policies and practices on the assessment of ELL students. Although some variations in states' policies were expected given states' specific ELL populations and different assessment systems, serious concerns were raised about threats to the validity of ELL testing. Among the areas reviewed, considerable variation was found across and within states with respect to defining ELL, redesignating ELL students, identifying ELP assessments, and determining the provision of accommodations. This variation may have resulted in part from the information provided in guidelines outlined by the states, which in many cases was limited or not specific enough. Another reason for variation is the fact that some states allowed local districts or schools to make academic decisions about ELL students without statewide guidelines (e.g., for identification, redesignation, and accommodation provision). This issue poses more complexity and difficulty when making a validation argument in assessing ELL students. In order to improve the quality of the assessment policies and practices, we suggest the following recommendations for each area reviewed.

1. States should define clearly the terms they use to refer to ELL students, making it explicit whether they include newly redesignated ELL students as well. We found that this was not always the case in current practice. Procedures for identifying ELL students and determining their levels of English language proficiency are considerably different across states, owing to the different sources 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.

2. By the same token, states should provide a policy for redesignating ELL students. Our review found that variations in this area are substantial. The number of criteria for making a redesignation decision ranged from one to seven across states. Moreover, many states leave the decision-making process to local districts without clear guidelines. Considering the consequences of the policies for redesignated ELL students (e.g., AYP reporting, accommodation policies), comparability issues arise in this area. In addition, our review revealed that it was unclear as to whether all states monitor newly redesignated ELL students and maintain a tracking system. 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.

3. As reported, the majority of states adopted newly developed ELP assessments within the past 2 years. With regard to using these new ELP assessments and making validation arguments, we recommend considering the following issues, as raised by our review.

   • Purposes: States should clearly identify the intended uses of ELP tests. It was found that not all assessments' purposes were publicly documented. Given that making a validation argument begins with the consideration of the assessment purposes, identifying the purposes is a prerequisite step for valid use of the assessment. It was also found that many states used their ELP test for multiple purposes. States should not only make a validation argument for each purpose but also ensure the appropriate use for each intended purpose. As noted, our review found that some states were not clear on how they reported the test results for each purpose.

   • Alignment: States should examine whether the constructs of their ELP assessments are aligned with constructs in their ELP standards. NCLB legislation makes it clear that states should provide an academic assessment of English proficiency. It also stipulates that the constructs should be aligned with the state's ELP standards in order to measure the progress of appropriate English language development. We found that there were different degrees to which states described the standards in incorporating the characteristics of academic English. Even though some states use the same ELP assessment (i.e., within consortia), their ELP standards were not necessarily the same. An examination of an alignment of the constructs will also provide informative guidelines for instruction of ELL students.

   • Levels of proficiency: States should have comprehensible guidelines to align and compare the levels of proficiency defined by the state's standards and the ELP assessment. Surprisingly, there are different numbers of levels and labels between the standards and the assessment for many states. As indicated above, having clear guidelines on the alignment between the two sets of proficiency levels is urgent in order to establish valid use of the test results. This will also provide useful information for test developers and classroom teachers.

4. States should provide comprehensive and specific accommodation guidelines for the effective use of accommodations for local districts and schools. Specifically, clear guidelines are needed for what and how accommodations should be implemented. In some cases, during our review, it was unclear who made decisions about accommodations. It seemed likely that decisions about what accommodations were available and who should receive them could vary substantially by district and school, raising serious problems in comparability of scores. Although most states documented a list of allowable accommodations and specified that those accommodations should be familiar to students, more information is needed about which accommodations would be effective for what kinds of ELL students. In the use of accommodations, states should consider the following aspects:

- States should provide detailed guidelines in making accommodation decisions. The guidelines should include who may be the decision maker(s), which students should be provided with accommodations, and what types of accommodations should be provided during what conditions. The conditions may include information about students' language proficiency levels and the content areas of the tests.

- States should specify both allowable and prohibited accommodations. The criteria for this classification may include the invariant construct and valid results without affecting the performance of students who do not need accommodations. Delineating detailed accommodations will be particularly important for states that allow local districts to make a provision decision.

States should have explicit criteria for determining the selection of accommodations. For example, principled use of ELP test results may provide a systematic procedure during accommodation decision making for academic achievement assessments. Inferences drawn from a student's poor performance on academic vocabulary may lead to implementation of a bilingual glossary with a subsequent science assessment, for instance.

A promising change is the development of a taxonomy for identifying ELL students and determining and keeping track of accommodations allowed for each individual student. The Selection Taxonomy for English Language Learner Accommodations (STELLA) is being developed by researchers at the University of Maryland, in collaboration with the South Carolina Department of Education. STELLA is a computerized decision-making system to help practitioners define ELL and identify ELL students, as well as match these students to the appropriate accommodations, and is expected to be implemented on a wide scale in the near future (Kopriva & Carr, 2006; Zehr, 2007).

And in all recommendations above, we feel that guidelines should be accessible (i.e., clearly written and easily found) by practitioners and the general public alike. Although many states appeared to make efforts to share policies regarding ELL students on their state

department of education Web sites, sometimes information was not easily found or was missing.

In addition to these recommendations for practitioners, we have the following recommendations for researchers.

1. Expand 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*—CRESST Tech. Rep. No. 731 (Wolf, Kao, et al., 2008), research has suggested that measuring academic English is necessary to better predict ELL students' readiness for mainstream classrooms. Although many 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 language demands of various subject areas at various grade levels for all four language domains. The findings from this line of research will provide some useful guidelines for states' policymakers and test developers to examine the constructs of their assessments and standards.

2. Investigate the confounding interaction between subject-area content and linguistic content. Further validity study should be included to investigate potentially confounding interactions between subject-area content and the linguistic content intended for ELP assessment. It has been argued that linguistic complexity may cause construct-irrelevant variance when measuring ELL students' content knowledge and skills. To what extent linguistic and content knowledge influence ELL students' performance both on content and ELP assessments remains unanswered.

3. Continue and expand accommodation research. Although many states provide language-related accommodations, their effects are inconclusive across research. The types of accommodations being examined to date are also limited. The current review found that some accommodations that many states listed as allowable have not always been investigated in previous research. Although many states endeavored to delineate ELL-specific accommodation types, some states still adopted accommodations based on the needs of students with disabilities. An expansion of research in ELL-specific accommodation types is critical to provide empirical evidence that practitioners can use to make provision decisions.

References

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

Abedi, J., Nambiar, V., & Porter, S. (in preparation). *English Language Proficiency Assessment in the nation: Current status and future practice*. Davis: University of California.

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

Bailey, A. L. (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.

Ballard & Tighe, Publishers. (2004). *IPT: The new Title III testing system: Language proficiency testing progress report*. Brea, CA: Author. Retrieved July 21, 2007, from http://www.nclb.ballard-tighe.com/

Buter, F. A., Lord, C., Bailey, A. (2004). *An approach to operationalizing academic language for language test development purposes: Evidence from fifth-grade science and math* (CSE Tech. Rep. No. 626). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J. S., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute. Retrieved July 6, 2007, from http://www.urban.org/publications/311230.html

Chamot, A. U. & O'Malley, J. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley.

CTB/McGraw-Hill. (2006). *LAS Links^{TM}: The technical manual*. Monterey, CA: Author.

Cummins, J. P. (1983). Language proficiency and academic achievement. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 108-126). Rowley, MA: Newbury House.

Hopstock, P. J. & Stephenson, T. G. (2003). *Descriptive study of services to LEP students and LEP students with disabilities. Special topic report #1: Native languages of LEP students.* Arlington, VA: Development Associates, Inc. Retrieved July 6, 2007, from http://www.ncela.gwu.edu/resabout/research/descriptivestudyfiles/native_languages1.pdf

Jepsen, C. & de Alth, S. (2005). *English learners in California schools.* San Francisco: Public Policy Institute of California. Retrieved May 8, 2007, from http://www.ppic.org/content/pubs/report/R_405CJR.pdf

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000-2001 Summary report*. Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.

Koenig, J. A. & Bachman, L. F. (2004). *Keeping score for all: The effects of inclusion and accommodation policies 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.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Northwest Educational Regional Laboratory. (2006). *The next step: Assessment and the English language learner*. Portland, OR: Author. Retrieved August 10, 2007, from http://www.nwrel.org/nwedu/11-03/step/

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.

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

Scarcella, R. (2003). *Academic English: A conceptual framework* (Tech. Rep. No. 2003-1). Santa Barbara: University of California, Linguistic Minority Research Institute.

Schleppegrell, M. J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, *12,* 431-459.

Wolf, M. K., Kao, J., Herman, J. L., Bachman, L. F., Bailey, A. L., Bachman, P. L., et al. (2008). *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., Herman, J. L., Bachman, L. F., Bailey, A. L., & Griffin, N. (2008). *Recommendations for assessing English language learners: English language proficiency measures and accommodation uses—Recommendations* (CRESST Tech. Rep. No. 737). 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

Appendix A – K4

(pgs 40–92)

Appendix A

Number and Percentage of LEP Enrollment Students
(2004–2005) [a] and Top Language and Percent Spoken,[b] by State

State	Number of LEP enrollment students	Percent of LEP enrollment students	Top language spoken by LEP students	Percent of LEP students speaking top language
AK	20,140	15.1	Yup'ik	38.6
AL	15,295	2.1	Spanish	74.7
AR	17,384	3.8	Spanish	87.6
AZ	155,789	15.1	Spanish	85.6
CA	1,591,525	25.7	Spanish	83.4
CO	90,391	11.8	Spanish	81.8
CT	27,580	4.8	Spanish	67.6
DC	4,771	6.4	Spanish	76.4
DE	5,094	4.3	Spanish	72.3
FL	299,346	11.3	Spanish	75.8
GA	50,381	3.2	Spanish	70.1
HI	18,376	10.1	Ilocano	31.8
IA	14,421	3.0	Spanish	62.3
ID	17,649	6.9	Spanish	78.8
IL	192,764	9.2	Spanish	77.6
IN	31,956	3.1	Spanish	64.4
KS	23,512	5.3	Spanish	81.3
KY	11,181	1.8	Spanish	47.3
LA	7,990	1.1	Spanish	48.5
MA[c]	49,923	5.1	Spanish	69.4
MD	24,811	2.9	Spanish	53.0
ME	2,896	1.4	French	16.8
MI	64,345	3.7	Spanish	44.8
MN	56,829	6.8	Hmong	34.1
MO	15,403	1.7	Spanish	44.2
MS	4,125	0.8	Spanish	60.4
MT	6,911	4.7	Blackfoot	25.2
NC	70,288	5.8	Spanish	77.6
ND	4,749	4.7	Native Am. (unsp)	85.9
NE	16,124	5.6	Spanish	76.8
NJ	61,287	4.4	Spanish	67.3

State	Number of LEP enrollment students	Percent of LEP enrollment students	Top language spoken by LEP students	Percent of LEP students speaking top language
NM	90,926	28.7	Spanish	78.8
NV	72,117	18.1	Spanish	91.5
NY	203,583	7.1	Spanish	62.2
OH	25,518	1.4	Spanish	39.2
OK	33,508	5.3	Spanish	51.7
OR	59,908	10.8	Spanish	72.5
PA	39,847	2.2	Spanish	52.9
RI	10,921	7.0	Spanish	69.8
SC[d]	15,396	2.2	Spanish	77.3
SD	5,847	4.8	Lakota	57.4
TN	19,355	2.1	Spanish	61.2
TX	684,007	15.5	Spanish[c]	93.4[d]
UT	56,319	11.4	Spanish	65.3
VA	67,933	5.6	Spanish	60.4
VT	1,393	1.4	Serbo-Croatian	26.7
WA	75,678	7.4	Spanish	60.9
WI[c]	35,871	4.1	Spanish	47.8
WV	1,236	0.4	Spanish	26.3
WY	3,742	4.4	Spanish	90.4
Reporting States	**4,479,576**	**9.3**	**Spanish**	**79.0**

Note. The abbreviation LEP (Limited English Proficient) is used here to maintain the source terminology.

[a]Source is obtained from http://www.ncela.gwu.edu/stats/3_bystate.htm

[b]Top Language Spoken and Percent, adapted from National Clearinghouse for English Language Acquisition, 2000–2001. Please see Kindler (2002) for reporting procedures.

[c]Data for Massachusetts, Texas, and Wisconsin are from 1999–2000.

[d]Data for South Carolina and Texas are from 2001–2002.

