

## **LINGUISTIC MODIFICATION**

### **Part I: Language Factors in the Assessment of English Language Learners: The Theory and Principles Underlying the Linguistic Modification Approach**

by Jamal Abedi, Ph.D.

University of California, Davis and the Advance Research and Data Analyses Center

### **Part II: A Guide to Linguistic Modification: Increasing English Language Learner Access to Academic Content**

by Edynn Sato, Ph.D.

Assessment and Accountability Comprehensive Center and WestEd

Prepared for the U.S. Department of Education: LEP Partnership

## Contents

### **Part I: Language Factors in the Assessment of English Language Learners: The Theory and Principles Underlying the Linguistic Modification Approach**

Introduction	4
Background and Rationale	5
How ELL Students are Performing Under the Current Educational System	7
How Can the Validity of Assessment for ELL Students be Improved?	11
What is Linguistic Modification/Language Simplification?	14
Linguistic Modification Approach: A Brief Overview	14
Vocabulary	16
Grammar	17
Style of Discourse	22
Problems Specific to Math	23
Research on the Effectiveness and Validity of Linguistic Modification as a Tool for Improving Validity of Assessment for ELL Students	25
Determining the Comparability of Scores from Linguistically Modified Assessments	30
Impact of Linguistic Modification of Assessment on Psychometric Characteristics of Assessments for ELL Students	32
What Does Research Suggest?	34
Discussion	35
References: Part I	37
Appendix A. Tables	43
Appendix B. Linguistic Features That May Affect Comprehension	47

**Part II: A Guide to Linguistic Modification:**

**Increasing English Language Learner Access to Academic Content**

Introduction	53
Linguistic Modification	54
Critical Considerations for Linguistically Modified Assessments	55
Consideration of the English Language Learner Population	56
Consideration of Access and Implications for Validity	57
Developing a Linguistically Modified Assessment: Increasing Access for ELLs	59
Defining the Assessed Domain and Constructs	60
Distinguishing Construct-Relevant from Construct-Irrelevant Language	62
Defining the English Language Learner Population	63
Specifying Linguistic Modification Strategies—Current Research and Recent Refinements	66
Current Research: General Guidelines for Linguistic Modification	66
Recent Research: Implications for Refining Linguistic Modification Strategies	68
Evaluating the Effectiveness of Linguistic Modification	70
Examining the Effect of Linguistic Modification	70
Content and Bias Reviews: Consideration of the ELL Population	72
User’s Guide: An Approach to Developing a Linguistically Modified Assessment	74
References: Part II	81
Tables	89

**Part I: Language Factors in the Assessment of English Language Learners:  
The Theory and Principles Underlying the Linguistic Modification Approach**

Jamal Abedi, Ph.D.

University of California, Davis

and

Advance Research and Data Analyses Center

**Introduction**

Federal legislation demands a fair assessment and accountability system for all children, including English language learner (ELL) students as the number of these children continues to grow rapidly. However, major issues remain regarding the assessment of ELL students. For example, research on the assessment and accommodation of ELL students shows a substantial performance gap between these students and their native English-speaking peers, which could be attributed at least partly to the impact of linguistic biases in the assessments. Many different accommodations have been proposed to remedy this situation. However, some of these accommodations do not appear to help ELL students specifically with their language needs. Among the language-related accommodations, linguistic modification is the most promising as it does not affect the validity of assessment and at the same time helps to narrow the performance gap between ELL and non-ELL students. This project aims to present a comprehensive view of the theory and application of the linguistic modification approach to assist states with their decision in selecting and using appropriate accommodations for ELL students. The project consists of two independent yet related parts. In Part I of this project, which is the subject of this report, the theory and principles underlying linguistic modification of assessment will be discussed

and a summary of recent studies on the effectiveness and validity of assessment using the linguistic modification approach will be presented. In Part II, practical considerations and guidelines for developing linguistically modified versions of assessments will be discussed.

### **Background and Rationale**

English language learners face serious challenges in their academic careers: the challenge of learning both social and academic English, while simultaneously learning academic content taught in English. Learning English as a second or additional language is a difficult task and requires time. This adds to the cognitive load that ELL students have to bear in their academic careers. It is quite a difficult task for these students to learn content in a language that they are still struggling to learn. Based on research findings, it takes from five to seven years and even more for most ELLs to gain sufficient mastery of academic English to join English-speaking peers in taking full advantage of instruction in English (Hakuta, Butler, & Witt, 2000). While ELL students are struggling to learn English, learning content-based knowledge cannot occur at the same rate as for a native speaker of English when that instruction is offered only in English. Limited English proficiency may also make it difficult for ELL students to benefit fully from the teacher's instructions and to understand assessment questions. Therefore, limited English proficiency affects both learning and assessment. To help close the performance gap between ELL and non-ELL students both learning and assessment conditions must be addressed.

Attention to the educational needs of ELL students is critical since these students, their teachers and their parents are under tremendous pressure to succeed under educational environments that may not be familiar to them culturally and linguistically. Moreover,

## LEP Partnership

examining issues concerning assessment and accountability of these students has high priority in the national agenda as the number of these students is increasing rapidly, particularly in recent years, and continues to grow. In 1990, 32 million people over the age of five in the United States spoke a language other than English in their home, comprising 14 percent of the total U.S. population. By 2000, that number had increased by 47 percent to 47 million, representing 18 percent of the population of the United States (U.S. Census Bureau, 2002). Thus, the population of English language learners is one of the fastest growing subgroups in the nation (Kindler, 2002) and now numbers over 5 million students, representing approximately 10 percent of all K-12 students nationally (GAO, 2006). Consequently, the fairness and validity of their assessment must be among the top priorities of the national education agenda.

With recent federal legislation (the No Child Left Behind [NCLB] Act of 2001, Goals 2000 and Titles I and VII of the Improving America's Schools Act of 1994 [IASA]) mandating inclusion of all students in state and national assessments, issues concerning assessment of ELL students become even more important. According to this legislation, states should provide reliable and valid annual academic assessment of students' English language proficiency, assess their content knowledge through tests that are free of cultural and linguistic biases, and provide testing accommodations to accurately and validly measure their academic achievement in state content assessments (GAO, 2006). However, as illustrated below, the reliability and validity of existing assessment instruments for ELL students may be questionable.

## **How ELL Students are Performing Under the Current Educational System**

Research findings on the assessment of ELL students have demonstrated a substantial performance gap between ELL and non-ELL students (see, for example, Abedi, 2006a; Abedi, Leon, & Mirocha, 2003; Solano-Flores & Trumbull, 2003). To illustrate this problem, we compared performance of ELL and non-ELL students using data from seven different locations nationwide. To maintain confidentiality, we refer to them as Site 1 through Site 7. Of the seven sites, four represent data from pre-NCLB assessments (1998-1999 school year) and three represent post-NCLB assessment data (2005-2006 school year). Comparing findings from these sites may help establish cross-validation evidence over time and, more importantly, provide information on the impact of NCLB in reducing the performance gap between ELL and non-ELL students.

Table 1 presents descriptive statistics for the ITBS subscales at the various grade and test level combinations based on the data from 1998-1999 in Site 2. As the data in Table 1 show, ELL students generally performed lower than their non-ELL peers. For non-ELL students the overall mean normal curve equivalent (NCE) scores were substantially higher than scores for ELL students. However, the gap between the test scores of ELL and non-ELL students depends on the grade level and the content of the assessment. The difference between the mean NCE scores of ELLs and non-ELLs was generally small for Grade 3 students except in reading (where there was over a seven-point difference) and favored the non-ELL group.

In Grade 4, the reading gap between ELL and non-ELL students was even larger. The mean reading score for Grade 4 non-ELL students was 45.44 (SD = 15.70) as

compared with a mean of 34.85 (SD = 12.77) for ELLs, a gap of more than two-thirds of a standard deviation.

The trend of increasing performance gaps between ELL and non-ELL students varies across the content/subsection areas. The largest gap between the two groups was in reading. This result was expected because the reading test items have presumably the highest language demand among the four content areas (math concepts and estimation, math problem solving, math computation and reading). Among these four content areas, the math computation subsection appears to have the lowest language demand.

Accordingly, the performance gap between ELLs and non-ELLs was the lowest on the math computation subsection. To compare ELL and non-ELL score differences across test, grade and content area, a *Disparity Index* (DI) of the performance of non-ELLs over ELLs was obtained. The DI was computed by subtracting the ELL subtest mean from the non-ELL subtest mean, dividing the difference by the ELL subtest mean, and multiplying the result by 100. The result gives the percentage by which the non-ELL mean exceeds the ELL mean on that particular subtest. A negative DI indicates that the ELL mean exceeds the non-ELL mean.

Table 2 presents the DIs comparing performance of ELLs with non-ELLs by grade and content area. The results in Table 2 show that, except for Grade 3 math computation, the DI percentages were all positive, indicating that in general the non-ELLs outperformed ELLs. Major differences between ELLs and non-ELLs were found for students in Grade 3 and above. The differences between the mean of ELL and non-ELL students increased sharply by grade. For example, in Grade 3 non-ELL students had Disparity Indices of 5.3% in math concepts and estimation, 11.1% in math problem solving and data



## LEP Partnership

interpretation, -3.1% in math computation (the ELL group performed better than the non-ELL group on this subtest), and 23.4% in reading. In Grade 4 these indices increased to 26.9% for math concepts and estimation, 19.3% for math problem solving and data interpretation, 6.9% for math computation, and 30.1% for reading. The indices further increased in Grade 5 to 36.5% for math concepts and estimation, 32.7% for math problem solving and data interpretation, 12.6% for math computation, and 41.1% for reading.

As the data suggest, the largest gap between ELLs and non-ELLs was in reading. The next largest gaps were in the content areas that appear to have more language content. For example, the math concepts and estimation and the math problem solving and data interpretation subsections seem to have more language content than the math computation subsection. The average DI for Grades 3 through 8 was 27.7% for math concepts and estimation. That is, the non-ELL group average in math concepts and estimation was 27.7% higher than the ELL group average. A similar trend was observed in math problem solving and data interpretation; the average DI for this subsection was 26.4%. However, the average DI for math computation was 9.0%, which is substantially lower than the corresponding DI for the other two math subsections. This lower performance gap between ELL and non-ELL students can be attributed to a lower level of language demand in the math computation subscale.

Analyses were also performed on the more recently administered assessments to enable us to compare the pre- and post-NCLB performance of ELL students and to examine possible changes in the performance gap between ELL and non-ELL students as a result of implementation of NCLB. The post-NCLB assessment came from Site 5, a state with a large enough number of ELL students to permit valid comparisons between the

## LEP Partnership

performance of ELL and non-ELL students. Comparisons were done on math, science and reading for Grades 5 and 8. Table 3 presents a summary of analyses for Grade 5 including the Disparity Indices, and Table 4 summarizes similar results for Grade 8.

Consistent with the pre-NCLB data presented above, the data show major performance gaps between ELL and non-ELL students in all content areas (reading, math and science) with a trend of higher performance gaps in higher grades. For Grade 5 students, the DI ranged from 21% for math to 37% for science. Once again, the performance gap is lower in math where there is less language demand in the assessments as compared with science where there is more language demand. Table 4 presents the summary analyses for Grade 8 students. As the data in Table 4 show, similar to what was presented for students in Grade 5 (Table 3), ELL students had lower performance, but the performance gap is higher for Grade 8 students as compared with the gap for Grade 5 students.

The DIs for students in Grade 8 (Table 4) were higher than those for Grade 5 students (Table 3), suggesting that the performance gap increases by grade level perhaps as the level of language complexity of assessments and instructional materials increases. To present a rough idea on performance gaps between ELL and non-ELL students before and after the passage of NCLB, we computed an average DI for Site 2 (pre-NCLB) and Site 5 (post-NCLB). The average DI for Site 2 was 30.3 (math computation was not included in the calculation of DI for this site) as compared with an average DI of 32.5 for Site 5. These analyses were replicated on the data from Sites 6 and 7 with the results being very consistent with those reported based on data from Site 5. This high level of consistency across multiple data sites over the nation suggests that the trend of performance gaps

between ELL and non-ELL students remains the same for many years and, in spite of the NCLB's efforts to make assessments more accessible to ELL students, these performance gaps remain a main concern. This is important since it suggests that something should seriously be done if ELL students are not to be left behind.

As indicated above, the results of analyses from other data sites in the nation were consistent with what was presented above and suggest that the trend of assessment outcome for ELL students is substantially lower than the trend for non-ELL students. Furthermore, the results of analyses show that the performance gap between ELL and non-ELL students can be explained mainly by the confounding of language with assessment outcomes. This gap increases as the level of language demand of items increases.

### **How Can the Validity of Assessment for ELL Students be Improved?**

Data presented above clearly suggest that ELL students perform far behind their non-ELL peers. However, there is no evidence to suggest that these students have less ability to learn content knowledge than non-ELL students. Therefore, nuisance variables such as linguistic and cultural biases may mainly be responsible for such performance gaps (Abedi, 2006a). To control for such nuisance variables and provide reliable and valid assessments for ELL students, various accommodations have been proposed and used in ELL assessments (Abedi, Kim-Boscardin, & Larson, 2000; Rivera, Stansfield, Scialdone, & Sharkey, 2000). However, research shows that some of these accommodations may not be relevant or may not even be effective for ELL students; thus, they may not produce valid assessment outcomes.

Based on Rivera's (2003) work, the most commonly used accommodations for ELL students are: extended time (42 of the 48 states examined), use of glossary (26 states); use

## LEP Partnership

of an English dictionary (33 states); use of a bilingual dictionary (22 states); and linguistically-simplified test items (12 states). Rivera's list of the most commonly used accommodations for ELL students included 73 accommodations. However, many of these accommodations do not directly address ELL students' needs. Abedi (2006b) evaluated the relevance of these 73 most commonly used accommodations for ELL students and found only 11 (15%) of them to be relevant for ELL students (the other accommodations were designed for students with cognitive or physical disabilities). These 11 accommodations provided direct language assistance to ELL students. They include English and bilingual dictionaries and glossaries, customized English and bilingual dictionaries, native language testing, and linguistically modified versions of the assessments.

Unfortunately, even some of the language-based accommodations that seem to be relevant for ELL students may not produce desirable outcomes. For example, studies have shown that some of these accommodations may alter the construct being measured and may thus render invalid results. A commercially published English dictionary, for example, may provide unfair advantage to the recipients since it may help with the correct response of test items (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Abedi, Courtney, & Leon, 2003; Abedi, Lord, Kim-Boscardin, & Miyoshi, 2001); therefore, the results of accommodated and non-accommodated assessment may not be aggregated because they represent different scales. Research also shows that translation of assessment tools into students' native language may not produce desirable results and may even provide invalid assessment results if the language of instruction and language of assessment are not aligned (Abedi, Lord, Hofstetter, & Baker, 2000.) Other studies have shown that using a glossary plus extra time as a form of accommodation has raised the

performance of both ELL and non-ELL students. The level of increase due to such accommodation strategies was higher for non-ELL students. This raised concern regarding the validity of accommodations (Abedi, Hofstetter, Lord, & Baker, 1998, 2000) since it suggests that content information was increased through these accommodations for both groups.

Studies by CRESST researchers along with other studies nationwide suggested that the linguistically modified version of the test's items is an effective and valid accommodation for ELL students. In fact, among many accommodations used in several experimentally controlled studies, the linguistic modification accommodation, which reduces unnecessary linguistic burden on students, was the only accommodation that reduced the performance gap between ELL and non-ELL students without compromising the validity of assessments (Abedi, Hofstetter, & Lord, 2004; Abedi, Lord, & Hofstetter, 1998; Abedi, Hofstetter, Lord, & Baker, 2000).

The purpose of this project is to inform states about the concept and principles of linguistic modification of assessments as a viable alternative to many commonly used accommodations for ELL students and to provide practical guidelines on how to incorporate this concept into the state assessment system without jeopardizing the integrity and validity of assessments for ELL students. This project consists of two different parts. Part I focuses on the theory and conceptual framework of *linguistic modification* and provides research-based evidence on the effectiveness and validity of language modification as a form of accommodation for ELL students for reducing the performance gap between ELL and non-ELL students. In Part II, practical guidelines will be provided

for developing linguistically modified assessments. This report (Part I) focuses on the theory and principles of linguistic modification of assessments.

### **What is Linguistic Modification/Language Simplification?**

The purpose of linguistic modification—sometimes referred to as language simplification—is to reduce or eliminate unnecessary linguistic complexity that confounds the content of assessment. This concept applies to content assessments in which content other than language is being assessed (e.g., math or science) and a language construct may be irrelevant to the purpose of assessment. Therefore, the main principle underlying language simplification is to reduce unnecessary linguistic complexity that is unrelated to the content of the test items. However, we prefer the term “*linguistic modification*,” as opposed to “*simplification*” which connotes a “dumbing down” of the assessment and is not the purpose of linguistic modification. There is a major distinction between the term “*modification*” used in “*linguistic modification*” as opposed to the term “*modification*” used by states to describe accommodations that may possibly change the construct being measured. In *linguistic modification* we believe (based on our research findings) that linguistic modification does not alter the construct being measured since it modifies only the unnecessary linguistic complexity that is irrelevant to the construct being measured. To present a comprehensive view of the linguistic modification approach, we first briefly introduce this concept and then present a summary of research on the effectiveness and validity of this approach in the assessment of ELL students.

### **Linguistic Modification Approach: A Brief Overview**

Assessments that are linguistically modified may facilitate students’ negotiation of language barriers. Linguistic modification of test items involves modifying the language of

a text while keeping the content intact. This may be accomplished by shortening sentences, removing unnecessary expository material, using familiar or frequently used words, using grammar that is more easily understood—including using present tense—and using concrete rather than abstract presentations.

While standardized achievement tests attempt to measure students' knowledge of specific content areas, analyses of mathematics and science subsections of the Test of Achievement and Proficiency (TAP) by Imbens-Bailey and Castellon-Wellington (Sept., 1999) show that two-thirds of the items include general vocabulary considered uncommon or used in an atypical manner. One-third of the items include syntactic structures that are evaluated as complex or unusual in their construction. In order to accurately assess knowledge within content areas, students must comprehend what the items are asking and understand the response choices.

Research has identified several linguistic features that appear to contribute to the difficulty of comprehending text. These features may slow down the reader, make misinterpretation more likely, or add to the reader's cognitive load and thus interfere with concurrent tasks. Indexes of language difficulty include unfamiliar vocabulary, complicated grammatical structures, and styles of discourse that include extra material, abstractions and passive voice (Abedi, Lord & Plummer, 1997). Some of these features<sup>1</sup> are discussed in detail below (see Appendix B for a more thorough discussion of linguistic features that affect performance outcomes of ELL students along with the citations to relevant research).

---

<sup>1</sup> See also the description of Language Demands (pp. 62-63) and Tables 1a and 1b in Part II: *A Guide to Linguistic Modification: Increasing English Language Learner Access to Academic Content* (Sato, 2007).

## VOCABULARY

### Unfamiliar Words/Idioms

Some words, word pairs, or groups of words still unfamiliar to English language learners might be used in a test item. They are unnecessary if they are not essential to the concept being tested. Idioms are words, phrases, or sentences that cannot be understood literally. Many proverbs, slang phrases, phrasal verbs and common sayings cannot be decoded by English language learners because they are not literal. Words that sound alike but have different meanings (there, their and they're) can also impede ELLs' understanding of text as can hyphenated words that may take on new or complex meaning when joined. On the other hand, words that are high on a general frequency list for English are likely to be familiar to most readers because they are often encountered.

#### *Unfamiliar words and pairs:*

According to the article, what role did some Navajo speakers play during World War II?

Circle the clumps of eggs in the illustration.

Patty expects that each tomato plant in her garden will bear 24 tomatoes. . .

In the last census, 80% of the households had one or more wage-earners.

#### *Idioms:*

In the story, who is in charge when the parents are gone?

James bought a used car from his uncle for \$5,000, but he put nothing down. If his loan is for 24 months, how much will his payments be?



## LEP Partnership

### False Cognates

False cognates are words that are similar in form or sound in two different languages but have two completely different meanings. A bilingual reviewer might recognize these:

In the story, how does Mack dote on his wife? (*la dote* in Spanish means dowry)

Examine the pie chart below to answer questions 4-8. (*pie* in Spanish (*foot*) is different than *pie* in English.)

### Overuse of Synonyms/Indefinite Pronouns/Missing or Unclear Antecedents

It is considered a feature of good writing to not use the same words over and over again. But a test writer's effort to infuse lexical variety into test items may result in confusion for English language learners. Synonyms like *table*, *chart*, and *matrix* should not describe the same thing on the same page. Using indefinite articles and substituting pronouns like *some* and *any*, are some of the potentially confusing ways we avoid repeating nouns. For example:

Marie had 10 marbles in a bag. **It** weighed 1 pound. Then she added **some more** until **it** weighed two pounds. . .

## GRAMMAR

### Long Phrases in Questions

Complex question types might have an opening phrase or clause that either replaces or postpones the question word.

At which of the following times should Ed feed the parking meter?

Of the following bar graphs, which represents the data?

According to the passage above, where do sea turtles lay their eggs?

### Complex Sentences

A complex sentence contains a main clause and one or more subordinating (dependent) clauses. Subordinating (dependent) words include *because, when, after, although, if, and since*. (More on *if* under Conditional Clauses.)

Because she wants to stay in touch, Peggy frequently \_\_\_\_\_.

When she came home, he \_\_\_\_\_ the letter.

Although the ship was \_\_\_\_\_, she was calm.

### Compound Sentences

A compound sentence consists of two or more clauses of equal importance. A coordinating conjunction (*and, or, but, so, for, yet, nor*) often connects the two clauses. Sometimes a conjunctive adverb (*however, therefore, moreover, nevertheless, as a result, accordingly, etc.*) combines the two ideas or begins a new sentence. When combined with other clauses or complex features, combining ideas in this way further complicates the language since it adds the consideration of relationship to comprehension.

Joe said that he ate more pizza than Ella, but Ella said they both ate the same amount.

We were hungry; however, there was no time to heat up the pizza.

Jim lost three marbles in the game and went home with 5 marbles. (note compound verb) Therefore, how many marbles did Jim start with?

When combined with other clauses or complex features, combining ideas in this way further complicates the language.

## LEP Partnership

### Logical Connectors: Conditional/Adverbial Clauses

Logical connectors are adverbial expressions that allow a listener/reader to infer connections between two structures. They mainly include dependent words (subordinating conjunctions—see above). In mathematics they often include conditional ‘if-then’ statements. Some take the form of complex sentences.

#### *Adverbial clauses:*

When the barber was finished with the haircut, he took the customer’s money.

While he was listening to music, he did his homework.

#### *Conditional Clauses:*

As long as you bring your own bedding, you can stay with us.

Given that  $a$  is a positive number, what is  $-a$ ?

If one pint will fill 2 cups, how many cups can be filled from 8 pints?

(vs. One pint will fill 2 cups. Eight pints will fill \_\_\_\_\_ cups.)

In Jean’s class there are twice as many boys as girls. If there are 10 girls in the class, how many boys and girls are there in the class?

### Unfamiliar Tenses

These include perfect tenses and modal auxiliaries; they are among the most difficult structures English language learners must interpret.

Perfect tenses use a helping verb and a past participle: *had gone, will have gone*:

She had been gone three hours when he arrived. (Past perfect progressive)

Modals are used to interject perspective or more subjectivity:

To give a proposition a degree of probability or express one’s attitude:

## LEP Partnership

He may pass the test.

The flashing light could mean that the batteries are low.

To convey the idea of necessity:

She ought to pass the test.

Modal:

What length other than the original estimate could have been the actual length of this dinosaur?

Perfect Tense plus Modal:

If Shirley had won three more marbles, could she have beaten Rodney in the game?

## Long Noun Phrases

Nouns sometimes work together to form one concept, such as a *pie chart* or *bar graph*.