Appendix B

Terms Used by States to Refer to ELL Students

State	LEP	ELL	Other terms used
AK	X	—	—
AL	X	X	—
AR	X	X	—
AZ	X	X	Primary or Home Language is other than English (PHLOTE)
CA	X	—	English Learner (EL)
CO	X	X	Primary or Home Language is other than English (PHLOTE), Non-English Proficient (NEP)
CT	—	X	—
DC	X	X	Primary or Home Language is other than English (PHLOTE), Non-English Proficient (NEP), Language Minority Student (LM), Linguistically and Culturally Diverse Student (LCD)
DE	X	X	—
FL	X	X	—
GA	X	X	Primary or Home Language is other than English (PHLOTE), Non-English Proficient (NEP), English Language Learner Monitored (ELL-M)
HI	X	X	Non English Proficient (NEP)
IA	X	X	—
ID	X	X	—
IL	X	—	Non-English Language Background (NELB)
IN	X	X	Language Minority (LM)
KS	X	X	—
KY	X	X	—
LA	X	—	—
MA	X	X	—
MD	X	X	—
ME	X	—	Non-English Proficient (NEP), Linguistically and Culturally Diverse Student (LCD)
MI	X	X	—
MN	X	X	—
MO	X	X	Language Minority (LM), Linguistically Diverse students (LD)
MS	X	X	Non-English Proficient (NEP), Potentially English Proficient (PEP), Language Minority (LM)
MT	X	X	—
NC	X	X	National Origin Minority (NOM)

State	LEP	ELL	Other terms used
ND	X	X	—
NE	X	X	—
NJ	X	—	—
NM	X	X	Primary or Home Language is other than English (PHLOTE)
NV	X	—	—
NY	X	X	—
OH	X	—	Primary or Home Language is other than English (PHLOTE)
OK	X	X	—
OR	X	X	Primary or Home Language is other than English (PHLOTE)
PA	X	X	—
RI	X	X	—
SC	X	—	Primary or Home Language is other than English (PHLOTE), Language Minority (LM)
SD	X	X	—
TN	X	X	Non-English Language Background (NELB)
TX	X	X	—
UT	X	X	—
VA	X	—	—
VT	X	X	Primary or Home Language is other than English (PHLOTE), Non-English Language Background (NELB), Non-English Proficient (NEP), Transitional English Proficient (TEP)
WA	X	X	Primary or Home Language is other than English (PHLOTE)
WI	X	—	—
WV	X	—	Primary or Home Language is other than English (PHLOTE)
WY	—	X	—
Total	**48**	**38**	**20**

Note. LEP = Limited English proficiency. ELL = English language learner.
— = state does not use.

Appendix C

State ELL Identification Criteria

State	HLS	ELP assessment[a]	Other criteria
AK	X	New IPT	Observation
AL	X	W-APT	School personnel input, previous school records
AR	X	ELDA	—
AZ	X	AZELLA	—
CA	X	CELDT	—
CO	X	CELA	—
CT	X	LAS Links	Interview, previous school records
DC	X	W-APT, Pre-LAS	Interview
DE	X	W-APT	—
FL	X	CELLA	State test performance (Grades 4 and up), school personnel input
GA	X	W-APT	—
HI	X	LAS Links	Parent or school personnel input
IA	X	ELDA	—
ID	X	IELA	—
IL	X	W-APT	—
IN	X	LAS Links	—
KS	X	KELPA	—
KY	X	W-APT	—
LA	—	ELDA	—
MA	X	District chosen	—
MD	X	LAS Links	Interview, parent or school personnel input, previous school records
ME	X	W-APT	—
MI	X	ELPA Initial Screening	—
MN	X	TEAE, MN-SOLOM, District chosen	State test performance, previous school records
MO	X	District chosen	—
MS	X	SELP	—
MT	X	District chosen	State test performance, previous school records, observation
NC	X	New IPT	—
ND	X	W-APT (ACCESS for ELLs® beginning 2007)	Observation, school personnel input

State	HLS	ELP assessment[a]	Other criteria
NE	—	District chosen	—
NJ	X	W-APT, IPT, MAC II or LAS, District chosen	—
NM	X	NMELPA	Observation
NV	X	PreLAS, LAS, District chosen	—
NY	X	LAB-R	Interview
OH	X	District chosen	—
OK	X	W-APT	—
OR	X	OR ELPA	State test performance, parent or school personnel input
PA	X	W-APT, District chosen	—
RI	X	W-APT	—
SC	X	LAS, IPT, Woodcock-Muñoz, District chosen	—
SD	—	LAS, IPT, District chosen	—
TN	X	ELDA[b]	—
TX	X	District chosen	State test performance, school personnel input
UT	X	UALPA, District chosen	—
VA	X	Virginia SELP	—
VT	X	W-APT	Parent input
WA	X	WLPT-II	Observation
WI	X	W-APT, District chosen	—
WV	X	Woodcock-Muñoz	—
WY	X	SELP, WELLA, District chosen	—
Total	**47**	**47**	**16**

Note. HLS = Home language survey. ELP = English language proficiency.
— = state does not use.

[a]This column indicates assessment acronyms. See Appendix G2 for complete assessment names.

[b]Tennessee adopted the ELDA in the Spring of 2007; however, a placement test is still under development, so the state has approved the use of the New IPT or the CELLA for initial screening of ELL students.

Appendix D

English Language Proficiency Levels in State Standards and ELP Assessments

State	Levels in state standards	Levels in ELP assessment[a]
AK	Beginner Low, Beginner High, Intermediate Low, Intermediate High, Proficient, Proficient High	(New IPT) Beginner Low, Beginner High, Intermediate Low, Intermediate High, Proficient, Proficient High
AL	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
AR	ELDA Levels	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
AZ	Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced	(AZELLA) Pre-emergent, Emergent, Basic, Intermediate, Proficient
CA	Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced	(CELDT) Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced
CO	Beginning, Intermediate, Advanced	(CELA) Level 1, Level 2, Level 3, Level 4, Level 5
CT	Beginning, Intermediate, Advanced	(LAS Links) Beginning, Early Intermediate, Intermediate, Advanced, Above Proficient
DC	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
DE	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
FL	Beginning, Intermediate, Advanced	(CELLA) Beginning, Intermediate, Advanced, Proficient
GA	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
HI	Level 1, Level 2, Level 3, Level 4, Level 5	(LAS Links) Beginning, Early Intermediate, Intermediate, Advanced, Above Proficient
IA	Level 1(*Pre-production*), Level 2 (*Early Production*), Level 3 (*Speech Emergence*), Level 4 (*Intermediate Fluency*), Level 5 (*Fluent*)	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
ID	IELA Levels	(IELA) Beginning, Advanced Beginning, Intermediate, Early Fluent, Fluent

State	Levels in state standards	Levels in ELP assessment[a]
IL	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
IN	Level 1(*Beginner*), Level 2 (*Early Intermediate*), Level 3 (*Intermediate*), Level 4 (*Advanced*), Level 5 (*Fluent English Proficient*)	(LAS Links) Beginning, Early Intermediate, Intermediate, Advanced, Above Proficient
KS	Beginning, Intermediate, Advanced	(KELPA) Beginning, Intermediate, Advanced, Fluent
KY	Beginning, Lower Intermediate, Upper Intermediate, Advanced	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
LA	ELDA Levels	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
MA	Beginning, Early Intermediate, Intermediate, Transitioning	(MEPA) Beginning, Early Intermediate, Intermediate, Advanced
MD	Low Beginning, High Beginning, Low Intermediate, High Intermediate, Advanced	(LAS Links) Beginning, Early Intermediate, Intermediate, Advanced, Above Advanced
ME	NA	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
MI	Level 1A (*Basic*), Level 1B (*Basic*), Level 2, Level 3 (*Intermediate*), Level 4 (*Intermediate*), Level 5 (*Proficient*)	(MI-ELPA) Basic, Intermediate, Proficient
MN	Beginning, Intermediate, Advanced, Transitional	(TEAE & MN-SOLOM) Level 1, Level 2, Level 3, Level 4, Level 5
MO	Beginning, Intermediate, Advanced	(MAC II) Basic Beginner, Beginner, Low Intermediate, High Intermediate, Advanced
MS	Pre-Production, Early Production, Emergent, Intermediate, High Intermediate, Transitional	(SELP) Pre-Emergent, Emergent, Basic, Intermediate, Proficient
MT	NA	(MontCAS ELP) Novice, Nearing Proficient, Proficient, Advanced
NC	Novice Low, Novice High, Intermediate Low, Intermediate High, Advanced, Superior	(New IPT) Beginner Low, Beginner High, Intermediate Low, Intermediate High, Proficient, Proficient High
ND	Preliterate, Beginning, Intermediate, Transitional, Proficient	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
NE	Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency

State	Levels in state standards	Levels in ELP assessment[a]
NJ	Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficient	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
NM	Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced	(NMELPA) Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced
NV	Level I, Level II, Level III, Level IV, Level V	(LAS Links) Beginning, Early Intermediate, Intermediate, Proficient, Above Proficient
NY	Beginning, Intermediate, Advanced, Proficient	(NYSESLAT) Beginning, Intermediate, Advanced, Proficient
OH	Pre-functional, Beginning, Intermediate, Advanced, Proficient/Trial-mainstream	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
OK	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
OR	Pre-production, Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced/Proficient	(Oregon ELPA) Pre-production, Beginning, Early Intermediate, Intermediate, Early Advanced, Advanced/Proficient
PA	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
RI	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
SC	NA	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
SD	Pre-emergent, Emergent, Basic, Intermediate, Proficient	(DELP) Pre-Emergent, Emergent, Basic, Intermediate, Proficient
TN	Beginner, High Beginner, Intermediate, High Intermediate, Advanced	(ELDA) Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
TX	NA	(TELPAS, [RPTE, TOP]) Beginning, Intermediate, Advanced, Advanced High
UT	Beginner, Pre-Emergent, Emergent, Intermediate, Advanced, Fluent	(UALPA) Beginner, Pre-Emergent, Emergent, Intermediate, Advanced, Fluent
VA	Level 1, Level 2, Level 3, Level 4	(Virginia SELP) Pre-Emergent, Emergent, Basic, Intermediate, Proficient

State	Levels in state standards	Levels in ELP assessment[a]
VT	ACCESS Levels	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
WA	Beginning, Advanced Beginning, Intermediate, Advanced, Transitional	(WLPT-II) Beginning, Advanced Beginning, Intermediate, Advanced, Transitional
WI	Preproduction, Production, Intermediate, Advanced Intermediate, Advanced, Formerly LEP	(ACCESS for ELLs®) Entering, Beginning, Developing, Expanding, Bridging
WV	Level 1, Level 2, Level 3, Level 4, Level 5	(WESTELL) Beginning, Intermediate, Advanced, Fully English Proficient
WY	Introductory, Beginning, Intermediate, Advanced	(SELP) for Grades K–2 Pre-Emergent, Emergent, Basic, Intermediate, Proficient (WELLA) for Grades 3–12 Introductory, Beginning, Intermediate, Advanced, Transitional

Note. ELL = English language learner. ELP = English language proficiency. NA = information was not available on the state's department of education Web site.

[a]This column includes assessment acronyms. See Appendix G2 for complete assessment names.

Appendix E

State ELL Redesignation Criteria

State	ELP test score[a]	Content test score[b]	District decision[c]	School personnel input[d]	Parent/ guardian input[e]	Other criteria
AK	X	—	—	—	—	—
AL	X	X	—	X	X	—
AR	X	X	—	X	X	Student grades
AZ	X	—	—	—	—	—
CA	X	X	X	X	X	Student grades
CO	X	X	X	X	X	Student portfolio
CT	X	X	—	—	—	—
DC	X	—	—	—	—	—
DE	X	—	—	—	—	—
FL	X	X	X	X	X	—
GA	X	X	—	—	—	—
HI	X	X	—	X	X	—
IA	X	X	X	X	X	—
ID	X	X	X	X	X	Student grades, observation, student portfolio
IL	X	X	X	X	X	—
IN	X	X	X	X	X	Student grades, observation, student portfolio
KS	X	—	X	—	—	—
KY	X	—	—	X	—	—
LA	X	X	—	—	—	—
MA	X	—	X	X	X	Student grades, formative assessments
MD	X	X	X	X	X	Student grades, student portfolio
ME	X	—	—	—	—	—
MI	NA	NA	NA	NA	NA	NA
MN	X	X	X	X	X	—
MO	X	X	X	—	—	Student grades, student portfolio
MS	X	X	X	X	X	Observation, student portfolio, classroom performance
MT	NA	NA	NA	NA	NA	NA
NC	X	—	—	—	—	—
ND	X	—	X	—	—	—
NE	—	—	X	—	—	—
NJ	X	X	—	X	—	Classroom performance

State	ELP test score[a]	Content test score[b]	District decision[c]	School personnel input[d]	Parent/ guardian input[e]	Other criteria
NM	X	—	—	—	—	—
NV	X	X	—	—	—	—
NY	X	—	—	—	—	—
OH	X	X	—	—	—	—
OK	X	—	—	—	—	—
OR	X	—	X	—	—	—
PA	X	X	—	—	—	Student grades
RI	X	X	—	X	—	—
SC	X	X	—	—	—	—
SD	X	—	—	—	—	—
TN	X	X	—	—	—	Observation, classroom performance
TX	X	X	—	X	—	Student grades
UT	X	—	—	—	—	—
VA	X	X	—	X	—	Classroom performance, cultural adjustment
VT	X	X	X	X	X	Student grades, student portfolios, interviews
WA	X	—	—	X	—	Student grades, school personnel input
WI	X	—	—	—	—	—
WV	X	X	—	—	—	—
WY	—	—	X	—	—	—
Total	**46**	**28**	**17**	**21**	**15**	**16**

Note. ELL = English language learner. ELP = English language proficiency. NA= information was not available on the state's department of education Web site.
— = state does not use.