Sometimes adjectives and nouns work together to create meaning: *high school diploma*, *income tax return*. To further complicate interpretation, strings of adjectives and nouns create subjects and objects: *freshwater pond*, *long-term investment*, *new word processing program*.

A loaded trailer truck weighs 26,643 kilograms. When the trailer truck is . . .

Of the following number pairs, which is the dimension of a 100-square-foot room?

To become next year's tennis team captain, how many votes will Sandra need?

## Relative Clauses

A relative clause is an embedded clause that provides additional information about the subject or object it follows. Words that lead a relative clause include *that*, *who*, and *which*.

Note: Often *that* is omitted from a relative clause.

## LEP Partnership

A bag that contains 25 marbles . . . (vs. One bag has 25 marbles. A second . . .)

Joe found the student who had loaned him the book.

When possible, relative clauses should be removed or recast.

### Prepositional Phrases

Prepositional phrases work as adjectives or adverbs to modify nouns, pronouns, verbs, adverbs, or adjectives. When they occur before question words, between the subject and the verb, or in strings, they can be especially confusing to English language learners.

Which of the following is the best approximation of the area of the shaded rectangle in the figure above if the shaded square represents one unit of area?

### Comparative Construction

Comparisons are made using *greater than*, *less than*, *n times as much as*, *as...as*—as well as by using certain verbs.

Jesse saw more mountains than he'd ever seen.

Who has more marbles than Carlos? Who has the most?

From which bag is he more likely to pull out a green marble?

If Bill runs 100 yards per hour faster than Peter . . .

Certain verbs imply comparison:

Joan underbid her hand.

Compared to Keith, Jen is short.

Note reduced clauses that can cause confusion:

John is taller than Mary. (than Mary is.)

The flour doesn't cost as much as the sugar. (as the sugar does.)

Mr. Jones' account is greater than that of Mr. Johnson. (than the account of . . .)

Even though structures such as these are useful and widespread in mathematical discourse, studies have shown that comparative structures are difficult for students to comprehend.

### **STYLE OF DISCOURSE**

#### Long Problem Statements/Unnecessary Expository Material

When the problem context set-up is long, students do not perform as well. Limit the story line to the essentials; however, do not reduce math items to bare computation.

#### Abstract (vs. Concrete) Presentation of Problem

Respondents show better performance when survey questions are presented in concrete rather than abstract terms. Information presented in narrative structures tends to be understood and remembered better than information presented in expository text.

The weights of two objects were measured. *vs.* The clerk weighed two suitcases.

#### Passive Voice

In active voice, the subject is the one performing an action. In passive voice, the one receiving the action is in the subject position. Often the “actor” is not stated.

He was given a ticket. *vs.* The officer gave him a ticket.

Girls’ ears were pierced in infancy. *vs.* Parents pierced infant girls’ ears.

When comparisons were made, the amounts in each jar had been reduced.

#### Complex Arrangement of Parts of Speech

At times the traditional subject-verb-object word order of English may be altered for expressing focus and emphasis. Passive voice is only one example.

The window was broken. (object-verb)

A Jedi knight you will be. (object-subject-verb)

John I can understand; the others just confuse me. (object-subject-verb)

## LEP Partnership

Ambitious she must have been, or she wouldn't have come. (adj-subject-verb)

### Negation

Several types of negative forms are confusing to English language learners:

Proper double negative:

Not all the workers at the factory are not male.

It's not true that all the workers at the factory are not male.

Negative question:

Which student will not finish in time?

Negative terms:

Ted can no longer drive over 40 mph in his truck.

### **PROBLEMS SPECIFIC TO MATH**

#### Confusing Order of Math Operations

The way a word question is composed might contribute to a formula being written incorrectly. Revise, unless the point of the item is to see if students can create a formula from a challenging word problem.

There are four times as many English teachers as math teachers in the school.

(common wrong answer:  $4E = M$ ; right answer:  $E = 4M$ )

If the unknown in the problem is clearly stated near the beginning of a problem statement, there is less likelihood of error. Some problems do not clearly denote the unknown:

In 7 years, John will be 18 years old.

Ordinary English Words Used as Math Words

Some words have a common meaning and a different meaning when used in math. Studies have shown that people are more likely to think about the more common interpretation first. Some examples are: *product, average, mean, table, column, odd, prove.*

Problems with Mathematical “Keywords”

Depending on their exposure to verbal and written cues for math operations, English language learners may not know all of the vocabulary associated with the four operations:

Addition: *plus, are added, is increased by, the sum of, more than, exceeds, altogether, gained, total*

Subtraction: *minus, subtract, subtracted from, decreased by, diminished by, reduced by, the difference between, less than, how much less, lost, left, how much larger than, how much smaller than, how much greater than, how much further than . . .*

Multiplication: *times, the product of, multiplied by*

Division: *divided by, the quotient of, one half the sum, the fractional part of (e.g., three-fourths of), average . . .*

Other challenges:

*Altogether* is sometimes interpreted by students as *each*:

Lisa and Diane have 5 yards of blue fabric altogether. How much fabric may . . .

*At the start* is sometimes interpreted as being the first number mentioned:

Dwayne finished the game with 10 marbles. He had lost a total of 5 marbles.  
How many marbles did Dwayne have **at the start**?



## LEP Partnership

In studies, students usually associated *more* with addition and *less* with subtraction. Note the following:

On Monday, the milkman brought 7 bottles of milk. That was 4 bottles **less than** he brought on Sunday. How many bottles did he bring on Sunday?

Ordinal numbers can confuse students:

Six times a number is equal to a second number.

Some students misunderstand the ordinal number *second* as the cardinal number *two* and write the equation  $6x=2$ .

### **Research on the Effectiveness and Validity of Linguistic Modification as a Tool for Improving Validity of Assessment for ELL Students**

Providing equitable access to special needs student populations in large-scale assessments is a major step toward fair assessment for all. Well-designed assessments are better measures for all students, including students with disabilities and English language learners. Linguistic modification as a form of accommodation makes the assessment more accessible to larger populations of students including English language learners.

As elaborated earlier, language factors affect performance outcomes, especially for English language learners. Students' content knowledge in areas such as mathematics, science, or social studies may not be truly assessed if students cannot understand the language of the test. Previous research shows that minor changes in the wording of content-related test items can raise student performance (Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Lord, & Plummer, 1997; Cummins, Kintsch, Reusser, & Weimer, 1988; De Corte, Verschaffel, & DeWin, 1985; Durán, 1989; Hudson, 1983; Riley, Greeno, & Heller, 1983). For example, rewording a verbal problem can make semantic

relations more explicit, without affecting the underlying semantic and content structure; thus the reader is more likely to construct a proper problem representation and to solve the problem correctly.

Findings from the analyses of many national and state studies clearly show the impact of language on the assessment outcomes for ELL students. For example, the results of analyses of NAEP extant data (Abedi, Lord, & Plummer, 1997) suggested that ELL students had difficulty with the test items that were linguistically complex. The study also found that ELL students exhibited a substantially higher number of omitted/not-reached test items since it took them a much longer time to read and understand assessment questions. Based on the findings of these studies, the linguistic modification approach was applied. In one study, researchers identified 48 linguistic features that may affect ELL students' performance and grouped them into 14 general categories (Abedi, Lord, & Plummer, 1997). The impact of these linguistic features on the performance of ELL students in content-based areas (math and science) was then examined. A short description of each of these 14 categories along with research evidence of the impact of these features on assessment of ELL students is presented in Appendix B of this report.

The effects of some of these linguistic features on a sample of 1,031 eighth-grade students in Southern California were examined (Abedi & Lord, 2001). In this study, the math items for Grade 8 students were modified to reduce the complexity of sentence structures and to replace potentially unfamiliar vocabulary with more familiar words without changing the content-related terminologies (mathematical terms were not changed). The results showed significant improvements in the scores of ELL students and also non-ELLs in low- and average-level mathematics classes, but changes did not affect

scores of higher-performing non-ELL students. Among the linguistic features that appeared to contribute to the differences were low-frequency vocabulary and passive voice verb constructions. These features contributed to the linguistic complexity of the text and made the assessment more linguistically complex for ELL students.

The outcome of this study was cross-validated in another study in which Abedi, Lord, and Hofstetter (1998) examined the impact of linguistic modification on the mathematics performance of ELLs and non-ELLs from a sample of 1,394 eighth graders in schools with high enrollments of Spanish speakers. Results confirmed findings of the earlier studies and showed that modification of the language of items contributed to improved performance on 49% of the items; the ELL students generally scored higher on shorter/less linguistically complex problem statements. The results of this study also suggest that lower-performing native speakers of English also benefited from the linguistic modification of the assessment.

Other studies were conducted to obtain cross-validation evidence on the contribution of the language modification approach in improving the validity of assessments for ELL students. A study (Abedi, Lord, Hofstetter, & Baker, 2000) on a sample of 946 eighth graders found that among four different accommodation strategies for ELL students, only the linguistically modified English form narrowed the score gap between English learners and other students.

The effectiveness of the language modification approach in reducing the performance gap between ELL and non-ELL students was the topic of another study (Abedi, Courtney, & Leon, 2003). This study examined 1,594 eighth-grade students using items from the National Assessment of Educational Progress (NAEP) and the Third

## LEP Partnership

International Math and Science Study (TIMSS). Students were given either a customized English dictionary (words were selected directly from test items), a bilingual glossary, a linguistically modified test version, or the standard test items. Only the linguistically modified version improved the ELL students' scores without affecting the non-ELL students' scores.

The results of the studies presented above are consistent with the findings of other studies and suggest that linguistic modification of assessment items provides a more valid and effective alternative to the conventional testing approach. Maihoff (2002) found linguistic simplification of content-based test items to be a valid and effective accommodation for ELL students. Kiplinger, Haug, and Abedi (2000) found linguistic modification of math items helped improve the performance of ELL students in math without affecting performance of non-ELL students. Rivera and Stansfield (2001) compared ELL performance on regular and simplified fourth- and sixth-grade science items. Although the small sample size in the Rivera and Stansfield study did not show significant differences in scores, the study did demonstrate that linguistic simplification did not affect the scores of English-proficient students, indicating that linguistic simplification is not a threat to score comparability.

However, some researchers and policymakers have expressed concern over the concept and application of language simplification of text used in the assessment and instruction of ELL students. For example, they argue that to be successful academically, ELL students must be proficient in academic language, which is not necessarily the same as conversational fluency. Proficiency in academic language includes the knowledge of less frequent vocabulary and the ability to interpret and produce complex written language.

## LEP Partnership

Students should be able to understand complex linguistic structure in content areas such as science, social sciences and mathematics. According to these researchers, reducing the complexity of language that is required to perform such complex tasks may not be productive.

Francis, Lesaux, Kieffer, and Rivera (2006) reviewed research on accommodations for English language learners, including studies on the validity and effectiveness of linguistic modification/language simplification of assessment as a form of accommodation for ELL students. By reviewing literature and comparing effect sizes of the linguistically modified version of the tests reported in the peer-reviewed papers, Francis et al. indicated that the findings supporting the effectiveness of simplified English were weak. The authors also added that “While it is possible that the effects of simplified English vary according to variables such as grade level, content area, and the nature of the assessment, the evidence does not currently support this conclusion (p 26).”

Findings from Francis et al’s reviews raise serious concern regarding methodological issues of the studies examining the impact of language simplification on the assessments for ELLs. There are several issues in the design of some of the studies that were reviewed. Among these issues are: (1) a small (and non-representative) sample for ELL students in the study; (2) lack of control of extraneous variables that could affect performance of ELL students such as their level of English proficiency; and (3) lack of operational definition of the linguistic simplification approach. For example, in some of these studies the number of ELL students was very small and did not provide enough subjects to assess the impact of linguistic modification. More importantly, the different studies applied the concept of linguistic modification of assessment items quite differently. That is, there was no uniform

approach to linguistic modification across the studies. As indicated earlier, Abedi, Lord and Plummer (1997) introduced the concept of linguistic modification and identified several linguistic features as indications of linguistic complexity in the assessments. Others used other linguistic features or even a simple editorial process in simplifying the language of assessments for ELL students.