[a]ELP Test Score = Student must achieve a specified score or proficiency level as indicated on an ELP assessment.

[b]Content Test Score = Student must achieve a specified score or proficiency level on all or parts of regular class content assessments (e.g., state criterion-referenced tests or norm-referenced tests, classroom/program assessments).

[c]District Decision = Local school district develops ELL exit and redesignation policies.

[d]School Personnel Input = School personnel (e.g., ESL teacher, regular classroom teacher, ELL committee, school administrators, etc.) contribute to exit and redesignation decisions.

[e]Parent/Guardian Input = Parent/guardian interviews, requests, or opinions contribute to exit and redesignation decisions.

Appendix F

Consortia Information and Consortia ELP Assessments

Mountain West Assessment Consortium (MWAC)	
Web site	www.measuredprogress.org/assessments/clients/MountainWest/MountainWest.html
Test name	—
States originally involved	AK, CO, ID, MI, MT, NV, NM, ND, OR, UT, WY
Other collaborating entities	Measured Progress
States currently using consortium test/items	ID, MI, MT, UT
Format of test	—
Grade bands	K–1, 1–2, 3–5, 6–8, 9–12
Domains	Listening, Speaking, Reading, Writing
Levels of proficiency	Pre-Emergent, Emergent, Intermediate, Fluent, Advanced
Administration procedures	—
Scoring system	—
Technical quality	—

Note. Consortium dissolved before assessment was finalized. Some states chose to develop their own assessments using some of the items and instruments developed by MWAC.
— = Does not apply to this category.

Pennsylvania Enhanced Assessment Group (PA EAG)	
Web site	www.fldoe.org/aala/cella.asp
Test Name	Comprehensive English Language Learning Assessment (CELLA)
States originally involved	FL, MD, MI, PA, TN
Other collaborating entities	Educational Testing Service, Accountability Works
States currently using consortium test/items	FL
Format of test	Multiple-choice for Listening and Reading Constructed-response and multiple-choice for Writing Constructed-response for Speaking
Grade bands	K–2, 3–5, 6–8, 9–12
Domains	Listening, Speaking, Reading, Writing
Levels of proficiency	Beginner, High Beginner, Intermediate, High Intermediate, Advanced
Administration procedures	Untimed (approx. 2.5 hrs., less for younger grades) Speaking section is individually administered Listening, Reading, and Writing sections are group administered Individual administration is recommended for Kindergarten

(table continues on next page)

Pennsylvania Enhanced Assessment Group (PA EAG)	
Scoring system (continued)	Speaking section scored by local administrator Multiple-choice scored by machine, by publisher Writing scored by rater, by publisher. States may opt to locally administer and score Writing Listening and Speaking cut scores are combined (495–835) Reading cut score (345–820) Writing cut score (515–850) Total Score cut score (1,355–2,505)
Technical quality	NA

Note. NA= Information was not available on the consortium's Web site.

State Collaborative on Assessment & Student Standards (SCASS) Consortium	
Web site	www.ccsso.org/projects/ELDA/
Test Name	English Language Development Assessment (ELDA)
States originally involved	AL, CA, HI, IN, IA, KY, LA, MI, NE, NV, NJ, NY, NC, OH, OK, OR, SC, TX, VA, WV
Other collaborating entities	University of Maryland, Council of Chief State School Officers, American Institutes for Research
States currently using consortium test/items	AR, IA, LA, NE, OH, SC, TN, WV
Format of test	Multiple-choice for Listening and Reading Constructed-response and multiple-choice for Writing Short constructed-response for Speaking K–2 inventory for all sections
Grade bands	K–2, 3–5, 6–8, 9–12
Domains	Listening, Speaking, Reading, Writing
Levels of proficiency	Beginner, Lower Intermediate, Upper Intermediate, Advanced, Full English Proficiency
Administration procedures	Speaking section is approximately 25 minutes Listening, Reading and Writing sections are each approximately 1 hour Speaking section is individually administered Listening, Reading and Writing sections are group administered
Scoring system	Speaking section responses recorded and scored off site States may opt to locally administer and score Speaking Multiple-choice scored by machine, by publisher Writing scored by rater, by publisher Cut scores vary by grade bands and levels of proficiency Listening cut score (22–53) Speaking cut score (11–29) Reading cut score (23–54) Writing cut score (9–32) Maximum points per domain (28–60)
Technical quality	See Appendix J1, J2

World-Class Instructional Design and Assessment (WIDA) Consortium	
Web site	www.wida.us/assessment/ACCESS.aspx
Test name	Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®)
States originally involved	AL, AR, DE, IL, ME, NH, RI, VT, WI
Other collaborating entities	Center for Applied Linguistics; Center for Equity and Excellence in Education; Second Language Acquisition
States currently using consortium test/items	AL, DE, DC, GA, IL, KY, ME, NJ, ND, OK, PA, RI, VT, WI
Format of test	Multiple-choice for Listening and Reading Constructed-response Writing Constructed-response for Speaking
Grade bands	Pre-K–K, 1–2, 3–5, 6–8, 9–12
Domains	Listening, Speaking, Reading, Writing
Levels of proficiency	Entering, Beginning, Developing, Expanding, Bridging
Administration procedures	Untimed (approx. 2.5 hrs.) Speaking section is individually administered Listening, Reading and Writing sections are group administered Kindergarten all sections individually administered
Scoring system	Speaking section scored by local administrator Multiple-choice scored by machine, by publisher Writing scored by rater, by publisher Kindergarten scored by local administrator Scale scores 100–600 for each section
Technical quality	See Appendices J1, J2

Appendix G1

State ELP Assessment Histories

State	ELP assessment[a]	Years used[b]	Starting school year	Previous ELP assessments
AK	New IPT	2	2005–2006	IPT
AL	ACCESS for ELLs®	3	2004–2005	LAS, IPT[c]
AR	ELDA	1	2006–2007	LAS, IPT
AZ	AZELLA (augmented SELP)	1	2006–2007	SELP
CA	CELDT	7	2001–2002	LAS
CO	CELA	2	2005–2006	NA
CT	LAS Links	2	2005–2006	LAS
DC	ACCESS for ELLs®	3	2004–2005	LAS, Pre-LAS
DE	ACCESS for ELLs®	2	2005–2006	LAS
FL	CELLA	1	2006–2007	Cadre of 7–12 state-approved tests
GA	ACCESS for ELLs®	2	2005–2006	Language Assessment Battery
HI	LAS Links	2	2005–2006	LAS, Pre-LAS, Basic Inventory of Natural Language
IA	ELDA	2	2005–2006	LAS, IPT
ID	IELA (Mountain West items)	2	2005–2006	WMLS, LAS, IPT, MAC II
IL	ACCESS for ELLs®	2	2005–2006	LAS, IPT, Language Proficiency Test Series, MAC II
IN	LAS Links	2	2005–2006	WMLS, IPT
KS	KELPA	2	2005–2006	LAS, IPT, Language Proficiency Test Series
KY	ACCESS for ELLs®	1	2006–2007	LAS, IPT
LA	ELDA	3	2004–2005	NA
MA	MEPA (MEPA-R/W, MELA-O)[d]	3	2004–2005	MELA-O
MD	LAS Links	2	2005–2006	IPT
ME	ACCESS for ELLs®	2	2005–2006	LAS, IPT
MI	MI-ELPA (Mountain West items)	2	2005–2006	LAS, Pre-LAS, IPT, Bilingual Verbal Ability Test, SELP, MAC II, Woodcock-Muñoz Language Survey
MN	TEAE MN-SOLOM[d]	7 7	2001–2002 2001–2002	NA NA
MO	MAC II	7	2001–2002	NA
MS	SELP	NA	NA	NA
MT	MontCAS ELP (Mountain West items)	2	2005–2006	NA
NC	New IPT	2	2005–2006	IPT
ND	ACCESS for ELLs®	1	2006–2007	LAS, IPT, Woodcock-Muñoz Language Survey

State	ELP assessment[a]	Years used[b]	Starting school year	Previous ELP assessments
NE	ELDA	3	2004–2005	LAS, IPT, Bilingual Verbal Ability Test, Language Proficiency Test Series, Woodcock-Muñoz Language Survey
NJ	ACCESS for ELLs®	1	2006–2007	LAS, IPT, MAC II
NM	NMELPA (augmented SELP plus Mountain West items)	1	2006–2007	LAS, IPT, Woodcock-Muñoz Language Survey
NV	LAS Links	2	2005–2006	Pre-LAS, LAS
NY	NYSESLAT	3	2004–2005	NA
OH	OTELA (augmented ELDA)	2	2005–2006	ELDA
OK	ACCESS for ELLs®	1	2006–2007	LAS, IPT
OR	Oregon ELPA	2	2005–2006	LAS, IPT, SELP, Woodcock-Muñoz Language Survey
PA	ACCESS for ELLs®	1	2006–2007	NA
RI	ACCESS for ELLs®	2	2005–2006	MAC II
SC	ELDA	3	2004–2005	LAS, IPT, Woodcock-Muñoz Language Survey
SD	DELP (augmented SELP)	2	2005–2006	SELP
TN	ELDA	1	2006–2007	CELLA
TX	TOP RPTE [d]	3 7	2004–2005 2000–2001	NA NA
UT	UALPA (Mountain West items)	1	2006–2007	NA
VA	Virginia SELP (augmented SELP)	3	2004–2005	SELP
VT	ACCESS for ELLs®	2	2005–2006	NA
WA	WLPT-II (augmented SELP)	NA	NA	NA
WI	ACCESS for ELLs®	2	2005–2006	LAS, IPT, MAC II, Woodcock-Muñoz Language Survey
WV	WESTELL (augmented ELDA)	3	2004–2005	NA
WY	SELP[e] WELLA (augmented SELP)[d]	NA 2	NA 2005–2006	NA SELP

Note. ELP = English language proficiency. ELL = English language learner. NA = information was not available on the state's department of education Web site.

[a]This column indicates assessment acronyms. See Appendix G2 for complete assessment names.

[b]Years used includes current (2006–2007) school year.

[c]IPT refers to versions prior to 2005 unless otherwise stated.

[d]Minnesota uses two assessments for ELP assessment: one for reading and writing, the other for listening and speaking.

[e]Wyoming uses SELP for Grades K–2, and WELLA for Grades 3–12.

Appendix G2

Current ELP Assessments in Use and Their Developers

ELP assessment	Developer and Web site
Arizona English Language Learner Assessment (AZELLA)	Arizona Department of Education, Harcourt Assessment http://www.ade.state.az.us/asd/lep/ Publisher: Harcourt Assessment
Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®)	Center for Applied Linguistics in collaboration with the WIDA Consortium http://www.wida.us Publisher: Metritech
California English Language Development Test (CELDT)	California Department of Education, CTB/McGraw Hill http://www.cde.ca.gov/ta/tg/el/ Publisher: CTB/McGraw Hill
Colorado English Language Assessment (CELA)	Colorado, with CTB-McGraw Hill as contractor http://www.ctb.com/netcaster/extranet/program_index.jsp?PROG_ID=1244 Publisher: CTB/McGraw Hill
Comprehensive English Language Learning Assessment (CELLA)	ETS, PA EAG Consortium, and Accountability Works (Washington, DC) http://www.fldoe.org/aala/cella.asp Publisher: ETS, Accountability Works
Dakota English Language Proficiency Test (DELP)	South Dakota Department of Education, Measurement Inc., AIR, CCSSO, LEP-SCASS (Based on the test item banks and scales from the ELDA) http://doe.sd.gov/octa/assessment/delp/index.asp
English Language Development Assessment (ELDA)	Measurement Inc., AIR, CCSSO, LEP-SCASS http://www.ccsso.org/projects/ELDA/
Idaho English Language Assessment (IELA)	(A modified version of an assessment developed for the Mountain West Consortium) Currently produced by BETA/TASA, Inc., with ongoing modification of items. http://www.boardofed.idaho.gov/lep/LEPAssessment.asp
Kansas English Language Proficiency Assessment (KELPA)	Center for Educational Testing and Evaluation (CETE) (Part of Kansas University) and Kansas State Department of Education http://www.ksde.org/Default.aspx?tabid=1636
Language Assessment Scales Links K–12 Assessment (LAS Links)	CBT/McGraw-Hill http://www.ctb.com/products/product_summary.jsp?FOLDER%3C%3Efolder_id=1408474395292398&bmUID=1170710441740

ELP assessment	Developer and Web site
Maculaitis Assessment of Competencies Test of English Language Proficiency (MAC II)	Touchstone Applied Science Associates, Inc http://store.cambiumlearning.com/ProgramPage.aspx?parentId=019005529&functionID=009000008&pID=&site=sw
Massachusetts English Proficiency Assessment – Reading, Writing, Oral (MEPA-R/W, MELA-O)	Massachusetts Department of Education in collaboration with the Evaluation Assistance Center (East) at The George Washington University http://www.doe.mass.edu/mcas/mepa/
Michigan English Language Proficiency Assessment (MI-ELPA)	Harcourt Assessment, Mountain West http://www.measuredprogress.org/assessments/clients/MountainWest/MountainWest.html Michigan Department of Education (MI) http://www.michigan.gov/mde/0,1607,7-140-22709_40192---,00.html
Minnesota Student Oral Language Observation Matrix (MN-SOLOM)	Minnesota Department of Education, adapted from San Jose, CA Unified School District's SOLOM http://education.state.mn.us/MDE/Accountability_Programs/Assessment_and_Testing/Assessments/ELL_Tests/index.html
Montana Comprehensive Assessment System English Language Proficiency (MontCAS ELP)	Measured Progress (with Mountain West items) http://www.opi.state.mt.us/Assessment/ELP.html
New IDEA Proficiency Test (New IPT) 2005	Ballard & Tighe http://www.ballard-tighe.com/eld/carousel.html
New Mexico English Language Proficiency Assessment (NMELPA)	Harcourt Assessment, Inc. (with Mountain West items) http://www.ped.state.nm.us/div/acc.assess/assess/index.html
New York State English as a Second Language Achievement Test (NYSESLAT)	New York State Education Department and Harcourt Assessment, Inc. http://www.emsc.nysed.gov/osa/nyseslat/ Publisher: Harcourt Assessment
Ohio Test of English Language Acquisition (OTELA)	Ohio Department of Education, Measurement Inc., AIR, CCSSO, LEP-SCASS (Based on the test item banks and scales from the ELDA) http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=1086&ContentID=8402&Content=22547
Oregon English Language Proficiency Assessment (Oregon ELPA)	Oregon Department of Education http://www.ode.state.or.us/search/page/?id=1224
Stanford English Language Proficiency (SELP)	Harcourt Assessment http://harcourtassessment.com/haiweb/cultures/en-us/productdetail.htm?pid=015-8429-206

ELP assessment	Developer and Web site
Test of Emerging Academic English (TEAE)	Minnesota Department of Education Language Learning Solutions, Ltd. Lidget Green, Inc., http://education.state.mn.us/MDE/Accountability_Programs/Assessment_and_Testing/Assessments/ELL_Tests/index.html Publisher: Metritech
Texas English Language Proficiency Assessment System (TELPAS) / Reading Proficiency Tests in English (RPTE)	Pearson Educational Measurement, Texas Education Agency http://www.tea.state.tx.us/student.assessment/admin/rpte/index.html
Texas English Language Proficiency Assessment System (TELPAS) / Texas Observational Protocols (TOP)	Pearson Educational Measurement, Texas Education Agency http://www.tea.state.tx.us/student.assessment/admin/rpte/index.html
Utah Academic Language Proficiency Assessment (UALPA)	Utah State Office of Education, Mountain West http://www.schools.utah.gov/eval/DOCUMENTS/UALPA_Coordinators_Manual.pdf Publisher: Measured Progress
Washington Language Proficiency Test (WLPT-II)	Washington Office of Superintendent of Public Instruction http://www.k12.wa.us/assessment/WLPTII/default.aspx Publisher: Harcourt Assessment
West Virginia Test of English Language Learning (WESTELL)	West Virginia Department of Education, Measurement Inc., AIR, CCSSO, LEP-SCASS (Based on the test item banks and scales from the ELDA) http://wvconnections.k12.wv.us/assessment.html
Woodcock-Muñoz Language Survey	Riverside Publishing http://www.riverpub.com/products/wmls/index.html
World-Class Instructional Design and Assessment - Assessing Comprehension and Communication in English State-to-State Placement Test (W-APT)	Center for Applied Linguistics in collaboration with the WIDA Consortium http://www.wida.us/assessment/W-APT.aspx Publisher: Metritech
Wyoming English Language Learners Assessment (WELLA)	Wyoming Department of Education http://www.k12.wy.us/FP/title3.asp Publisher: Harcourt Assessment

Note. ELP = English language proficiency. ELL = English language learner.

Appendix H

Purposes of ELP Assessments

ELP assessment	Test purposes[a]
Arizona English Language Learner Assessment (AZELLA)	▶ To assess students' general acquisition of English. ▶ To determine students' readiness for redesignation.
Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs®)	▶ To help determine the language proficiency levels of the ELL. ▶ For schools to determine ELL progress and reclassification.
California English Language Development Test (CELDT)	▶ To identify new students who are English learners in K–12. ▶ To determine their level of English language proficiency. ▶ To annually assess their progress in acquiring listening, speaking, reading, and writing skills in English.
Colorado English Language Assessment (CELA)	▶ To place students into instructional programs. ▶ To explore students' English language abilities through an analysis of linguistic characteristics (i.e., vocabulary, grammar, and syntactic structures). ▶ To obtain additional diagnostic information on learners.
Comprehensive English Language Learning Assessment (CELLA)	▶ To provide evidence of program accountability in accordance with NCLB. ▶ To chart ELL student progress. ▶ To provide information to make decisions about ELL student proficiency levels and redesignation. ▶ To provide diagnostic information about ELL students' individual strengths and weaknesses in English.
English Language Development Assessment (ELDA)	▶ To measure annual progress in the acquisition of ELP skills among non-native English speaking student in Grades K–12.
Idaho English Language Assessment (IELA)	▶ To fulfill the requirements of NCLB. ▶ To assess English proficiency in listening, speaking, reading, and writing. ▶ To report scores in each language domain and a total score.
Kansas English Language Proficiency Assessment (KELPA)	▶ To identify and measure students' English language proficiency.
Language Assessment Scales Links K–12 Assessment (LAS Links)	▶ To measure "English language skills in reading and writing necessary for functioning in a mainstream academic environment." ▶ To provide information for classifying ELL students and monitoring their progress in acquiring English.
Maculaitis Assessment of Competencies Test of English Language Proficiency (MAC II)	▶ To inform decisions about individual ELL students, such as identification, program placement, placement review, and program exit. ▶ To inform classroom instruction. ▶ To provide information for program evaluation and school accountability.

ELP assessment	Test purposes[a]
Massachusetts English Proficiency Assessment (MEPA)	▸ To provide data and resources to strengthen curriculum, instruction, and classroom assessment. ▸ To measure the progress made by LEP students toward English proficiency over time. ▸ To identify LEP students who have achieved proficiency in English.
Michigan English Language Proficiency Assessment (MI-ELPA)	▸ For official reporting of Title III annual performance of ELL student progress in acquiring English language skills. ▸ To measure the English language proficiency levels of students who are learning English as a second language.
Minnesota Student Oral Language Observation Matrix (MN-SOLOM)	▸ To rate, through teacher observation, the progress of ELL students in listening and speaking English proficiency.
Montana Comprehensive Assessment System English Language Proficiency (MontCAS ELP)	▸ To assess the proficiency level of LEP students in the five areas of listening, speaking, reading, writing, and comprehension. ▸ To provide Montana educators with proficiency scores for use in their schools, systems, and state, as mandated by NCLB.
New IDEA Proficiency Test (New IPT) 2005	▸ To provide "comprehensive assessment for the initial identification and redesignation of [LEP]." ▸ To evaluate students' oral language, reading, and writing in English. ▸ For identification, placement, and redesignation of ELL students. ▸ To measure the proficiency and progress of ELL students. ▸ To provide language proficiency testing data and important diagnostic information about the student.
New Mexico English Language Proficiency Assessment (NMELPA)	▸ To document ELL/LEP students' annual progress and attainment of English language proficiency in the domains of listening, speaking, reading, writing, and comprehension.
New York State English as a Second Language Achievement Test (NYSESLAT)	▸ To measure the English language arts proficiency of ELL students, across the state, Grades K–12, from year to year. ▸ To identify ELL students' English language proficiency level. ▸ To identify those ELL students who have achieved a Proficient level and no longer require ESL and/or bilingual services.
Ohio Test of English Language Acquisition (OTELA)	▸ To determine the ELP for Ohio LEP students in Grades K–12.
Oregon English Language Proficiency Assessment (Oregon ELPA)	▸ To measure the development of students in English language proficiency and their progress toward proficiency, including academic language skills. ▸ To provide school staff with data that can contribute to evaluating student progress individually.
Stanford English Language Proficiency (SELP) Test	▸ To evaluate K–12 ELL listening, reading, comprehension, writing, and speaking skills. ▸ To identify and place ELL students. ▸ To measure outcomes and progress. ▸ To evaluate program effectiveness. ▸ To assist in instructional planning.

ELP assessment	Test purposes[a]
South Dakota English Language Proficiency Test (DELP)	▸ To measure annual progress in the acquisition of ELP skills among non-native English speaking student in Grades K–12. ▸ To annually measure the ELP of all identified LEP students, documenting their progress in acquiring ELP.
Test of Emerging Academic English (TEAE)	▸ To assess yearly progress of ELL students in reading and writing English proficiency. ▸ To generate information on program accountability.
Texas English Language Proficiency Assessment System (TELPAS) / Reading Proficiency Tests in English (RPTE)	▸ To measure English reading ability according to a language proficiency continuum, based on the stages of language development of second language learners. ▸ To provide a measure of the ELL student's current reading proficiency levels and annual progress in learning to read in English.
Texas English Language Proficiency Assessment System (TELPAS) / Texas Observational Protocols (TOP)	▸ To meet NCLB requirements for assessing the English language proficiency of Grades K–12 ELL students in listening, speaking, reading, and writing. ▸ To meet NCLB accountability requirements for meeting the educational needs of ELL students.
Utah Academic Language Proficiency Assessment (UALPA)	▸ To assess the proficiency level of English language learners. ▸ To provide educators with a total proficiency score for use in their schools, districts, and state, as mandated by NCLB. ▸ To address the goals of NCLB. ▸ To assess academic English acquisition.
Washington Language Proficiency Test (WLPT-II)	▸ Annual assessment to measure growth of ELL students.
West Virginia Test of English Language Learning (WESTELL)	▸ To measure annual progress in the acquisition of ELP skills among non-native English speaking student in Grades K–12.
Woodcock-Muñoz Language Survey	▸ To establish language proficiency levels in English.
World-Class Instructional Design and Assessment - Assessing Comprehension and Communication in English State-to-State Placement Test (W-APT)	▸ To identify ELL students for program placement. ▸ To determine academic English proficiency level of new ELL students. ▸ To assign ELL students to correct Tier grouping for ACCESS scale.
Wyoming English Language Learners Assessment (WELLA)	▸ Annual assessment to measure growth of ELL students. ▸ Fall identification and placement.

Note. LEP = Limited English proficiency. ELL = English language learner. ELP = English language proficiency.

[a]Information was compiled from available test developer or related state documents.

Appendix I

The Constructs of Each ELP Assessment Used by States

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: process, understand, interpret, and evaluate spoken language in various situations for 5 content areas: social and instructional, language arts, math, science, and social studies *Speaking*: engage in oral communication in a variety of situations for an array of purposes and audiences for 5 content areas: social and instructional, language arts, math, science, and social studies *Reading*: process, interpret, and evaluate written language, symbols and text with understanding and fluency for 5 content areas: social and instructional, language arts, math, science, and social studies *Writing*: engage in written communication in a variety of forms for an array of purposes and audiences for 5 content areas: social and instructional, language arts, math, science, and social studies
Arizona English Language Learner Assessment (AZELLA)	Pre-K–K 1–2 3–5 6–8 9–12	Listening Speaking Reading Writing Writing Conventions	(Descriptions could not be found.) The final consideration in developing AZELLA was the specific language to be tested. Language, especially spoken language, is fluid and ever changing. It is essential that English Language Proficiency tests focus on *fresh, vibrant language*, the language that is actually used in classrooms and the community.
California English Language Development Test (CELDT)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: respond to social and classroom commands, understand spoken information in academic settings, answer questions about a short story *Speaking*: know how to use nouns, action words, adjectives, and adverbs, using language for specific tasks, state a preference and give two reasons, tell a story based on a series of pictures *Reading*: analyze patterns and structure of word, know a range of word definitions, and understand facts, inferences, and critical analysis of written stories *Writing*: use Standard English grammatical structure and writing conventions, construct sentences on specific topics, compose short compositions on specific topics
Colorado English Language Assessment (CELA)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: assess general comprehension and inferential and critical thinking skills at a discourse level *Speaking*: assess vocabulary, social and academic language and grammar *Reading*: evaluate phonemic awareness for recognizing words and developing vocabulary and demonstrate sentence level reading ability *Writing*: evaluate language usage and fluency