In general, the research evidence shows linguistic complexity as a major source of measurement error in the assessment results for ELL students. Research findings also suggest that reducing the level of unnecessary linguistic complexity of assessments may help improve assessment validity and reliability for these students. Improvement of the assessment validity and reliability can be linked to reducing unnecessary linguistic complexity of assessment. On the other hand, some people argue that reducing the complexity level of academic contents may change the construct being taught and being assessed. However, the complex linguistic structures that are related to the content of assessment and instruction must be distinguished from the unnecessary linguistic complexity of the text in both assessment and instruction.

### **Determining the Comparability of Scores from Linguistically Modified Assessments**

The principle underlying the concept of linguistic modification is based on the premise that the unnecessary linguistic complexity of content-based assessment (e.g., math and science) is considered as a construct-irrelevant source of variance, which undermines assessment validity for ELL students. That is, scores for many ELL students underrepresent these students' knowledge and skills. As elaborated in the studies investigating linguistic modification (see for example, Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997) tests with modified language are usually constructed with the advice of

content experts to avoid changing the construct the items are intended to address, which is a first step in maintaining comparability between the modified language test and the source (general) test. In the studies cited above, ELL students' scores increase when unnecessary linguistic complexity is reduced. The study results suggest that ELL students' scores from linguistically modified assessments are, as compared to general assessments, more on par with English-proficient students' scores in terms of representing knowledge and skills (Abedi, Lord, & Plummer, 1997; Abedi, Leon & Mirocha, 2003).

Several studies have been conducted to investigate the validity (comparability) of the inferences made from the linguistically modified test scores with the scores from the original tests. In these studies, the validity of linguistically modified assessments has been thoroughly examined. These studies are based on the assumption that the linguistically modified assessment is considered valid if it does not alter the construct being measured. Thus, any indication of a significant change in the performance of non-ELL students taking the modified version of the assessment may jeopardize the validity of this approach.

For examining the validity (comparability) of linguistically modified assessments, in many studies, both ELL and non-ELL students were tested under both original and linguistically modified assessments. The results of these studies consistently indicated that the performance of non-ELL students was not affected by the linguistic modification of assessments (see for example, Abedi & Lord, 2001; Abedi, Courtney, & Leon, 2003; Abedi, Lord, & Hofstetter, 1998; Abedi, Lord, & Plummer, 1997; Abedi, Lord, Hofstetter, & Baker, 2000; Maihoff, 2002; Kiplinger, Haug & Abedi, 2000; Rivera & Stansfield, 2001). Findings of these studies that have been cross-validated clearly support the notion of comparability of scores from the linguistically modified test scores and those from the

original tests.

A linguistically modified version of assessments could lend valid results if it does not change the construct being measured. To examine the validity of assessment under linguistically modified tests, many different studies compared performance of non-ELL students using the linguistically modified version of the tests with non-ELL students who were tested under the standard testing condition using the original version of the tests. The main assumption here was that linguistic modification should not change the performance of non-ELL students for whom the language of assessment should not impact their performance outcome. Such comparisons did not yield any significant results, which suggests that the linguistic modification approach did not affect the performance of non-ELL students (see for example, Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Kiplinger, Haug, & Abedi, 2000; Maihoff, 2002; Hansen & Mislevy, 2004).

### **Impact of Linguistic Modification of Assessment on Psychometric Characteristics of Assessments for ELL Students**

By reducing the impact of language barriers on content-based assessments, the assessments' validity and reliability can be improved, resulting in fairer assessments for all students, including ELLs (see Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Kiplinger, Haug, & Abedi, 2000; Maihoff, 2002; Hansen & Mislevy, 2004). When math test items were modified to reduce the level of linguistic complexity, over 80% of middle-school students who were interviewed preferred the linguistically modified over the original English version of the test items (see Abedi et al., 1997).

Recent studies on the assessment of ELL students have demonstrated that the unnecessary linguistic complexity of content-based assessments (e.g., math and science) is



a likely source of measurement error, differentially impacting the reliability of assessment for the ELL subgroup. The linguistic complexity of test items as a source of construct-irrelevant variance may also influence the construct validity of the assessment for these students (Abedi, 2006a). Results of analyses of existing data from several locations nationwide show a substantial gap in reliability (internal consistency) and validity (concurrent validity) between ELL and non-ELL students on test items that are linguistically complex (Abedi, 2006a; Abedi, Leon, & Mirocha, 2003).

To illustrate the impact of linguistic modification on the reliability and validity of assessments for ELL students, we analyzed data from one of the data sites mentioned earlier in this report (Site 2). Table 5 presents the results of these analyses. As Table 5 shows, the gap in the reliability and validity coefficients reduces as the level of language demand of the assessment decreases. The reliability coefficients (alpha) for non-ELL students ranged from .898 for math to .805 for science and social science. For ELL students, however, alpha coefficients differed considerably across the content areas (see Table 5, the difference in alpha). In math, where language factors might not have much influence on performance, the alpha coefficient for ELL (.802) was only slightly lower than the alpha for non-ELL students (.898). In English language arts, science, and social science, however, the gap in alpha between English-only and ELL students was large. Averaging over English language arts, science, and social science results, the alpha for non-ELL was .808 as compared to an average alpha of .603 for ELL students. Thus, language factors introduce a source of measurement error affecting ELL students' test outcomes, while they may not have much impact on students who are native or fluent speakers of English (for a more detailed description, see Abedi, 2006a; Abedi, Leon, &

Mirocha, 2003). As the level of linguistic complexity in science and social science tests decreased, the gap in the reliability coefficient was reduced substantially.

As indicated earlier in this report, the validity of the assessments for ELL students was also affected by the unnecessary linguistic complexity as a source of construct-irrelevant variance. Results of a multiple group confirmatory factor analysis model showed lower construct validity for ELL students as compared with non-ELL students (see Abedi, Leon & Mirocha, 2003).

### **What Does the Research Suggest?**

In order to accurately assess knowledge within content areas, students must comprehend what the items are asking and understand the response choices. Research to date suggests that a productive teaching and assessment approach for helping students, particularly ELL students, must be examined for unnecessary linguistic complexities. What can we say with confidence at this time?

1. The performance gap between ELL students and other students on content area tests can be narrowed by modifying the language of the test items to reduce the use of low-frequency vocabulary and language structures that are incidental to the content knowledge being assessed. This strategy has been shown to be effective in reducing the performance gap between high- and low-performing students as well.
2. All students should have content area assessments that use clear language and provide sufficient time for them to demonstrate their knowledge.
3. The development of future instructional materials and large-scale content area assessments should consider ELL students from the outset rather than as an

afterthought. The use of clear language, free of unnecessary complexity, can and should be a part of good instructional planning and assessment practice.

4. The specific language demands of academic materials and assessment tools should be identified and provided to teachers so that they can ensure that students have the language resources to demonstrate their content-area knowledge and skills.

### **Discussion**

Research findings presented in this report showed substantial performance gaps between English language learners and their native English speaking peers. While there is no evidence to suggest any difference between ELL and non-ELL students on their ability to learn, these differences are alarming. Research-based evidence suggests that lower performance of ELL students is mainly due to the impact of language factors on the instruction and assessment of these students. It is extremely challenging for ELL students to be instructed and assessed in a language that they are not quite proficient in and are struggling to learn. To ameliorate this situation, federal legislation mandates the use of reliable and valid assessments using appropriate testing accommodations. The main issue, however, is identifying appropriate accommodations for ELL students. In this report we cited many accommodations that are used for ELL students, many of which may not even have face validity. Among the most commonly used accommodations, only those that could help ELL students directly with their language needs may be relevant to these students. Even among these language-related accommodations some of them may not produce valid assessment outcomes, as they may alter the construct under measurement.

In this report, we proposed linguistic modification as a form of accommodation that is relevant to the assessment of ELL students and at the same time does not alter the construct being measured, thereby providing valid assessment outcomes for ELL students. We presented research findings to support the use of this approach and presented a brief overview of this approach. This report provides research-based evidence to support the use of linguistic modification of assessment and help states to utilize this approach in the assessment of ELL students. We also discussed concerns expressed by some researchers on the possible negative impact of linguistic modification on the quality of assessments. In response to such concerns, the report provided evidence that the linguistic modification approach only targets unnecessary linguistic complexity of assessment, the language that is not directly related to the content being measured.

Findings of the studies cited in this paper clearly suggest that the performance of non-ELL students tested under linguistically modified assessments are not different than their performance under the original version. These findings provide assurances that linguistic modification of assessment that is used as an effective accommodation for ELL students does not compromise the validity of assessment for these students.

We hope this report provides useful information for states in their efforts to establish and administer valid and fair assessments for ELL students. The linguistic modification approach suggests developing assessments with language that clarifies the message of the construct of the assessment, rather than mystifying it. In that sense, everyone can benefit from assessments that are clear.

**References: Part I**

- Abedi, J. (2006a). Language Issues in Item-Development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 377-398). Mahwah, NJ: Erlbaum.
- Abedi, J. (2006b). *Accommodations for English language learners that may alter the construct being measured*. Paper (symposium) presented at the 2006 Annual Meeting of the American Educational Research Association in San Francisco, CA: April, 2006.
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234.
- Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Tech. Rep. No. 586). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodation for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification* (CSE Rep. No. 666 ). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C., & Lord, C. (1998). *Impact of selected background variables on students' NAEP math performance*:(CSE Rep. No. 478). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.
- Abedi, J., Hofstetter, C., Lord, C., & Baker, E. (1998). *NAEP math performance and test accommodations: Interactions with student language background*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Hofstetter, C., Lord, C., & Baker, E. (2000). *NAEP math performance and test accommodations: Interactions with student language background* (CSE Tech. Rep. No. 536). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Abedi, J., Kim-Boscardin, C., & Larson, H. (2000). *Summaries of research on the inclusion of students with disabilities and limited English proficient students in large-scale assessment*. Los Angeles: University of California, Center for the Study of

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

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data* (CSE Tech. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Tech. Rep. No. 478). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., & Plummer, J. (1997). *Final Report of language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No. 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16-26.

Abedi, J., Lord, C., Kim-Boscardin, C., & Miyoshi, J. (2001). *The effects of accommodations on the assessment of LEP students in NAEP* (CSE Tech. Rep. No. 537). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Baugh, J. (1988, August). [Review of the article Twice as less: Black English and the performance of black students in mathematics and science]. *Harvard Educational Review*, 58(3), 395-404.

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, 1(3), 79-132.

Botel, M., & Granowsky, A. (1974). A formula for measuring syntactic complexity: A directional effort. *Elementary English*, 1, 513-516.

Celce-Murcia, M., & Larsen-Freeman, D. (1983). *The grammar book: An ESL/EFL teacher's book*. Rowley, MA: Newbury House.

Chall, J. S., Jacobs, V. S., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20, 405-438.