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
Comprehensive English Language Learning Assessment (CELLA)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: listen and match, picture description, short talks, extended listening *Speaking*: oral vocabulary, speech functions, personal opinion, story retelling, graph interpretation *Reading*: synonym, antonym, idiom, root and affix, main idea, detail, inference/prediction, reference, rhetorical elements, vocabulary in context *Writing*: grammar, structure and written expression, paragraph choices, recognizing errors, writing sentences, writing paragraphs
English Language Development Assessment (ELDA)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: comprehend spoken instructions; determine main idea/purpose; identify important supporting ideas; determine speaker's attitude/perspective; comprehend key vocabulary/phrases; draw inferences, predictions, conclusions *Speaking*: connect, tell, explain, reason *Reading*: demonstrate pre-/early reading skills; comprehend key vocabulary/phrases; comprehend written instructions; determine main idea/purpose; identify important supporting ideas; draw inferences, predictions, conclusions; determine writer's attitude/perspective; analyze style/form *Writing*: planning and organizing; writing a draft text: narrative, descriptive, expository, persuasive; revising; editing; writing conventions
Idaho English Language Assessment (IELA)	K 1–2 3–5 6–8 9–12	Listening Speaking Reading Writing	(Descriptions could not be found.)
Kansas English Language Proficiency Assessment (KELPA)	K–1 2–3 4–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: follow directions; identify beginning, middle, ending sounds; discern between correct and incorrect sentence; evaluate listening comprehension (based on story or a listening passage) *Speaking*: answer short questions, answer more detailed questions, describe what's happening in a single picture, describe what's happening in a picture sequence *Reading*: identify rhyming words, identify initial/ending sounds, assess short story reading comprehension, complete cloze sentences, identify compound words, synonyms/antonyms definitions; fact/opinion, and analogies; assess reading comprehension of passages *Writing*: write letters/numbers based on oral prompt, complete the cloze sentence, compose sentence rewriting (correct syntax of incorrect sentence), circle correctly spelled word, identify vocabulary (write the word to label a picture), assess grammar/vocabulary usage (adjectives, prepositions,　　*(continues on next page)*

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
(continued) (KELPA)	K–1 2–3 4–5 6–8 9–12	Listening Speaking Reading Writing	verbs + verb tenses, comparative/superlative, adverbs, pronouns, plurals), identify synonyms/antonyms, punctuation, and syntax, write essay based on either picture or written prompt
Language Assessment Scales (LAS Links) Links K–12 Assessment	K–1 2–3 4–5 6–8 9–12	Oral Language (includes Speaking & Listening) Reading Writing	*Listening*: follow common, explicit oral directions to participate in diverse academic or social tasks, respond to idiomatic expressions to participate in diverse academic or social tasks including phrasal verbs with idiomatic meaning, identify main ideas, identify supporting ideas, predict and inference based on known information

Speaking: provide information, express opinions and preferences, make requests, ask questions, request clarification & negotiate for understanding, identify an object & describe its purpose or use, use words or phrases, identify an academic or social situation & describe it using sentences, describe processes, compare & explain preferences, interpret, narrate & paraphrase events using visual information

Reading: identify rhyming words, apply letter-sound relationships to read English words and phonemes, apply knowledge of morphemes and syntax to word meaning, classify words, demonstrate vocabulary & reading comprehension, identify important literary features of text, read critically & apply learning strategies to interpretation

Writing: use singular & plural, subject/ verb agreement, tense agreement, conjunctions, pronouns, prepositional phrases, and auxiliary verbs, capitalize beginning of sentences & proper names, use sentence-ending marks, commas in series and dates, apostrophes in contractions & possessives, differentiate complete sentences from fragments, use articles, form statements and questions, differentiate complete sentences from run-ons, use adjectives and adverbs, write simple sentences to describe and explain, write to describe, explain, report, compare, narrate, persuade, or express |
| Maculaitis Assessment of Competencies Test of English Language Proficiency (MAC II) | K–1 2–3 4–5 6–8 9–12 | Listening Speaking Reading Writing | Test content depends on grade cluster.

Listening: understanding words (nouns, verbs), classroom directions, stories, questions and answers, oral instructions, and conversations

Speaking: naming things, answering questions, asking social questions, telling a story

Reading: (optional for K) recognizing letters, reading words, reading and understanding words, vocabulary, reading sentences and passages (DRP)

Writing: (optional for K) writing words, sentences, a story, an essay, completing a form and grammar |

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
Massachusetts English Proficiency Assessment (MEPA) Comprised of Massachusetts English Proficiency Assessment-Reading and Writing (MEPA-R/W) and Massachusetts English Language Assessment-Oral (MELA-O)	MELA-O K–12 MEPA-R/W 3–4 5–6 7–8 9–12	Listening Speaking Reading Writing	**MEPA-R/W** *Reading*: vocabulary, beginning to read, comprehension, literary elements/expository text *Writing*: writing, editing **MELA-O** *Listening*: to assess, through multiple observations in academic and social activities, each student's listening comprehension *Speaking*: to assess, through multiple observations in academic and social activities, each student's speech production, each student's fluency, vocabulary, pronunciation, grammar
Michigan English Language Proficiency Assessment (MI-ELPA)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing Comprehension (reading & listening)	(Descriptions could not be found.) ELPA assesses the forms of the language (i.e., knowledge of syntax, vocabulary and morphology) and the functions of the language (i.e., ability to use English to express ideas and get things done; ideational and manipulative).
Minnesota Student Oral Language Observation Matrix (MN-SOLOM)	K–2 3–12	Listening Speaking (Reading) (Writing)	Teacher observations of students in social (informal) and academic (formal) settings using checklist *Listening*: academic comprehension, social comprehension *Speaking*: fluency, vocabulary, pronunciation, grammar *Reading*: (K–2 only) (Descriptions could not be found.) *Writing*: (K–2 only) (Descriptions could not be found.)
Montana Comprehensive Assessment System English Language Proficiency (MontCAS ELP)	K–12	Listening Speaking Reading Writing	(Descriptions could not be found.) Based on ELPA test, developed in part with Mountain West Consortia.
New IDEA Proficiency Test (New IPT) 2005	Pre-K–K 1–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: give a physical response to a rhyming pattern, repeat a rhyming pattern; choose a picture that corresponds to a high-frequency/sight word; follow directions; listen to a story and answer questions; identify a picture that corresponds to a word or sentence; identify letters that correspond to the beginning sounds of words; identify letters that correspond to the ending sounds of words, identify *(continues on next page)*

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
(continued) (New IPT) 2005	Pre-K–K 1–2 3–5 6–8 9–12	Listening Speaking Reading Writing	picture of students who have followed the directions; answer picture-based comprehension questions about the extended texts; identify picture of students who have followed the directions; choose a math expression that corresponds to a math word problem; answer comprehension questions about classroom dialogue; answer comprehension questions about the academic texts *Speaking*: give directions, retell a story; identify common objects and actions using high-frequency words; describe a classroom activity presented in a series of pictures; explain simple math operations; describe a common school activity presented in a series of pictures; describe and compare information in a graph, express an opinion on one of the three topics *Reading*: identify the beginning/ending sound of a word, distinguish different phonemes, choose the letter or combination that corresponds to a beginning/ending sound, sound out unfamiliar words; recognize parts of a book, name and order alphabet letters and group words by letters, read high frequency words and simple sentences for understanding, read aloud a story and answer comprehension questions; identify a word that describes a picture, identify a sentence that describes a picture, answer comprehension questions about a graph, answer comprehension questions about extended texts; identify a word that describes a picture, answer comprehension questions about a graph, answer comprehension questions about extended texts *Writing*: copy letters from a story; retell a story; choose correct spelling or punctuation that completes a sentence, choose an answer that completes a sentence grammatically, write a word in the blank in a sentence that describes a picture, write a narrative story about a common school activity/event presented in a series of pictures; write two sentences to describe a picture, write a narrative essay in response to a topic; write a paragraph to describe a picture or graph related to content areas, write a paragraph to explain a process shown in a series of pictures, write a persuasive essay in response to a written prompt
New Mexico English Language Proficiency Assessment (NMELPA)	Pre-K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: assess word/phrase/sentence comprehension, respond to authentic task-based items, comprehend conversational language *Speaking*: repeat or read aloud to assignments, sentence completion, discuss storytelling and social interaction *Reading*: assess word reading, sentence reading, passage comprehension (fiction, informational, and functional passages); evaluate initial understanding, interpretation, vocabulary and idioms in context, identify fact and opinion, evaluate literary analysis *(continues on next page)*

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
(continued) (NMELPA)	PreK–2 3–5 6–8 9–12	Listening Speaking Reading Writing	***Writing***: assess phonemic understanding, sentence structure, mechanics (capitalization, punctuation); evaluate early writing skills, copying, and dictation; generate ideas in the writing process ***Extended Response Writing***: evaluate accuracy and variety of sentence structure, identify appropriate and precise word choice; assess degree of fluency
New York State English as a Second Language Achievement Test (NYSESLAT)	K–1 2–4 5–6 7–8 9–12	Listening Speaking Reading Writing	***Listening***: comprehend authentic conversational English; comprehend and synthesize information students hear through task-based questions ***Speaking***: differentiate pronunciation, rate of speech, intonation, and general intelligibility; complete a sentence for an appropriate response, provide more elaborate descriptions and show relationships between ideas; accuracy of language, lexical appropriateness, and structure; exhibit sociolinguistic competence ***Reading***: understand basic reading skills, simple word and sentence recognition; understand and interpret vocabulary/idioms in context; understand directly stated details or relationships; extend meaning and infer relationships among the ideas suggested by the text; predict meanings of unknown words or idioms from clues in surrounding text ***Writing***: understand phonemes and associate letters of the alphabet; apply principles that form effective writing; understand English language structure, spelling, capitalization, and punctuation; generate ideas; use appropriate and precise vocabulary, construct sentences that are varied in structure and length, and organize sentences into paragraphs
Ohio Test of English Language Acquisition (OTELA)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	(Descriptions could not be found.) Based on ELDA.
Oregon English Language Proficiency Assessment (OR ELPA)	K–1 2–3 4–5 6–8 9–12	Listening Speaking Reading Writing	(Descriptions could not be found.)
Stanford English Language Proficiency Test (SELP)	Pre-K K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	***Listening***: evaluate student comprehension of conversational English or Spanish, analyze sentence level and longer discourse ***Speaking***: evaluate phonemic and linguistic accuracy, informational appropriateness, and overall intelligibility in the language ***Reading***: evaluate word understanding and different varieties of text and comprehension; understand directly stated details or relationships, *(continues on next page)*

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
(continued) (SELP)	Pre-K K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	extend meaning and infer relationships among ideas, predict meanings of unknown words or idioms ***Writing***: recognize English (or Spanish) language grammatical structure, recognize correctly spelled words, recognize conventional punctuation and capitalization
Test of Emerging Academic English (TEAE)	3–4 5–6 7–8 9–12	Reading Writing	***Reading***: activate prior knowledge on the reading topic, reflect typical classroom reading activities, understand literal or inferential questions, understand narrative and expository texts ***Writing***: use any writing strategy or style, express ideas in writing, the ability to stay on topic when writing, use details and precise language when writing, order and make connections between ideas, edit writing for spelling and punctuation
Test of Emerging Academic English: Listening and Speaking (TEAELS)	K–2 3–5 6–8 9–12	Listening Speaking	***Listening***: understand instructions; understand explanations, descriptions, sequences, compare/contrast and hypothesize; identify causal & temporal connections ***Speaking***: speak to paraphrase, define, enumerate, exemplify, sequence & hypothesize, inquire, suggest, inform, request; speak to describe, explain, describe cause & effect, give opinions, compare & contrast; speak to narrate, predict
Texas English Language Proficiency Assessment System (TELPAS) Comprised of Reading Proficiency Tests in English (RPTE) and Texas Observational Protocols (TOP)	TOP K–12 RPTE 3 4–5 6–8 9–12	TOP: Listening Speaking Reading Writing RPTE: Reading	**TOP**: (Teacher observations and ratings of students in formal and informal academic tasks and interactions over time. The writing ratings are based on student writing collections.) ***Listening***: to gather information about a student's English language proficiency level; to reflect on how well the student understands the English he or she hears during activities such as reacting to oral presentations, responding to text read aloud, following directions, cooperative group work, informal, social discourse with peers, large-group and small-group interactions in academic setting, one-on-one interviews ***Speaking***: to gather information about a student's English language proficiency level; to reflect on how well the student speaks English during activities such as cooperative group work, oral presentations, informal and social discourse with peers, large-group and small-group interactions in academic settings, one-on-one interviews, classroom discussions, articulation of problem-solving strategies, individual student conferences ***Reading***: (K–2 Only) to observe student's reading during academic settings and informal and spontaneous reading that occurs naturally, to reflect on how well each student understands the English used during activities such as paired reading, sing-alongs and read-alongs, shared reading, guided reading with leveled readers/text, reading subject-area texts, *(continues on next page)*