## LEP Partnership

- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 27,11-20, 28, 37-54.
- De Corte, E., Verschaffel, L., & DeWin, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, 77(4), 460-470.
- Duran, R. P. (1989). Assessment and instruction of at-risk Hispanic students. *Exceptional Children*, 56(2), 154-158.
- Finegan, E. (1978, December). *The significance of syntactic arrangement for readability*. Paper presented to the Linguistic Society of America, Boston, MA.
- Forster, K. I., & Olbrei, I. (1973). Semantic heuristics and syntactic trial. *Cognition*, 2(3), 319-347.
- Francis, D. J., Lesaux N., Kieffer, M. Rivera, H. (2006). Practical guidelines for the education of English language learners. Research-based recommendations for the use of accommodations in large-scale assessments. Houston: Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston for the Center on Instruction.
- Freeman, G. G. (1978, June). *Interdisciplinary evaluation of children's primary language skills*. Paper presented at the World Congress on Future Special Education, First, Stirling, Scotland. (ERIC Document Reproduction Service No. ED157341).
- GAO (2006). No Child Left Behind Act. Assistance from education could help states better measure progress of students with limited English proficiency: Washington, DC: United States Government Accountability Office.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
- Haiman, J. (1985). *Natural syntax: Iconicity and erosion*. New York: Cambridge University Press.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara: University of California Linguistic Minority Research Institute.
- Halliday, M. A. K., & Martin, J. R. (1994) *Writing science: Literacy and discursive power*. Pittsburgh, PA: University of Pittsburgh Press.
- Hansen, E. G. & Mislavy, R. (2004, April). *Towards a unified validity framework for ensuring access to assessments by individuals with disabilities and English language learners*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

- Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development, 54*, 84-90.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels* (Research Rep. No. 3). Urbana, IL: National Council of Teachers of English.
- Hunt, K. W. (1977). Early blooming and late blooming syntactic structures. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging*. Urbana, IL: National Council of Teachers of English.
- Imbens-Bailey, A., & Castellon-Wellington, M. (1999, September). *Linguistic demands of test items used to assess ELL students*. Paper presented at the annual CRESST Conference, Los Angeles, CA.
- Improving America's Schools Act of 1994, Pub. L. No. 103-382, 108 Stat. 3518 (1994).
- Jones, P. L. (1982). Learning mathematics in a second language: A problem with more and less. *Educational Studies in Mathematics, 13*, 269-87.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixation to comprehension. *Psychological Review, 87*, 329-354.
- Kindler, A. L. (2002). *Survey of the states' limited English proficient students & available educational programs and services, 2000-2001 Summary Report*. Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580-602.
- Kiplinger, V. L., Haug, C. A., & Abedi, J. (2000, April). *Measuring math – not reading – on a math assessment: A language accommodations study of English language learners and other special populations*. Presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*, 62-102.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day English*. Providence, RI: Brown University Press.
- Larsen, S. C., Parker, R. M., & Trenholme, B. (1978). The effects of syntactic complexity upon arithmetic performance. *Educational Studies in Mathematics, 21*, 83-90.
- Lemke, J. L. (1986). *Using language in classrooms*. Victoria, Australia: Deakin University Press.
- MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*, 692-715.



## LEP Partnership

- MacGinitie, W. H., & Tretiak, R. (1971). Sentence depth measures as predictors of reading difficulty. *Reading Research Quarterly*, 6, 364-377.
- Maihoff, N. A. (2002, June). *Using Delaware data in making decisions regarding the education of LEP students*. Paper presented at the Council of Chief State School Officers 32nd Annual National Conference on Large-Scale Assessment, Palm Desert, CA.
- Mestre, J. P. (1988). The role of language comprehension in mathematics and problem solving. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 200-220). Hillsdale, NJ: Erlbaum.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Orr, E. W. (1987). *Twice as less: Black English and the performance of black students in mathematics and science*. New York: W. W. Norton.
- Pauley, A., & Syder, F. H. (1983). Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics*, 7, 551-579.
- Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153-196). New York: Academic Press.
- Rivera, C. (2003). *State assessment policies for English language learners*. Presented at the 2003 Large-Scale Assessment Conference.
- Rivera, C., & Stansfield, C. W. (2001, April). *The effects of linguistic simplification of science test items on performance of limited English proficient and monolingual English-speaking students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Rivera, C., Stansfield, C. W., Scialdone, L., & Sharkey, M. (2000). *An analysis of state policies for the inclusion and accommodation of English language learners in state assessment programs during 1998-1999*. Arlington, VA: The George Washington University, Center for Equity and Excellence in Education.
- Savin, H. B., & Perchonock, E. (1965). Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 4, 348-353.
- Schachter, P. (1983). *On syntactic categories*. Bloomington: Indiana University Linguistics Club.
- Shuard, H., & Rothery, A. (Eds.). (1984). *Children reading mathematics*. London: J. Murray.
- Slobin, D. I. (1968). Recall of full and truncated passive sentences in connected discourse. *Journal of Verbal Learning and Verbal Behavior*, 7, 876-881.

## LEP Partnership

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher*, 32(2), 3–13.

Spanos, G., Rhodes, N. C., Dale, T. C., & Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking & J. P. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 221-240). Hillsdale, NJ: Erlbaum.

U.S Census Bureau, 2002

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 9, 398-404.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

### Appendix A. Tables

Table 1  
Mean, Standard Deviation and Number of Students for ITBS Subsection Scores at the Different Grade/Level Combinations (NCE Scores)

Grade	ELL Status	Math Concepts & Estimation	Math Prob. Solv. & Data Interp.	Math Computation	Reading
3	Non-ELL				
	Mean	44.14	40.52	50.21	37.92
	N	28,733	28,694	28,740	28,745
	SD	20.08	21.49	23.89	17.93
	ELL				
	Mean	41.89	36.47	51.84	30.72
	N	7,248	7,254	7,260	7,261
	SD	19.14	20.57	23.27	17.10
4	Non-ELL				
	Mean	44.12	45.47	56.08	45.44
	N	24,908	24,904	24,915	24,910
	SD	20.41	17.77	24.13	15.70
	ELL				
	Mean	34.84	38.31	52.58	34.85
	N	5,226	5,220	5,225	5,221
	SD	18.81	15.67	23.90	12.77
6	Non-ELL				
	Mean	45.20	43.94	50.82	42.66
	N	24,935	24,915	24,924	24,942
	SD	20.53	18.57	21.02	16.14
	ELL				
	Mean	35.41	33.69	45.60	29.73
	N	3,338	3,335	3,337	3,330
	SD	17.57	14.30	18.47	12.50
8	Non-ELL				
	Mean	48.36	47.50	49.13	46.59
	N	23,036	23,033	23,039	23,071
	SD	19.31	15.97	16.39	15.19
	ELL				
	Mean	37.08	35.94	43.52	32.69
	N	2,300	2,300	2,303	2,291
	SD	16.07	13.59	14.77	12.52

Table 2

Disparity Indices of Non-ELL over ELL Students on Reading and Math Subsections

Primary Grade	Math Concepts & Estimation	Math Problem Solving & Data Interpretation	Math Computation	Reading
3	5.3	11.1	-3.1	23.4
4	26.9	19.3	6.9	30.1
5	36.5	32.7	12.6	41.1
6	27.5	30.9	11.8	43.7
7	39.4	32.7	12.9	39.6
8	30.5	31.7	12.9	42.7
Average	27.7	26.4	9.0	36.8

LEP Partnership

Table 3

Mean, standard deviation and number of students for math, science and reading tests by students' ELL status for Grade 5

Content	ELL students			Non-ELL students			DI
	M	SD	N	M	SD	N	
Math	252.55	61.57	4973	304.93	74.04	27938	21
Science	226.44	66.46	4958	310.64	68.91	27903	37
Reading	215.41	61.45	4969	288.53	66.72	27949	34
Total	231.57	55.72	4953	301.50	63.82	27877	30

Overall Total: M=290.9518; SD=67.48482; N=32830

Table 4

Mean, standard deviation and number of students for math, science and reading tests by students' ELL status for Grade 8

Content	ELL students			Non-ELL students			DI
	M	SD	N	M	SD	N	
Math	220.81	85.05	3793	305.43	95.23	28811	38
Science	230.47	75.22	3765	314.72	82.79	28669	37
Reading	236.80	65.98	3795	305.41	69.13	28817	29
Total	229.78	66.443	3754	309.06	75.489	28569	34

Overall Total: M=290.9518; SD=67.48482; N=32830

LEP Partnership

Table 5  
Site 2 Stanford 9 Sub-scale Reliabilities (Alpha), Grade 9

Sub-scale (Items)	Non-ELL	LEP	Difference in alpha
Reading, N=	181,202	52,720	
-Vocabulary (30)	.835	.666	0.169
-Reading Comp (54)	.916	.833	0.083
Average Reliability	.876	.750	0.126
Math, N=	183,262	54,815	
-Total (48)	.898	.802	0.096
Language, N=	180,743	52,863	
-Mechanics (24)	.803	.686	0.117
-Expression (24)	.812	.680	0.132
Average Reliability	.813	.683	0.130
Science, N=	144,821	40,255	
-Total (40)	.805	.597	0.208
Social Science, N=	181,078	53,925	
-Total (40)	.805	.530	0.275

## Appendix B

### Linguistic Features That May Affect Comprehension<sup>2</sup>

Below is a summary of linguistic features that may affect comprehension, from Abedi, Lord, and Plummer (1997). These features slow down the reader, make misinterpretation more likely, and add to the reader's cognitive load, thus interfering with concurrent tasks.

Before discussing each of these individual linguistic features, it must be noted that this list is by no means exhaustive of the linguistic features that could affect students' performance. Researchers, linguists, and measurement experts may add other pertinent features. The purpose of this list is not to provide a comprehensive linguistic analysis of content-based test items. Rather, it is to introduce the concept of linguistic complexity and to discuss methodology to control for such threats to the reliability and validity of assessment tools for all students, particularly for those with greater language needs.

It must also be noted that there will naturally be circumstances where the grammar in an item is too challenging to modify. In such cases, we can only remind ourselves to not create items in such a manner for the future, and try our best to modify other parts of the item. The more we can linguistically modify items to make them easier to understand the greater the chances of adequately gauging an English language learner's content knowledge. Later we will provide examples of linguistic modification that specifically address these language concerns.

1. Word frequency/familiarity. Word frequency was an element in early formulas for readability (Dale & Chall, 1948; Klare, 1974). Words that are high on a general frequency list for English are likely to be familiar to most readers because they are encountered often. Readers who encounter a familiar word will be likely to interpret it quickly and correctly, spending less cognitive energy analyzing its phonological component (Adams, 1990; Chall, Jacobs, & Baldwin, 1990; Gathercole & Baddeley, 1993). On a test with math items of equivalent mathematical difficulty, eighth-grade students scored higher on the versions of items with vocabulary that was more frequent

---

<sup>2</sup> See also the description of Language Demands (pp. 62-63) and Tables 1a and 1b in Part II: *A Guide to Linguistic Modification: Increasing English Language Learner Access to Academic Content* (Sato, 2007).

and familiar; the difference in score was particularly notable for students in low-level math classes (Abedi et al., 1997).

2. Word length. As frequency of occurrence decreases, words tend to be longer. Accordingly, word length can serve as an index of word familiarity (Zipf, 1949; Kucera & Francis, 1967). Additionally, longer words are more likely to be morphologically complex. In one study, language minority students performed better on math test items with shorter word lengths than items with longer word lengths (Abedi et al., 1997).
3. Sentence length. Sentence length serves as an index for syntactic complexity and can be used to predict comprehension difficulty; linguistic definitions of complexity are based on the assumption that word depth correlates with sentence length (Bormuth, 1966; MacGinitie & Tretiak, 1971; Wang, 1970).
4. Voice of verb phrase. People find passive voice constructions more difficult to process than active constructions (Forster & Olbrei, 1973), and more difficult to remember (Savin & Perchonock, 1965; Slobin, 1968). Furthermore, passive constructions can pose a particular challenge for non-native speakers of English (Celce-Murcia & Larsen-Freeman, 1983). Passive voice constructions tend to be used less frequently in conversation than in formal writing such as scientific writing (Celce-Murcia & Larsen-Freeman, 1983). In one study, eighth-grade students (native and non-native English speakers) were given equivalent math items with and without passive voice constructions; students in average math classes scored higher in the versions without passive constructions (Abedi et al., 1997).
5. Length of nominals. Noun phrases with several modifiers have been identified as potential sources of difficulty in test items (Spanos, Rhodes, Dale, & Crandall, 1988). Long nominal compounds typically contain more semantic elements and are inherently syntactically ambiguous; accordingly, a reader's comprehension of a text may be impaired or delayed by problems in interpreting them (Halliday & Martin, 1994; Just & Carpenter, 1980; King & Just, 1991; MacDonald, 1993). Romance languages such as Spanish, French, Italian, and Portuguese make less use of compounding than English does, and when they do employ such a device, the rules are different. Consequently, students whose first language is a Romance language may have difficulty interpreting compound nominals in English (Celce-Murcia & Larsen-Freeman, 1983).