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
(continued) (TELPAS) (RPTE) (TOP)	TOP K–12 RPTE 3 4–5 6–8 9–12	TOP: Listening Speaking Reading Writing RPTE: Reading	independent reading, literature circles, cooperative group work, reading response journals, sustained silent reading *Writing*: (K–1 Only) to observe student's writing during academic settings, to reflect on how well each student writes in English during activities such as journal writing for personal reflections, shared writing for literacy and content-area development , language experience dictation, organization of thoughts and ideas through prewriting strategies, publishing and presenting, making lists for specific purposes, labeling pictures, objects, and items from projects, cooperative group work, first drafts, revising and editing skill application *Writing*: (Grades 2–12) to examine a student's writing collection, which is representative of the writing the student does during language and literacy instruction and in a variety of academic content areas RPTE (Descriptions could not be found.)
Utah Academic Language Proficiency Assessment (UALPA)	K 1–2 3–6 7–8 9–12	Listening Speaking Reading Writing Comprehen- sion (reading & listening)	(Descriptions could not be found.)
Washington Language Proficiency Test (WLPT-II)	K–2 3–5 6–8 9–12	Listening, Writing Conven- tions, Reading, Direct Writing, Speaking	*Listening*: Word/Phrase/Sentence Comprehension, Responding to authentic task-based items, Comprehension of Conversational Language *Writing*: Sentence structure, Mechanics, Capitalization, Punctuation, Usage, Spelling *Reading*: Word Reading, Sentence Reading, Comprehension of printed discourse (Narrative, informative, and functional passages) *Speaking*: Repeat/Read Aloud, Sentence Completion, Storytelling, Academic Narrative, Social Interaction
WIDA ACCESS Placement Test (W-APT)	K–2 3–5 6–8 9–12	Listening Speaking Reading Writing	*Listening*: process, understand, interpret, and evaluate spoken language in various situations *Speaking*: engage in oral communication in a variety of situations for an array of purposes and audiences *Reading*: process, interpret and evaluate written language, analyze symbols and text with understanding and fluency *Writing*: engage in written communication in a variety of forms for an array of purposes and audiences

ELP assessment 2006–2007	Grade band	Domains	Test constructs[a]
Woodcock-Muñoz Language Survey	No grade bands, may be used for test takers age 2 and up	Listening Speaking Reading Writing	***Listening***: recall increasingly-complex stories that are presented using an audio recording, listen to audio-recorded instructions and follow directions by pointing at objects; recall increasingly-complex stories that are presented using an audio recording, assess listening, lexical knowledge and working memory ***Speaking***: name the familiar and unfamiliar pictured objects that involve breadth and depth of school-related knowledge and experience; assess oral language, including language development and lexical knowledge, complete oral analogies requiring verbal comprehension and reasoning, assess reasoning using lexical knowledge ***Reading***: read familiar and unfamiliar letters and words; measure letter-word identification skills; comprehending passages ***Writing***: respond in writing to questions which require verbal comprehension, knowledge of letter forms, spelling, punctuation, capitalization, and word usage; measure prewriting skills (for early items) and ability to respond in writing to a variety of questions
Wyoming English Language Learner Assessment (WELLA)	K–2 3–12	Listening Reading Writing	Augmented SELP SELP: ***Listening***: understand conversational discourse, discern academic and social contexts, respond to authentic task-based questions ***Reading***: understand literary, informational, and functional passages, assess initial understanding, interpretation, vocabulary and idioms in context ***Writing***: understand descriptive and narrative prompts, identify opinion and persuasive prompts

Note. ELP = English language proficiency. ELL = English language proficiency.

[a]Information was compiled from available test developer or related State documents.

Appendix J1

Summary of Validity Evidence of 13 ELP Assessments

Common validation activities based on the reviewed technical manuals for 13 English language proficiency (ELP) assessments are summarized in the table below using the following abbreviations:

Reliability:
Corr (Correlation) = studies based on some type of correlational statistics such as Cronbach's alpha internal consistency estimate, or based on estimates derived from item response theory (IRT) analyses. These estimates examine the relationships between scores on particular test items and performance on the overall test.
RR (rater reliability) = studies of rater agreement, employing designs such as generalizability theory or simple agreement percentage. These studies are usually employed for performance assessments such as assessments of writing or speaking.

Content / Construct Validity:
Content = Content validity is typically established through reviews by a panel of experts such as senior teachers and curriculum experts. The panel reviews items and compares the content to curriculum and State standards. This type of study is typically qualitative.
Construct = Construct validity is most typically established using factor analytic techniques if it is investigated separately from criterion or predictive validity.

Criterion Validity:
C (Correlation) = studies based on correlational statistics such as the Pearson correlation coefficient, usually correlating test scores with an external measure of achievement. They may be of predictive or concurrent type, or both.
P (Predictive) = Predictive studies designed to see how well the test predicts relevant educational outcomes. For example, a study may examine how well a new test predicts students' end-of-year course grades, or students' status in the following years.
Con (concurrent) = Concurrent studies designed to compare simultaneous measures of the same or very closely related constructs. For example, many studies involve students' taking a new assessment and an established assessment concurrently and examining the relationship between the two tests.

Bias / Fairness:
Rev (Expert Review): The most common type of item bias or fairness study involves bias experts reviewing items for various types of cultural, linguistic, gender, etc. construct-irrelevant content. Items that are flagged by these experts are typically removed or rewritten.
DIF (Differential Item Functioning): Various statistical techniques, such as logistic regression and IRT are used to identify items that perform differentially for different groups, after taking test takers' overall ability into account. Items identified by this technique are then either removed without further review or subjected to expert scrutiny to determine the cause of this bias.

Test name	State	Reliability	Content / construct validity	Criterion validity[a]	Bias / fairness study
Consortium tests					
ACCESS	WIDA Consortium States[b]	Corr, RR: High reliability	Content: Expert panel review	C, Con: Multiple comparisons	Rev, DIF: Comprehensive bias review
ELDA	ELDA Consortium States[c]	Corr: Moderate to high reliability	Content: Expert panel review	C, P, Con: Multiple comparisons	DIF: Few items problematic and removed
Commercial tests					
LAS Links	Connecticut, Indiana, Hawaii	Corr, RR: High reliability	Content: Expert panel review	C, Con: LAS Links vs. older LAS; moderate correlation	Rev: Expert panel item review DIF: examined for gender bias
MAC II	Missouri	Corr, RR: Moderate to high reliability	Construct: Test internal structure based on linguistic theory	C, P, Con: MAC II vs. various proficiency tests, ELL classifications	Rev: Bias review panel during item development
SELP	Virginia, Wyoming, Mississippi	Corr: High reliability	Construct: Test internal structure based on linguistic theory	C, Con: ELL / Non-ELL classification accuracy	Rev: Bias review panel during item development
State-developed tests					
CELDT	California	Corr, RR: Moderate reliability	NA	C, P: CELDT v. ELP level classifications	Rev: Expert bias review
ELPA	Michigan	Corr: Low to high reliability	Content: Expert panel review	C, P: Comparison with NS sample, SELP, SAT9	DIF: Items flagged for DIF examined by panel
IELA	Idaho	Corr: Moderate to low reliability	NA	C, Con: predict student's current ELL status	NA
MEPA-R/W and MELA – O	Massachusetts	Corr: High reliability	Content: Expert panel review	NA	DIF: Items flagged for DIF examined by panel

Test name	State	Reliability	Construct / content validity	Criterion validity[a]	Bias / fairness study
State-developed tests (continued)					
Oregon ELPA	Oregon	NA	NA	C, Con: ELPA versus New IPT, Woodcock-Muñoz, LAS proficiency designations	NA
TEAE	Minnesota	NA	Content: Expert panel review	NA	Rev: Bias review panel during item development
TELPAS	Texas	Cor: High reliability	Content: Expert panel review	C, Con: ELP scores v. their English language arts (ELA) scores	Rev: Expert panel review
WLPT-II	Washington	Cor, RR: High reliability	Content: Expert review Construct: Internal structure, PCA, SEM[d]	NA	Rev: Expert panel review

Note. ELP = English language proficiency. ELL = English language proficiency. NA = Sufficient information was not made available for this report. CRESST does not know if adequate work has been done in this area. SAT-9 = Stanford Achievement Test, Ninth Edition. PCA = Principal component analysis. SEM = Structural equation modeling.

[a]We include here studies of criterion, predictive, and concurrent validity.

[b]WIDA Consortium consists of 14 states and the District of Colombia: Alabama, Delaware, Georgia, Illinois, Kentucky, Maine, New Jersey, New Hampshire, North Dakota, Oklahoma, Pennsylvania, Rhode Island, Vermont, and Wisconsin.

[c]ELDA consortium consists of 7 states: Iowa, Kentucky, Louisiana, Nebraska, Ohio, South Carolina, and West Virginia.

[d]Duran, R. P. & Lee, Y. (2007, June). *English language proficiency tests, one dimension or many?: The WLPTI*. Presented at CCSSO Large-Scale Assessment Conference, Nashville, TN.

Appendix J2

Examples of Validation Studies

1. WIDA Consortium — Assessing Comprehension and Communication in English State–to–State for English Language Learners (ACCESS for ELLs®)

The Assessing Comprehension and Communication in English State–to–State for English Language Learners (ACCESS for ELLs®) test was developed in 2003–05 by a consortium of states working in concert with the Center for Applied Linguistics (CAL) as a measure of English language development in schools. It consists of multiple-choice items testing listening and reading, and performance tasks of writing and speaking. Consortium member states first agreed on a common set of English language proficiency standards, which became both the official standards for each state and the theoretical construct of the ACCESS for ELLs® test. Teams of content experts and professional item writers created items according to Harcourt specification. Four grade bands were identified, and separate forms were created for each grade band: K–2, 3–5, 6–8, and 9–12. Field testing was conducted in 2004 in two of the member states, Illinois and Wisconsin, and 6662 students participated in the field test, approximately evenly distributed among the four grade bands (the kindergarten portion was field tested separately). Sixty-one percent of the field sample participants were native speakers of Spanish, with the remainder coming from many diverse language groups (Among speakers were the second most common group, representing 13% of the total).

A number of reliability analyses were carried out, at the item, subscore, and composite score levels. Stratified Cronbach's alpha estimates were used to estimate the internal consistency of each subscore and for each grade band, using samples of students from the first operational test administration. This information was further compiled in an estimate of the accuracy of the overall cut score. All estimates for the overall cut scores were above .90 with most around .940. The accuracy of cut scores was estimated by examining the standard errors of measurement (SEM) at each cut score for each subsection. Additional analyses were conducted at the item level using the Rasch measurement model used to develop the test. In summary, the developers were satisfied that ACCESS for ELLs® was sufficiently reliable to use as the basis for making high-stakes decisions.

A number of criterion-related validity studies have been performed by CAL and member states, and according to the technical report, other studies are currently in progress. In the first study, ACCESS for ELLs® scores were compared to a priori ELP categorizations from field test participants. The mean scores on ACCESS for ELLs® increased as students' ELP category increased for all modalities, except in some cases where sample sizes were small. This was interpreted as evidence that ACCESS for ELLs® can be used to place students into hierarchical ability levels. In the second study, concurrent validity was investigated by concurrently giving ACCESS for ELLs® and one "older generation" ELP test to 4,985 students in two member states. The tests used were the New IDEA Proficiency Test (New IPT), the Language Assessment Survey (LAS), the Maculaitis Assessment of Competencies Test of English Language Proficiency (MAC II), and the Language Proficiency Test

Series (LPTS). Overall, the study found moderate to strong correlations between ACCESS for ELLs® and the older tests, with .60 representing an approximate average. The findings were consistent with the developers' views of the differences between the academic English that the ACCESS for ELLs® test was intended to measure and the more social language on the older tests' constructs.

A study of content validity compared items, a priori proficiency level (the proficiency level the item was designed to target) against the items' difficulty. For every modality, items developed to measure higher proficiency levels had higher average difficulty than items developed for lower proficiency levels. Approximately 6,500 students' scores were used for this study from the field test data. Results were interpreted as evidence that the items were empirically ordered by difficulty as predicted by the WIDA standards. Additionally, items were reviewed for content validity by a panel of member state education experts at the item development stage, and this process is ongoing as new items are developed for ACCESS for ELLs® (33% annual replacement rate for test items). After items pass the content review stage, they are examined for bias/fairness by a separate panel of bias experts from diverse ethnic groups. The test's internal structure was studied by examining the correlations among subscales; the test developers found a moderately strong correlation among the various modalities, consistent with theory and prior research suggesting that while related, these modalities may be considered to some extent separate abilities, which may be tested separately.

Overall, the ACCESS for ELLs® test is an excellent example of current best practice in ELD testing. The test developers closely aligned the test to state standards (in this case, actually introducing new standards), and conducted rigorous, well documented studies in all major areas of testing validity. Further, they committed to ongoing review and item replacement and have committed substantial resources to this important project. Overall, the studies demonstrate convincingly that ACCESS for ELLs® may be an appropriate measure for the purpose of tracking ELL students' progress. If states desire to use ACCESS for ELLs® based on this comprehensive and careful validity work, however, the issue of standards should be addressed: What is the relationship between a state's current ELP standards and the WIDA consortium standards? If sufficient overlap is found, or if a state decides to adopt WIDA standards, ACCESS for ELLs® would be a recommended tool.