6. Complex question phrases. Longer question phrases occur with lower frequency than short question phrases, and low-frequency expressions are in general harder to read and understand (Adams, 1990).
7. Comparative structures. Comparative constructions have been identified as potential sources of difficulty for non-native speakers (Jones, 1982; Spanos et al., 1988) and for speakers of non-mainstream dialects (Orr, 1987, but see also Baugh, 1988).
8. Prepositional phrases. Students may find interpretation of prepositions difficult (Orr, 1987; Spanos et al., 1988). Languages such as English and Spanish may differ in the ways that motion concepts are encoded using verbs and prepositions (Slobin, 1968).
9. Sentence and discourse structure. Two sentences may have the same number of words, but one may be more difficult than the other because of the syntactic structure or discourse relationships among sentences (Freeman, 1978; Finegan, 1978; Larsen, Parker, & Trenholme, 1978).
10. Subordinate clauses. Subordinate clauses may contribute more to complexity than coordinate clauses (Botel & Granowsky, 1974; Hunt, 1965, 1977; Wang, 1970).
11. Conditional clauses. Conditional clauses and initial adverbial clauses have been identified as contributing to difficulty (Spanos et al., 1988; Shuard & Rothery, 1984). The semantics of the various types of conditional clauses in English are subtle and hard to understand even for native speakers (Celce-Murcia & Larsen-Freeman, 1983). Non-native speakers may omit function words (such as “*if*”) and may employ separate clauses without function words. Separate sentences, rather than subordinate “*if*” clauses, may be easier for some students to understand (Spanos et al., 1988). In fact, some languages do not allow sentences with the conditional clause in sentence-final position (Haiman, 1985). Consequently, this positioning may cause difficulty for some non-native speakers.
12. Relative clauses. Since relative clauses are less frequent in spoken English than in written English, some students may have had limited exposure to them. In fact, Pauley and Snyder (1983) argue that the relative clauses in literature differ from those in spoken vernacular language (Schachter, 1983).
13. Concrete vs. abstract or impersonal presentations. Studies show better performance when problem statements are presented in concrete rather than abstract terms

(Cummins et al., 1988). Information presented in narrative structures tends to be understood and remembered better than information presented in expository text (Lemke, 1986).

14. Negation. Mestre (1988) observed that a considerable number of research studies indicate that sentences containing negations (e.g., no, not, none, never) are harder to comprehend than affirmative sentences. One of the reasons for its complexity may be because there is a lack of parallelism in the use of negation between English and other languages. In Spanish, for example, double negative constructions retain a negative meaning instead of reverting to an affirmative meaning, as would be the case in grammatically correct English. Mestre found that Spanish-speaking students processed negations from left to right, which works for natural discourse but does not always work for mathematics texts.

**Part II: A Guide to Linguistic Modification:**

**Increasing English Language Learner Access to Academic Content**

Edynn Sato, Ph.D.

Assessment and Accountability Comprehensive Center and WestEd

## Contents

### Part II: A Guide to Linguistic Modification:

#### Increasing English Language Learner Access to Academic Content

Introduction	53
Linguistic Modification	54
Critical Considerations for Linguistically Modified Assessments	55
Consideration of the English Language Learner Population	56
Consideration of Access and Implications for Validity	57
Developing a Linguistically Modified Assessment: Increasing Access for ELLs	59
Defining the Assessed Domain and Constructs	60
Distinguishing Construct-Relevant from Construct-Irrelevant Language	62
Defining the English Language Learner Population	63
Specifying Linguistic Modification Strategies—Current Research and Recent Refinements	66
Current Research: General Guidelines for Linguistic Modification	66
Recent Research: Implications for Refining Linguistic Modification Strategies	68
Evaluating the Effectiveness of Linguistic Modification	70
Examining the Effect of Linguistic Modification	70
Content and Bias Reviews: Consideration of the ELL Population	72
User’s Guide: An Approach to Developing a Linguistically Modified Assessment	74
References: Part II	81
Tables	89

**Part II: A Guide to Linguistic Modification:**  
**Increasing English Language Learner Access to Academic Content**

Edynn Sato, Ph.D.

Assessment and Accountability Comprehensive Center and WestEd

**Introduction**

Educators and policymakers seeking to fairly and accurately measure the achievement of diverse learners and meet the requirements of the No Child Left Behind Act of 2001 (NCLB) need assurance that test content is as accessible as possible to all students, the test is equitable for all students, and the test results are valid and reliable. The focus of this guide is on improving these aspects of assessment for English language learners (ELLs) through *linguistic modification*. Linguistic modification is a method for increasing ELLs' access to assessment content and subsequently increasing assessment validity. [Textbox<sup>3</sup>— *Linguistic modification* is a process for increasing English language learners' access to assessment content and subsequently increasing assessment validity.] Linguistic modification is a viable assessment approach for states because, relative to other strategies such as portfolios and native language translations, it lends itself more readily to standardization and broader application across language groups. Thus, linguistic modification offers benefits to states in terms of both cost and feasibility. [Textbox— Relative to other strategies such as portfolios and native language translations, linguistic modification lends itself more readily to standardization and broader application across language groups.]

---

<sup>3</sup> Notes related to “textboxes” are for layout purposes. Textboxes are intended to highlight key information presented in the paper (e.g., as in textbooks or manuals).

As mentioned previously, this guide is one of two parts of an undertaking that focuses on linguistic modification. Part I presented the theory and principles underlying linguistically modified assessments and included a discussion of recent research on the effectiveness and validity of such assessments. Part II, which is this guide, underscores the critical considerations necessary for developing and implementing linguistically-modified assessments and presents frameworks and strategies for linguistic modification. [Textbox— This guide presents critical considerations necessary for developing and implementing linguistically-modified assessments and presents frameworks and strategies for linguistic modification.]

The frameworks and strategies presented in this guide will assist developers and consumers of academic content assessments for ELLs (e.g., state department of education staff, test developers, policymakers) meet relevant Title I requirements under NCLB. More important, this guide will inform the development of more valid linguistically-modified assessments and enable states to evaluate the degree to which assessment items address the linguistic access needs of ELLs.

### **Linguistic Modification**

*Linguistic modification* is a theory- and research-based approach in which the language in test items, directions, and/or response options is modified in ways that clarify and simplify the text without simplifying or significantly altering the construct that is being tested. To facilitate comprehension, linguistic modification reduces the construct-irrelevant language demands (e.g., semantic and syntactic complexity) of text through strategies such as reducing sentence length and complexity, using common or familiar words, and using concrete language (Abedi, Courtney, Mirocha, Leon, & Goldberg, 2005; Abedi, Lord, &

Plummer, 1997; Rivera & Stansfield, 2001). Linguistic modification is not simply “good editing” practice and does not result in “simpler” items. [Textbox—Linguistic modification is not simply “good editing” practice and does not result in “simpler” items.] Rather, it is a linguistically-based, systematic means for purposefully targeting and reducing the irrelevant variance in test performance that is attributable to individual differences in English proficiency so that ELLs are able to demonstrate fully their knowledge and skills related to the tested content. That is, construct-irrelevant language is minimized in order to facilitate ELL access to construct-relevant information and enhance students’ ability to focus on and demonstrate their knowledge and skills relevant to the assessed construct, yielding a *more valid* assessment for these students (Sato & Rabinowitz, forthcoming-b).

### **Critical Considerations for Linguistically Modified Assessments**

Assessing the achievement of ELLs poses a range of challenges to states working to satisfy the high expectations of NCLB; the unique characteristics of the ELL population and its particular access needs must be considered in order to ensure the fair and accurate measurement of student achievement (Sato & Rabinowitz, forthcoming-a). To assess this population, states typically implement one or more of the following practices: (1) incorporate principles of Universal Design (Bowe, 2000; Rochester, 2004); (2) provide accommodations that suit the particular needs and characteristics of this group of students (Center for Equity and Excellence in Education, 2005); and/or (3) develop alternate assessments based on the same content and performance standards as those used for the general education student population. However, each of these approaches has considerable technical challenges (Rabinowitz, Ananda, & Bell, 2005). Although states throughout the nation are working hard to meet NCLB requirements relevant to their ELLs, federal Peer

## LEP Partnership

Review reports have identified serious technical shortcomings of assessments for this student population. These significant technical limitations include validity, reliability, comparability, and the impact of accommodations on the meaningfulness of the test scores (U.S. Department of Education, 2006).

### Consideration of the English Language Learner Population

The ELL population is unique in a number of key ways that have implications for the valid assessment of this population (Liu, Anderson, Swierzbina, & Thurlow, 1999; Sato & Rabinowitz, forthcoming-a). The population's unique characteristics are related to factors that can limit the accurate measurement of the knowledge and skills of ELLs. One such factor is that ELLs lack sufficient literacy or fluency skills in the language of the test and lack the experience to understand the contextual content of items (Cummins, 1981; Solano-Flores & Li, 2006). Thus during assessment, ELLs may focus more on deciphering the language of the test than on addressing the tested content. Focusing on the language of the test that is not related to the tested content (construct-irrelevant language) may confound student performance and invalidate inferences about students' academic content achievement (Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliot, 2000; Rivera & Collum, 2004). The confounding effect of construct-irrelevant language (and linguistic load more generally) in testing is supported by evidence that the performance gap between ELL students and English-proficient students narrows on math items that have relatively lower language demands, such as math computation items (Abedi, 2001; Rivera & Stansfield, 2001). Thus, an implication for assessment is that the construct-irrelevant language of the



## LEP Partnership

test should be minimized in order to increase linguistic access<sup>4</sup> for ELLs and enable them to appropriately focus on the content of the test.

### Consideration of Access and Implications for Validity

Lack of appropriate *access* has consistently created the greatest threat to validity because it can contribute to construct-irrelevant variance and under-representation of the ELLs' knowledge and skills (Abedi, Courtney, & Leon, 2003; Abedi, Hofstetter, & Lord, 2004; Bielinski, Sheinker, & Ysseldyke, 2003; Elliot, Kratochwill, McKeivitt, 2001; Helwig, Rozek-Tedesco, Tindal, & Heath, 1999; Kopriva, Samuelson, Wiley, & Winter, 2003; Messick, 1993; Sireci, Li, & Scarpati, 2002; Thurlow & Wiener, 2000). [Textbox—Lack of appropriate *access* has consistently created the greatest threat to validity.]

Strategies that facilitate access take into account the particular needs of students to understand and to demonstrate their knowledge of content—whether those needs are cognitive or physical for students with disabilities, or linguistic for ELLs (Sato, Rabinowitz, & Gallagher, forthcoming).

**Access** is improved by systematically minimizing or removing sources of construct-irrelevant variance—without significantly altering the assessed construct—in order to

---

<sup>4</sup> *Access* refers to the minimization or removal of sources of construct-irrelevant variance, without significant alteration of the assessed construct, in order to facilitate students' ability to demonstrate their construct-relevant knowledge and skills. Strategies that facilitate access consider the particular needs of students (e.g., linguistic, cognitive, physical). When access is constrained, it could result in the measurement of sources of variance that are not related to the intended test constructs (*construct irrelevance*). Limited access could allow construct-irrelevant abilities to interfere with that student's ability to fully demonstrate what he or she knows and can do; consequently the test results underestimate that student's construct-relevant achievement. Inappropriate access could affect the construct such that the assessment no longer sufficiently represents the assessed domain (*under-representation*); that is, a student's proficiency on the assessment cannot be generalized to proficiency across the domain (Johnstone, 2003; Messick, 1993; Sato, Rabinowitz, & Gallagher, forthcoming).

facilitate students' ability to demonstrate their construct-relevant knowledge and skills. Simply providing students the opportunity to learn or exposing them to grade-level content may *not* increase access sufficiently, nor may the broad application of principles of Universal Design necessarily address ELLs' access needs adequately. [Textbox— *Access* is improved by systematically minimizing or removing sources of construct-irrelevant variance—without significantly altering the assessed construct. Providing students the opportunity to learn, exposing them to grade-level content, or applying broadly the principles of Universal Design do not necessarily address the particular access needs of English language learners adequately.]