2. SCASS Consortium — ELDA

The English Language Development Assessment (ELDA) was developed by member states of the Assessing Limited English Proficient Students group of the State Collaborative on Assessment and Student Standards (LEP-SCASS). The ELDA was funded jointly by the LEP-SCASS member states, Council of Chief State School Officers (CCSSO), and the U.S. Department of Education. The ELDA was designed to measure annual progress in English language proficiency for ELL students. It consists of separate tests for listening, speaking, reading, and writing at each of three grade clusters: 3–5, 6–8, and 9–12. Item types include multiple-choice, short constructed response, and extended constructed response. ELP standards were developed based on previously existing state-level standards, which became the theoretical construct of the ELDA. The K–2 instruments are composed of teacher surveys and skills inventories as opposed to multiple-choice or constructed

response items. Pilot testing took place in 2003; at this stage, 310 students in total across 31 schools participated, and the pilot test was intended to check test feasibility and the clarity of instructions and item types. A multi-state field test was conducted in 2004, and further field testing along with the first operational use of the ELDA was conducted in 2005. Item difficulty ranged from .54 to .81, and item discrimination, calculated using the point biserial correlations between items and total scores, ranged from .47 to .87.

For the Grades 3–12 test forms, Cronbach's alpha internal consistency estimates ranged from .76 to .95, with generally lower results for the writing sections, possibly due to the lower number of items and greater variety of items types there. The technical report did not detail what actions, if any, were taken to improve the reliability of the writing sections.

A number of validity studies were carried out. First, latent class analysis was used to estimate the differences in proportion correct for students across five levels of ability; items were analyzed for their ability to distinguish between students at different ability levels (the latent classes). To examine criterion-related validity, several studies were conducted. Teacher ratings of all field test participants were used to group participants into five ability levels, and the mean item scores (percent correct) were found to increase as teacher ratings of participants increased. Next, items were analyzed by applied linguistics experts and were classified according to "developmental level," or the general ability level that the item appeared to be targeting. All of these data were combined in a matrix to identify "strong" and "weak" items — in other words, the degree to which items appeared to be discriminating appropriately at various levels of student ability. Finally, an analysis of variance (ANOVA) was used to estimate the relationships between item difficulties and the teacher ratings, as well as between item difficulties and item developmental levels. Main effects were significant in all cases; for all grades bands and forms, teacher ratings and item developmental level ratings predicted student performance. This was seen as strong evidence of criterion-related validity.

In addition to the above validity analyses, the relationships between the ELDA, the Language Assessment Survey (LAS), the New IDEA Proficiency Test (New IPT), and teacher ratings of student proficiency were examined using multitrait-multimethod analyses. ELDA scores were more closely associated with the teacher ratings, which was seen as evidence of test validity because ELDA was designed to measure classroom language ability, as opposed to the social language constructs of the other tests. A number of further analyses were carried out, which are detailed in the ELDA/SCASS technical report.

Item bias/fairness was first addressed at the item development stage, where a trained and certified bias review expert examined potential test items and recommended modifications or the removal of potentially problematic items. Further analyses were conducted using the field test data; all items on all forms were examined for DIF by gender, Spanish L1 vs. other L1, and current/exited ELL groupings. Relatively few items were found to have DIF and of these, only a few were removed after close consideration of the item content.

In summary, the ELDA developers have addressed the major validity questions through rigorous applied research, and have articulated how the findings of their studies address these questions. In particular, the attention paid to bias and fairness is noteworthy, in that the developers incorporated attention to bias at both the item design stage and at the field trial stage, and used a rigorous screening process. In addition, the developers compared the ELDA against multiple validity criteria (teacher ratings, student ELP levels, other ELP tests) rather than against a singe criterion. Longer term, predictive validity studies as well as qualitative studies of test-taker response processes were not detailed in the technical report and are recommended for further validity study.

3. California — California English Language Development Test (CELDT)

The California English Language Development Test (CELDT) was developed in 2000 by the California State Department of Education (CDE) in conjunction with CTB McGraw Hill in order to comply with California laws passed in 1997 and 1999. It was first administered in 2001, and in addition to the initial test development validity work, yearly analyses are carried out by CTB McGraw Hill and made available on the CDE Web site.

Field testing and initial item analyses were conducted in 2000. Items were derived from two sources: Some items were carried over from the Language Assessment Scales (LAS) test and other items were created by the development team in order to comply with state ELP standards. The test was built around the four grade spans identified in the state ELP standards: K–2, 3–5, 6–8, and 9–12. Participants for the field test were sampled from school districts across the state and from a wide range of language backgrounds, including non-ELL students, and all grades. The majority of participants across all subtests (reading, writing, listening-speaking) and grades were Spanish speaking students. The total sample size (including all language backgrounds and grades) was 13,947 for listening-speaking and 12,067 for reading and writing. Items were analyzed using two different Item Response Theory (IRT) models: the three-parameter dichotomous model for the multiple-choice items, and the two-parameter partial credit model for the constructed-response items. Subsequent to the item analysis, standard setting was conducted by 95 content experts and teachers using the bookmark method. Cut scores were established and used for the first operational trial of the CELDT in 2001.

In 2001, the first operational year, approximately 1.6 million students took the test, and a number of studies were conducted at that time. Reliability was estimated using Cronbach's alpha for the three subtests and grade spans separately. Estimates ranged from .64 (listening-speaking, Grades 6–8) to .91 (reading, Grades 3–5). Estimates were generally lower for listening-speaking than for reading or writing. Standard errors of measurement (SEM) were also examined with analogous results (higher SEM for listening-speaking). For the 2003–2004 administration, a study of rater agreement (reliability) in the writing section found rater agreement levels between 69% and 89% for constructed response items, depending on grade level and item type (sentence writing and essay). Test-retest reliability was determined to be between .85 and .90 using embedded items to simulate a parallel-forms reliability design, again depending on grade level and test domain. The technical report does

not discuss any actions that may have been taken in response to the findings of lower reliability for some sections and grades.

Yearly criterion-related validation studies have also been conducted by CTB McGraw Hill. For the 2001 test, a validation study was conducted in which 22 experts each compared the performance of 40 students on external criteria against CELDT classifications. Cut scores were found to be reasonably accurate for Grades 6–8 and for Grades 9–12. The panel recommended that cut scores for Grades 3–5 be studied further. K–2 was not studied. A cut score validation study in 2003 compared qualitative assessments of 600 ELL students' language ability with their CELDT scores and concluded that the two assessments achieved comparable results, with disagreement of more than +/- 1 proficiency level occurring less than 10% of the time, supporting the validity of the cut scores for the various CELDT proficiency levels. Qualitative analyses included an item content review by an expert panel. The technical report does not discuss any actions that may have been taken in response to this validation work. To date, no comprehensive studies of content validity or bias/fairness have been reported for the CELDT beyond any analyses which may have been conducted during the initial item selection/development phase.

Finally, modifications and corresponding studies have been made to the CELDT in each year. In 2002, the reading section was redesigned to flow better and be administered more quickly. A qualitative study was conducted wherein classrooms were observed taking the newly designed form, and it was concluded that the new design was satisfactory. The listening and speaking test was revised in 2003 in response to requests from the field to ease test administration. The listening portion was redesigned to be group administered, and the speaking portion was redesigned to take less time (about 10 minutes per student). A small-scale trial was conducted at this time.

Although sophisticated statistical work has been conducted on CELDT test items, and test score reliability seems to be well studied, substantial work remains to be done on this extremely large-scale test. No comprehensive bias or fairness study appears to have been done; the technical reports, for example, do not mention any study of differential item functioning by ethnicity or by SES, to mention two of the most common potential areas of DIF. Further, the criterion-related validity evidence of the CELDT appears to be limited to two studies of cut score validity. A comprehensive link between CELDT scores and external criteria such as scores on similar tests, classroom teacher evaluations, etc., has not been established in the CELDT technical literature. If such work has been done, either by CDE or by outside researchers, it should be reported in this literature. Finally, the available literature does not discuss what decisions or modifications, if any, were made to CELDT as a result of the existing research findings.

4. CTB McGraw-Hill — Language Assessment Scales Links (LAS Links)

The Language Assessment Scales Links (LAS Links) was developed by CTB McGraw-Hill in response to NCLB legislation. It is intended to measure English language development from kindergarten to 12th-grade across the four modalities on a common scale, in order to accurately measure growth and help states comply with NCLB mandates. State standards and content experts

were consulted in order to ensure that LAS Links content corresponded to typical state ELP standards; the test was not designed to align with any one state's standards. Five grade spans are identified and separate sets of forms are used for each: K–1, 2–3, 4–5, 6–8, and 9–12. During field testing, more than 1,000 test takers took each form at each grade level. Diverse ethnic groups and students with various first language backgrounds participated in the field testing.

Reliability on LAS Links was calculated as the Standard Error of Measurement (SEM) for each subsection and grade band. The technical report lists the SEM for each raw score at each grade band and subsection. Additionally, rater agreement was calculated for those sections requiring constructed responses. This agreement was calculated as a series of intra-class correlations, and reported individually for each constructed response item; correlations ranged from .87 to .99 with most values at .95 or above.

The evidence for criterion-related validity for LAS Links came from a study in which students in each grade band took both the older LAS and the new test concurrently. Sample sizes ranged from 307 to 819, but no information was provided as to participant characteristics other than grade. Correlations between the two tests ranged from .48 for the oral section in Grades K/1 to .92 for the reading and writing section in Grades 4 and 5. Content validity was addressed during test development by consultations with experts in the fields of applied linguistics and education, as well as attention to the APA, AERA, and NCME (1999) guidelines.

Test bias/fairness was addressed through continuous attention to test content at the item development stage, as well as a series of tryouts with education experts of various ethnicities who examined the test items for possible content bias. Finally, DIF analyses were conducted for gender using the IRT method and the tryout data; no information was provided as to results of these DIF analyses.

The LAS Links test appears to be carefully designed using modern principles of test design and sophisticated technical analyses. Further, it is encouraging that the test publishers noted their attention to research in second language acquisition and education. Further, the test appears to have high reliability, though some means of interpreting the raw data provided in the technical manual should be provided. However, more work on test validity, specifically criterion-related validity, is recommended. In addition, states using the LAS Links should study the degree of correspondence between the LAS Links content and their particular state ELP standards; no such studies have been found, yet this correspondence is crucial in determining whether decisions based on LAS Links can be justified.

Appendix K1

Accommodations Involving Changes to the Timing or Scheduling Provided to ELL Students by State

State	Extended time	Frequent breaks	Testing over several sessions/days	Flexible scheduling	Tested at time most beneficial to students
AK	X	X	X	—	—
AL	X	—	X	X	X
AR	X	—	—	X	—
AZ	X	X	X	—	—
CA	X	X	—	X	—
CO	X	—	—	—	—
CT	X	—	—	—	—
DC	—	X	X	X	—
DE	—	X	X	—	—
FL	X	X	—	X	—
GA	X	X	—	—	—
HI[a]	—	—	—	—	—
IA	X	X	—	—	—
ID	X	X	X	X	—
IL	X	—	—	—	—
IN	X	X	X	—	—
KS	X	—	—	—	—
KY	X	—	—	—	—
LA	X	X	X	—	—

Appendix K1-81

State	Extended time	Frequent breaks	Testing over several sessions/days	Flexible scheduling	Tested at time most beneficial to students
MA	—	—	—	—	—
MD	X	X	X	X	X
ME	X	X	X	X	X
MI	X	X	—	X	X
MN	—	—	—	—	—
MO	X	—	X	—	—
MS	X	X	X	—	—
MT	X	X	—	—	X
NC	X	—	X	—	—
ND	—	—	—	—	—
NE	X	X	—	—	X
NJ	X	—	—	—	—
NM	X	X	X	X	X
NV	X	—	—	—	—
NY	X	—	—	—	—
OH	X	—	X	—	—
OK	X	X	X	—	—
OR	X	X	X	—	X
PA	X	X	X	—	X
RI	X	X	—	—	X
SC	X	X	X	—	—
SD	X	X	X	X	X
TN	X	—	—	—	—

Appendix K1-82

State	Extended time	Frequent breaks	Testing over several sessions/days	Flexible scheduling	Tested at time most beneficial to students
TX	—	—	—	—	—
UT	X	X	—	—	—
VA	—	X	X	X	—
VT	—	X	—	—	X
WA	—	X	X	—	X
WI	X	X	X	—	—
WV	—	X	—	X	—
WY	—	X	X	—	—
Total	**38**	**31**	**24**	**13**	**13**

Note. Allowable accommodations may vary by specific grades or subsections of achievement tests, depending on state.

— = state does not allow.

[a]For Hawaii, specific accommodations allowed were not found.

Appendix K2

Accommodations Involving Changes to the Setting Provided to ELL Students by State

State	Test individually administered	Test administered in small group	Test administered in separate location[a]	Test administered in ESL or bilingual classroom	Test taker provided preferential seating	Test administered by other school personnel[b]	Test administered in non-school setting[c]	Special or adaptive classroom equipment, furniture or lighting provided	Other[d]
AK	X	X	—	—	—	X	—	—	—
AL	X	X	X	X	—	X	—	—	X
AR	X	X	X	—	X	—	—	—	—
AZ	X	X	X	—	X	X	—	X	—
CA	X	X	X	—	—	—	—	—	—
CO	—	—	—	—	—	—	—	—	—
CT	—	—	X	—	—	—	—	—	—
DC	X	X	—	—	X	—	—	—	—
DE	X	X	—	—	—	—	—	—	—
FL	—	X	X	—	—	X	—	—	—
GA	X	X	X	X	X	—	—	—	—
HI[a]	—	—	—	—	—	—	—	—	—
IA	X	X	X	—	—	—	—	—	—
ID	X	X	X	X	X	—	—	—	X
IL	—	—	X	—	—	—	—	—	—
IN	X	X	—	—	—	—	—	—	—
KS	—	X	—	—	—	—	—	—	—
KY	—	X	—	—	—	—	—	—	—

Appendix K2-84

State	Test individually administered	Test administered in small group	Test administered in separate location[a]	Test administered in ESL or bilingual classroom	Test taker provided preferential seating	Test administered by other school personnel[b]	Test administered in non-school setting[c]	Special or adaptive classroom equipment, furniture or lighting provided	Other[d]
LA	X	X	—	—	—	X	—	—	—
MA	—	—	—	—	—	—	—	—	—
MD	X	X	X	—	X	X	—	X	X
ME	X	X	X	—	—	X	X	—	—
MI	—	X	—	X	X	X	—	—	X
MN	—	—	—	—	—	—	—	—	—
MO	X	X	—	—	—	—	—	—	—
MS	—	X	X	—	—	X	—	—	X
MT	X	X	X	—	—	X	X	—	X
NC	X	X	X	—	—	—	—	—	—
ND	—	—	—	—	—	—	—	—	—
NE	—	—	X	—	—	—	—	—	—
NJ	—	—	—	—	—	—	—	—	—
NM	X	X	X	X	X	—	—	—	—
NV	X	X	X	—	—	X	—	—	—
NY	X	X	X	—	—	—	—	—	—
OH	—	—	X	—	—	—	—	—	—
OK	X	X	X	X	—	—	—	—	—
OR	X	X	X	—	X	—	—	X	X
PA	—	X	X	—	—	—	—	X	—
RI	—	X	X	—	X	X	X	X	X

Appendix K2-85

State	Test individually administered	Test administered in small group	Test administered in separate location[a]	Test administered in ESL or bilingual classroom	Test taker provided preferential seating	Test administered by other school personnel[b]	Test administered in non-school setting[c]	Special or adaptive classroom equipment, furniture or lighting provided	Other[d]
SC	X	X	—	—	—	X	—	—	—
SD	X	X	X	—	X	X	—	X	—
TN	—	—	—	—	—	—	—	—	—
TX	—	—	—	—	—	—	—	—	—
UT	X	X	X	—	—	X	—	—	—
VA	X	X	X	—	X	—	—	—	—
VT	X	X	X	—	X	X	X	X	—
WA	X	X	X	—	X	—	—	X	—
WI	X	X	X	—	—	—	X	X	X
WV	—	—	—	—	—	—	—	—	—
WY	—	X	X	—	—	—	—	—	—
Total	**29**	**37**	**31**	**6**	**14**	**16**	**5**	**9**	**9**

Note. Allowable accommodations may vary by specific grades or subsections of achievement tests, depending on state.

— = state does not allow.

[a]Including in a study carrel, non-regular classroom, or somewhere quiet with minimal distractions.

[b]Test administered by ESL or bilingual teacher or person familiar with test taker.

[c]For example, at home.

[d]For example, opportunity for student to move, stand or pace; supporting physical position of student; test administered with teacher facing student; test administered with student seated in front of classroom.

[e]For Hawaii, specific accommodations allowed were not found.

Appendix K3

Accommodations Involving Changes to the Presentation Provided to ELL Students by State

State	Directions read[a] aloud in English	Test items read aloud in English[b]	Directions simplified or paraphrased	Directions clarified or explained	Directions translated (oral or written)	Dictionaries, glossaries, or word lists (bilingual)	Dictionaries, glossaries, or word lists (other[c])	Dual language or bilingual test versions	Translated or native language test versions	Test items read aloud in native language	Audio recording or similar technology	Other[d]
AK	X	X	X	X	X	X	—	—	—	—	—	X
AL	—	—	—	—	X	X	X	—	—	—	—	X
AR	X	X	—	—	—	X	—	—	—	—	—	X
AZ	X	X	X	X	X	X	—	—	X	—	—	X
CA	—	—	—	—	X	X	—	—	—	—	—	X
CO	X	X	—	—	X	X	X	—	—	X	—	—
CT	—	X	—	—	X	X	—	—	—	—	—	—
DC	X	—	X	—	—	—	—	—	—	—	—	X
DE	X	—	X	—	X	—	X	X	X	X	—	—
FL	—	—	—	X	—	X	—	—	—	—	—	X
GA	X	X	—	X	—	X	—	—	—	—	—	X
HI[e]	—	—	—	—	—	—	—	—	—	—	—	—
IA	X	X	—	X	X	—	—	X	X	—	X	—
ID	X	X	X	X	X	X	X	—	—	—	X	X
IL	—	—	—	—	X	—	—	—	—	—	X	X
IN	X	X	—	—	—	X	—	—	—	—	—	—
KS	—	X	—	—	X	X	—	X	X	X	X	X
KY	X	X	X	—	X	X	—	—	—	X	—	X

State	Directions read[a] aloud in English	Test items read aloud in English[b]	Directions simplified or paraphrased	Directions clarified or explained	Directions translated (oral or written)	Dictionaries, glossaries, or word lists (bilingual)	Dictionaries, glossaries, or word lists (other[c])	Dual language or bilingual test versions	Translated or native language test versions	Test items read aloud in native language	Audio recording or similar technology	Other[d]
LA	X	X	—	—	X	X	—	—	—	—	—	—
MA	—	—	—	—	—	X	—	—	X	—	—	—
MD	X	X	—	—	—	X	—	—	—	—	X	X
ME	X	X	—	—	—	X	—	—	—	—	X	X
MI	X	X	—	X	X	X	—	—	—	X	X	X
MN	X	X	—	—	X	X	—	—	X	—	X	—
MO	—	X	—	—	—	X	—	—	—	X	—	—
MS	X	X	X	—	—	X	X	—	—	—	—	—
MT	X	X	—	—	X	X	—	—	—	—	—	X
NC	—	X	—	—	—	X	—	—	—	—	—	X
ND	—	X	X	X	—	—	—	—	—	—	—	—
NE	—	—	—	—	—	X	—	X	—	—	—	X
NJ	—	—	—	—	X	X	—	—	—	—	—	—
NM	X	X	—	—	X	X	—	—	—	—	—	—
NV	X	X	—	—	—	X	X	—	—	X	X	X
NY	—	—	—	X	—	X	—	X	X	X	—	X
OH	—	X	—	—	—	X	X	X	—	X	X	X
OK	X	X	X	X	X	X	—	—	—	X	X	—
OR	X	X	X	X	X	X	—	X	X	—	X	X
PA	X	X	X	—	X	—	X	X	—	—	X	X
RI	X	X	—	—	X	X	—	—	—	—	—	X

State	Directions read aloud[a] in English	Test items read aloud in English[b]	Directions simplified or paraphrased	Directions clarified or explained	Directions translated (oral or written)	Dictionaries, glossaries, or word lists (bilingual)	Dictionaries, glossaries, or word lists (other[c])	Dual language or bilingual test versions	Translated or native language test versions	Test items read aloud in native language	Audio recording or similar technology	Other[d]
SC	X	X	X	—	X	X	—	—	—	—	—	X
SD	X	X	X	—	—	X	—	—	—	—	—	—
TN	X	X	—	—	—	X	—	—	—	—	—	—
TX[f]	—	X	—	—	—	X	—	X	X	X	—	X
UT	X	X	X	X	X	X	—	—	X	X	X	X
VA	X	X	X	X	—	X	—	—	—	—	—	X
VT	—	X	—	X	X	X	—	—	—	—	—	X
WA	X	X	—	—	X	X	X	—	—	—	—	X
WI	X	X	X	—	X	X	X	—	—	X	X	X
WV	X	X	X	—	—	X	—	—	—	—	—	—
WY	X	X	X	—	X	X	—	—	—	X	—	X
Total	**33**	**39**	**18**	**14**	**28**	**43**	**10**	**9**	**10**	**14**	**15**	**32**

Note. Allowable accommodations may vary by specific grades or subsections of achievement tests, depending on state.

— = state does not allow.

[a]Including repeated or re-read.

[b]In some subsections (not in Reading).

[c]English only or not specified.

[d]Other includes using place markers; highlighting key words or phrases; providing electronic translators; allowing the student to self-vocalize; using templates or color overlays; administering using sheltered English; providing written versions of oral directions; using noise buffers.

[e]For Hawaii, specific accommodations allowed not found.

[f]Accommodations in this table and other tables for Texas are those allowed for students designated as LEP-exempt, which is a select category of ELL students determined by the school's Language Proficiency Assessment Committee (LPAC). They are offered Linguistically Accommodated Testing (LAT) only for math tests and only in Grades 3–8 and 10. No general ELL accommodations were found.

Appendix K3-89

Appendix K4

Accommodations Involving Changes to the Response Provided to ELL Students by State

State	Test taker marks answers in booklet	Test taker dictates to or uses a scribe to respond in English	Test taker dictates responses in native language	Test taker's response recorded	Test taker uses computer, word processor, typewriter	Test taker points to or indicates response	Provision of spelling or grammar checker	Test taker verifies understanding of directions	Other
AK	—	—	—	—	—	—	—	X	—
AL	X	—	—	—	X	—	—	—	—
AR	—	—	—	—	—	—	—	—	—
AZ	—	—	—	—	—	—	—	—	—
CA	—	—	—	—	—	—	—	—	—
CO	—	—	X	—	—	—	—	—	—
CT	—	—	—	—	—	—	—	—	—
DC	—	—	—	—	—	—	—	—	—
DE	—	X	—	—	—	—	—	—	X
FL	—	—	—	—	—	—	—	X	—
GA	X	X	—	—	—	—	—	—	—
HI[a]	—	—	—	—	—	—	—	—	—
IA	X	—	X	—	—	X	X	—	—
ID	X	X	X	—	—	—	—	X	—
IL	—	X	—	—	—	—	—	—	—
IN	—	—	—	—	—	—	—	—	—
KS	—	—	—	—	—	—	—	—	—

State	Test taker marks answers in booklet	Test taker dictates to or uses a scribe to respond in English	Test taker dictates responses in native language	Test taker's response recorded	Test taker uses computer, word processor, typewriter	Test taker points to or indicates response	Provision of spelling or grammar checker	Test taker verifies understanding of directions	Other
KY	—	X	—	—	X	—	—	—	X
LA	—	—	—	—	—	—	—	—	—
MA	—	—	—	—	—	—	—	—	—
MD	X	X	—	X	X	—	X	—	X
ME	—	—	—	—	X	—	—	X	—
MI	X	X	—	—	—	—	—	X	—
MN	—	X	—	—	—	—	—	—	—
MO	—	X	—	X	X	—	—	—	—
MS	X	X	—	—	—	—	—	—	—
MT	—	X	—	—	X	—	—	X	—
NC	—	—	—	—	—	—	—	—	—
ND	—	—	—	—	—	—	—	—	—
NE	—	X	—	X	X	—	—	—	X
NJ	—	—	—	—	—	—	—	—	—
NM	—	—	X	—	—	—	—	—	—
NV	—	—	—	—	—	—	—	—	—
NY	—	—	—	—	—	—	—	—	X
OH	—	X	X	—	X	—	—	—	—
OK	X	X	—	—	—	—	—	—	—
OR	X	X	—	X	X	X	—	—	X
PA	X	X	X	—	X	—	—	—	X

Appendix K4-91

State	Test taker marks answers in booklet	Test taker dictates to or uses a scribe to respond in English	Test taker dictates responses in native language	Test taker's response recorded	Test taker uses computer, word processor, typewriter	Test taker points to or indicates response	Provision of spelling or grammar checker	Test taker verifies understanding of directions	Other
RI	X	X	—	—	X	X	—	—	—
SC	X	—	—	—	—	—	—	—	—
SD	—	—	—	—	—	—	—	—	—
TN	—	—	—	—	—	—	—	—	—
TX	—	—	—	—	—	—	—	—	—
UT	—	—	—	—	—	—	—	—	X
VA	X	X	—	X	—	—	—	—	—
VT	X	X	—	X	X	X	—	—	—
WA	—	—	—	—	—	—	—	—	X
WI	X	—	X	—	—	—	X	—	—
WV	X	X	—	—	X	—	—	—	—
WY	—	—	—	—	—	—	—	—	—
Total	**16**	**20**	**7**	**6**	**13**	**4**	**3**	**6**	**9**

Note. Allowable accommodations may vary by specific grades or subsections of achievement tests, depending on state.

— = state does not allow.

[a] For Hawaii, specific accommodations allowed were not found.