Applying general principles of Universal Design<sup>5</sup> (UD), while necessary, is not sufficient to maximize access vis-à-vis the needs of the ELL population. The few studies that have examined the application of UD in large-scale assessment do not provide generalizable or conclusive evidence supporting the facilitation of access to typically underperforming students (Blackorby, Wagner, Cameto, Davies, Levine, Newman, Marder, & Sumi, 2005; Johnston, 2003; Johnstone, Thompson, Moen, Bolt, & Kato, 2005). Also, it is unclear whether universally designed items produce differential access for students in different subgroups (e.g., students with disabilities vs. ELLs), as suggested by the findings of Johnstone and his colleagues (2005). Nor is it clear whether the intended benefits of UD interact with test content (e.g., math vs. English language arts). Existing studies tend to rely

---

<sup>5</sup> *Universal Design* is a method for developing assessments in which user diversity is considered at each step of the process. Its goals are to maximize accessibility, regardless of student ability or disability. Universal Design is guided by a set of principles created to ensure access and strengthen the validity of inferences based on the assessment's results. For more information about Universal Design, see Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large-scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

on post-performance methods (item ranking, item intercorrelations, point biserials, DIF) for interpreting problems identified with items, since these methods are primarily intended to identify disparities in subgroup responses (or response patterns) that warrant further review (Sireci, Li, & Scarpati, 2002). Even if subsequent examinations of items suggest bias (i.e., the presence of a factor other than the construct being assessed that systematically advantages one subgroup over another), determination of the source of bias in such studies may be confounded by interactions among item characteristics (e.g., complexity, format), content idiosyncrasies, and student characteristics (e.g., language proficiency). Therefore, it is critical to consider the specific access needs of the student population (discussed in greater detail in a subsequent section of this guide) and their implications for test validity, rather than simply rely on general UD item development and review procedures. Linguistic modification is one such approach for addressing the particular access needs of ELLs and for increasing the validity of assessment results.

### **Developing a Linguistically Modified Assessment: Increasing Access for ELLs**

In this section, we begin describing an effective process for developing a linguistically modified assessment (for Title I). The process includes the following general steps:

- Defining the assessed domain and constructs;
- Distinguishing language related to the assessed content (construct-relevant language) from language not related to the assessed content (construct-irrelevant language);
- Defining the student population;
- Specifying the linguistic modification strategies that will be used; and

- Evaluating the effectiveness of linguistic modification.

Each step is discussed in greater detail below. Strategies for linguistic modification (Table 2a and Table 2b) and examples of linguistically modified items (Table 3) are presented. A *User's Guide* to this approach also is provided.

### Defining the Assessed Domain and Constructs<sup>6</sup>

Development of a valid assessment depends on clarity of the assessed domain (e.g., English language arts, mathematics) and consistency between this domain definition and the definitions of the assessed content and skills. Therefore, it is critical that state departments evaluate (a) the clarity of the parameters of the assessed domain (breadth, depth, range of complexity) and (b) the clarity and coherence of the definitions of the relevant content and skills—these definitions drive the development of the assessment.

[Textbox— It is critical that state departments verify the clarity of the assessed domain and the cohesiveness of the definitions driving the development of the assessment for English language learners.]

State departments ought to consider the following two questions:

1. What are the targeted skills/content within the assessed domain (content-related skills, concepts, language)?
2. What breadth, depth, and range of complexity of content are intended to be assessed within the domain?

---

<sup>6</sup> It is critical to note that when the assessment's targeted content/construct is language, as in English language arts, caution must be taken against altering the construct-relevant language of the assessment.

While addressing these questions is necessary for valid assessment development regardless of the student population tested, there are specific considerations that facilitate the development of a linguistically modified assessment to address the access needs of ELLs.

For ELL assessment, states should consider the following additional questions:

1. What is a language skill versus a content skill?
  - a. With regard to language, what are the more general or foundational skills and functions versus the more specific academic skills and functions?
  - b. How interdependent are language and content skills (i.e., can the same construct/content be assessed using a range of linguistic complexity)?
2. What are the expectations for student performance (i.e., proficient vs. not proficient; levels of proficiency; proficiency expectations of ELLs and non-ELLs)?

Since language affects ELLs' ability to demonstrate what they know and can do, it is important to distinguish language central to the assessed content (construct-relevant language) from language that is not (construct-irrelevant language—e.g., language used in general test directions to students). Therefore, defining the domain and assessed constructs is necessary, but not sufficient for developing linguistically modified assessments. Distinguishing construct-relevant from construct-irrelevant language also is critical to developing such assessments.

Distinguishing Construct-Relevant from Construct-Irrelevant Language

**Language demands**<sup>7</sup> is a research- and theory-based framework that operationalizes the language needed by ELLs for academic achievement and helps clarify where *language* and *content* converge/diverge (language skills and/or content skills) as well as helps distinguish construct-relevant from construct-irrelevant language. [Textbox—*Language demands* are research- and theory-based and provide a framework for defining where *language* and *content* converge/diverge.] Language demands facilitate the systematic analysis of the specific linguistic skills and language functions in standards, curriculum and instructional materials, and assessments (Sato, Lagunoff, Worth, Bailey, & Butler, 2005; Bailey, Butler, & Sato, 2007).

Table 1a and Table 1b provide lists of language demands and their definitions. These components of language are influential to academic achievement and can be categorized in terms of linguistic skills and academic language functions (Butler, Lord, Stevens, Borrego, & Bailey, 2004; Cummins, 2005; Sato, Lagunoff, Worth, Bailey, & Butler, 2005; Schleppegrell, 2001; Bailey, Butler, & Sato, 2007). A language demand is categorized as a **linguistic skill** if it is fundamental to the development and use of language (Table 1a). A language demand is categorized as an **academic language function** if it is a contextual application of language (Table 1b). As a framework, *Language Demands* can be applied across content areas and grade levels just as, for example, *Depth of Knowledge* (Webb, 1997) as a framework can be applied across content areas.

[Insert Table 1a: Language Demands—Linguistic Skills]

---

<sup>7</sup> See also the description of linguistic features that may affect comprehension in Appendix B of Part I: *Language factors in the assessment of English language learners: The theory and principles underlying the linguistic modification approach* (Abedi, 2007).

[Insert Table 1b: Language Demands—Academic Language Functions]

Language demands help articulate student expectations related to language; that is, help specify language students need to meaningfully engage with—to understand and to demonstrate understanding of—the content. For example, students often are expected to “know” or “understand” aspects of content. Language demands help clarify expectations of *how* students are to demonstrate “knowing” and “understanding” content through their use of language (e.g., labeling, defining, explaining, hypothesizing, summarizing, synthesizing, persuading). [Textbox—As a framework, *Language Demands* can be applied across content areas and grade levels. Language demands help clarify the expectations of *how* students are to demonstrate “knowing” and “understanding” content through their use of language.] The specific expectations of language use in the content areas have implications for the amount (e.g., word, phrase, sentence, discourse) and complexity of language (e.g., structure, register) needed by ELLs to meaningfully and successfully engage with academic content. Thus language demands help distinguish language skills from content skills and construct-relevant from construct-irrelevant language—such distinctions are critical to developing linguistically modified assessments. Strategies for linguistic modification can be applied to reduce the complexity of the construct-irrelevant language in order to increase linguistic access for ELLs (strategies for linguistic modification are discussed in greater detail in a subsequent section).

### Defining the English Language Learner Population

ELLs are diverse in their educational history, background experiences, sociocultural practices, socioeconomic status, national origin, cultural values, ethnicity, fluency in native or home language, cognitive strengths and limitations, and levels of motivation. This

## LEP Partnership

diversity may affect how different groups of ELLs approach and interpret test items (Abedi, 2004; Abedi & Dietel, 2004; Garcia, 2000; Goh, 2004; Kopriva, 2000; LaCelle-Peterson & Rivera, 1994; Scribner, 2002; Solano-Flores & Trumbull, 2003). The dialect spoken, the amount of formal elementary and secondary schooling in the student's home country and in the U.S., the depth and breadth of their academic knowledge base, mobility, and the consistency of language instruction also impact ELLs' levels of language literacy and their interpretation of the contextual content of items (Albus, Bielinski, Thurlow, & Liu, 2001; Liu et al., 1999; Solano-Flores & Li, 2006). In many cases, the only commonality among groups of ELLs within a state is that they all are facing a language barrier in English-speaking schools. Because of group heterogeneity, the meaning of scores may differ across subgroups of linguistically diverse test takers and may invalidate assumptions about individual performance (Yansen & Shulman, 1996). These factors also may confound the determination of effectiveness of programs, interventions, and accommodations (Abedi 1999; Abedi, Lord, Hofstetter, & Baker, 2000).

Therefore, student characteristics that are unique to ELLs and may impact measurement of their knowledge and skills ought to be considered when implementing linguistic modification strategies.

English language learner characteristics that may impact the measurement of student knowledge and skills include:

- Language background/primary language
- Type and amount of language other than English spoken in the home
- Cultural background



- Age
- Degree of prior education
- Literacy in native language (academic)
- Number of years in the United States
- Number of years in American schools
- Amount of instruction in English
- Number of English-only classes taken
- Migration patterns
- Parent education
- Participation in the free and reduced lunch program

Consideration of the ELL population and characteristics that impact their access to tested content is critical to developing linguistically modified assessments; that is, for determining how linguistic modification strategies should be implemented in order to effectively increase linguistic access for these students (strategies for linguistic modification are discussed in greater detail in a subsequent section). Consideration of ELL population characteristics is *not* intended to suggest that different assessments be developed for each ELL subgroup. Rather, it is important that the assessment be accessible to all students. That is, the assessment and its tasks (e.g., graphics, cultural/situational references in a test item, construct-irrelevant language) should reflect sensitivity to the backgrounds and experiences (e.g., sociocultural, sociolinguistic, geographic) of ELLs that may affect their performance on the assessment.

Specifying Linguistic Modification Strategies—Current Research and Recent Refinements

The assessed domain/construct definition, distinctions between construct-relevant and construct-irrelevant language, and ELL population characteristics, particularly as they impact student access needs, are inter-related and affect the specific manner in which linguistic modification strategies are implemented.<sup>8</sup> The following is a description of linguistic modification strategies based on current research and revisions/refinements to these strategies based on recent research on facilitating ELL access to academic content.

*Current Research: General Guidelines for Linguistic Modification*

Currently, linguistic modification strategies address one or more of the following aspects of a test item/test:<sup>9</sup>

- Context
- Graphics
- Vocabulary/wording
- Sentence structure
- Format/style

The strategies aim to reduce the effects of language that are not central to the assessed content (construct-irrelevant), thereby facilitating access for ELLs without affecting the

---

<sup>8</sup> States are encouraged to work with their test development partners and advisors (e.g., Technical Advisory Committee) to ensure the linguistic modification strategies are (a) operationalized and implemented in a manner that addresses the linguistic access needs of the state's ELL population, and (b) that the linguistic modifications strategies do not alter the assessed constructs.

<sup>9</sup> *Context, graphics and format/style* are included among the aspects of an item addressed by linguistic modification because they are methods for purposefully presenting items such that linguistic access is facilitated and the assessed construct is not altered (e.g., supporting meaning, segmenting elements of language in bulleted lists, emphasizing key words using ALL CAPS or **bold-faced** font).

## LEP Partnership

assessed construct and subsequently the attributions of achievement that can be drawn from student scores.

Currently, strategies for linguistic modification follow general guidelines such as the following:

- Include items that use context (names, settings, scenarios) that will be familiar to students from diverse cultural and linguistic backgrounds and experiences.
- Use simple, clear vocabulary in test items (directions, stems, answer choices) and associated stimulus materials (e.g., passages, tables, graphs).
- Avoid using trade-specific jargon and culture-specific idioms and colloquialisms. Such language can be misleading to students who rely on literal interpretation.
- Remove superficial details and complex phrases or sentence structures. Students then may dedicate their cognitive resources to selecting or generating the correct response instead of translating text.
- Include items that provide clear, culture-neutral visual support (e.g., illustrations) as a reference, so students can verify their interpretation of text.

Table 2a presents current strategies for linguistic modification according to the aspects of a test item/test that each test item/test addresses (see Table 2b for refinements to these strategies based on recent research).

[Insert Table 2a: Linguistic Modification Guidelines and Strategies: What Current Research Suggests]

Specifying the linguistic modification strategies prior to assessment development (e.g., which strategies will be used and the parameters for their use/application) is important to

## LEP Partnership

ensure that ELL access is systematically and consistently addressed in the assessment/across the assessment tasks.

### *Recent Research: Implications for Refining Linguistic Modification Strategies*

The actual effectiveness of current practices for making assessments accessible, equitable, and valid for ELLs is limited (Abedi, Hofstetter, & Lord, 2004; Butler & Stevens, 2001; Castellon-Wellington, 2000; Holmes & Duron, 2002; National Research Council, 2002, 2004; Rivera & Collum, 2004; Rivera & Stansfield, 2001). However, recent research has begun to systematically examine the intended and actual impact of linguistic modification on ELL access (Sato & Rabinowitz, forthcoming-b). [Textbox—The actual effectiveness of current practices for making assessments for accountability accessible, equitable, and valid for English language learners is unclear. However, recent research has begun to systematically examine the *intended* and *actual* impact of linguistic modification on English language learner access.] This section presents additional information about linguistic modification that can inform the development of state assessments so that the test items increase access for ELLs (i.e., linguistic access). As mentioned previously, specification of the linguistic modification strategies prior to assessment development (e.g., which strategies will be used and the parameters for their use/application) is important to ensure that ELL access is systematically and consistently addressed in the assessment/across the assessment tasks.

One study by Sato and Rabinowitz (forthcoming-b) employs a combination of qualitative (e.g., expert judgment, cognitive interviews) and quantitative (e.g., analysis of student performance data) methods to begin systematically examining the comparability between original test items and their linguistically modified counterparts. The study

demonstrates that linguistic modification performed properly does not simplify the test content or significantly alter the tested construct.

Table 2b presents refinements to the strategies presented in Table 2a. These refinements are based on recent research on linguistic modification strategies and their impact on ELL access. These strategies and their outcomes (linguistically modified items) were validated by experts with a range of knowledge and experience (i.e., assessment, measurement, standards, instruction, curriculum, the academic content areas, applied linguistics, English language development, English language learner population).

[Textbox—Refinements to current research-based access strategies were validated by experts with a range of knowledge and experience (i.e., assessment and measurement, standards, instruction and curriculum, the academic content areas, applied linguistics, English language development, English language learner population).] In the process of refining the strategies and reviewing the linguistically modified items vis-à-vis the original items, the experts verified that the content and format of each item supported linguistic modification (e.g., the item had construct-irrelevant language that could be linguistically modified), but that linguistic modification did not significantly simplify the content such that the construct was altered (Sato & Rabinowitz, forthcoming-b).

[Insert Table 2b: Linguistic Modification Guidelines and Strategies: Refinements Based on Recent Research Findings]

Table 3 presents example items that were linguistically modified (original and linguistically modified items). Each example includes notes regarding the linguistic and population-specific considerations underlying each item's modifications.

[Insert Table 3: Examples of Original and Linguistically Modified Items]

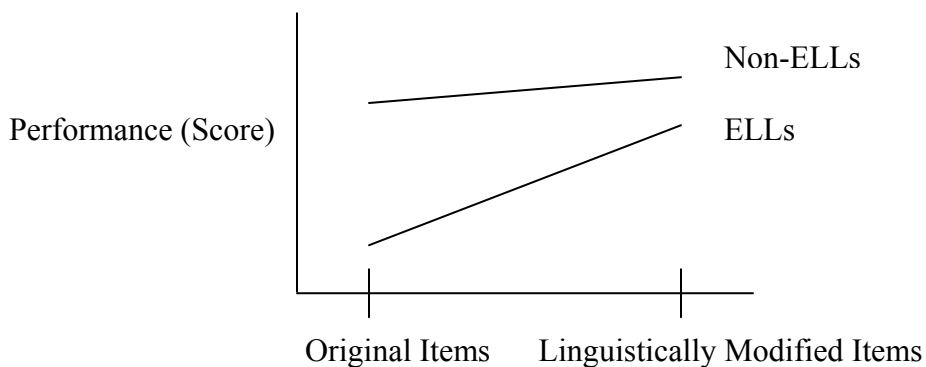
## LEP Partnership

Specifying the linguistic modification strategies (e.g., which strategies will be used and the parameters for their use/application) and having examples of linguistically modified items (i.e., illustrative examples of the application/operationalization of linguistic modification strategies) prior to assessment development will help ensure that ELL access is systematically and consistently addressed in the assessment/across the assessment tasks.

### Evaluating the Effectiveness of Linguistic Modification

#### *Examining the Effect of Linguistic Modification*

Linguistically modified items may improve performance for all students since the process may result in clearer, better-defined items. However, the expectation is that the improvement in performance on items (i.e., linguistically modified vs. original) is greater for the population of students who are ELLs because of their greater language difficulties. The figure below illustrates the expected effect of linguistically modified items on the performance of ELLs and non-ELLs.



The process of developing a linguistically modified assessment should include some evaluation of the effectiveness of the linguistic modification strategies (i.e., evaluation of the degree to which the linguistic modification strategies increase ELL access

as intended). The degree to which linguistically modified items increase ELL access as intended can be determined in a number of ways. [Textbox— The process of developing a linguistically modified assessment should include some evaluation of the effectiveness of the linguistic modification strategies.]

In addition to expert reviews of items (e.g., individuals with expertise in English language development, the ELL population, assessment, standards), there are various statistical analyses that can be conducted to examine the degree to which linguistic modification impacts student performance gains. States (and/or test contractors) should consider using one or more methods (qualitative, quantitative) to examine the effectiveness of their linguistically modified assessments.

Analysis of Variance Student performance scores, disaggregated by group (ELLs vs. non-ELLs) and item set version (linguistically modified vs. original) can be summarized to provide information about how each group performs on each set of items. Mean differences in scores for the two student populations can be compared in order to examine if ELLs are better able to demonstrate their ability on the linguistically modified set of items. If linguistic modification provides ELLs greater access to the content, then the score difference between original and linguistically modified item sets should be greater for the ELLs than for the non-ELLs.

Differential Item Functioning Analysis An analysis of differential item functioning (DIF) will address whether items in their original form that were flagged for possible bias against ELLs are no longer flagged in their linguistically modified form.

Factor Analyses Exploratory factor analysis can be conducted to estimate the number of constructs assessed by the set of items and the underlying measurement

structure (correlations) of the unobservable (latent) factor(s). Comparisons can be made for the original item set versus the linguistically modified item set.

Cognitive Interviews With a purposeful sample of students (e.g., by gender, race, primary language), a trained researcher can guide individual students through a think-aloud protocol as students work on original and linguistically modified items (Paulsen & Levine, 1999). Findings from the cognitive interviews can be used to understand the strategies students use to access the content of each item—verifying the degree to which the intended impact of linguistic modification strategies on student access is consistent with the actual impact on student access.

*Content and Bias Reviews: Consideration of the ELL Population*

Research suggests that there is substantial, but not total, overlap between the procedures and criteria found appropriate and essential for establishing the technical adequacy of ELL assessments versus those of their general education counterparts (Rabinowitz & Sato, 2005). Some criteria do not transfer directly or are less critical for establishing the technical adequacy of ELL assessments. For example, the definition of the referent groups for sampling or DIF analyses should be sensitive to the heterogeneity of the ELL population. Similarly, the procedures and criteria widely used for reviews of assessments for non-ELL populations and establishment of the validity of these assessments may need to be modified before they are applied to more specialized populations, such as ELLs.

For ELL assessment, unless the individuals involved in the test development process (from design through implementation) truly understand the ELL population and how to minimize the influences of construct-irrelevant factors (e.g., construct-irrelevant



language, unfamiliar cultural/contextual factors), validity will suffer and *real* student achievement will be underestimated. [Textbox— For ELL assessment, unless the individuals involved in the test development process truly understand the English language learner population and how to minimize the influences of construct-irrelevant factors, validity will suffer and *real* student achievement will be underestimated.] Therefore, in designing, implementing, and evaluating a linguistic modification approach to test development, individuals with relevant knowledge and experience must be included in all key aspects of an assessment’s development, including item and form review.

These guidelines present information on linguistic modification as a viable assessment approach for states. Linguistic modification is a strategy supported by both theory and research, and relative to other strategies, such as portfolios and native language translations, linguistic modification lends itself more readily to standardization and broader application across language groups. The following is a *User’s Guide* that outlines steps in the process for developing a linguistically modified assessment.

## **USER’S GUIDE: AN APPROACH TO DEVELOPING A LINGUISTICALLY MODIFIED ASSESSMENT**

For all students, *access* to test content is necessary to ensure the validity of the assessment results, and valid assessments are especially critical if results are used for instructional and/or accountability purposes. When access is constrained in some way (e.g., linguistically, cognitively), a student may be prevented from fully demonstrating what he or she knows and can do; subsequently the test results could underestimate or misrepresent the student’s achievement. When assessing English language learners’ knowledge of academic content, it is critical to determine whether their academic performance is reflective of their understanding of the content versus their understanding of the English language. There is an interaction between the assessed content and the access needs of the English language learner population. This interaction affects validity of the interpretation of the assessment results.

Linguistic modification is an approach for addressing the particular access needs of English language learners such that the variance in test performance that is attributable to individual differences in English proficiency is reduced, and the students are able to more fully demonstrate their knowledge and skills related to the tested content. The following approach is intended to help consumers and developers of assessments consider key characteristics of the content and the student population as they develop a linguistically modified assessment.

### **STEP I: DEFINE THE DOMAIN AND CONSTRUCTS**

Articulate the purpose(s) of your assessment. Consider the range of ways the assessment results will be used and the intended outcomes.

Recommendation: Given the purpose(s) of the assessment and the population assessed, it is recommended that the following individuals be involved in this step: content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and individuals with knowledge of the English language learner population.

**Purpose(s):**

The assessment results will be used for the following purpose(s):

---

---

**Assessed Domain:**

The assessment will measure students' knowledge of:

---

---

Evaluate:

- Does the assessed domain match the stated purpose(s) of your assessment?

**Assessed Constructs—Content and Skills:**

More specifically, the assessment will measure the following constructs (content and skills) related to the domain:

---

---

---

---

---

Evaluate:

- Do the content and skills reflect the intended breadth, depth, and range of complexity of the assessed domain?
- Are the verbs used in the standards statements for what students are expected to know and do specific enough to guide assessment development (e.g., identify, describe, compare), or are the verbs vague in the standards statements (e.g., know, understand)? If the latter, consider how students are expected to demonstrate knowledge and understanding.

**Content-related Language—Language Demands:**

The following language demands are associated with the content and skills that will be assessed (see **Table 1a** and **Table 1b** for Language Demands: Linguistic Skills and Academic Language Functions):

---

---

---

---

---

Evaluate:

- Have both *linguistic skills* and *academic language functions* been considered?
- Is the range of language demands consistent with the breadth, depth, and range of complexity of the assessed domain (content, skills)?

**Content-related Language—Specific Vocabulary and Terminology:**

The following vocabulary and terminology are specific to the grade-level content assessed; therefore, they should *not* be linguistically modified:

---

---

---

---

---

Evaluate:

- The vocabulary and terminology identified is consistent with the intent of the grade-level content standards.

## STEP II: DEFINE THE POPULATION

Articulate the key characteristics and access needs of the English language learner population. Since this group of students is especially diverse and heterogeneous, you may need to first specify key subgroups of students within the state.

Recommendation: Given the purpose(s) of the assessment and the population assessed, it is recommended that the following individuals be involved in this step: content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and individuals with knowledge of the English language learner population.

### Student Population:

The English language learner population tested can be characterized as follows (see pp. 63-65 for a discussion and list of English language learner characteristics):

---

---

---

---

### Student Access Needs:<sup>10</sup>

Given the characteristics of the English language learner population, their access needs include:

Context: What topics, themes, locations, situations, illustrations, etc. are familiar to these students?

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

---

<sup>10</sup> *Context* and *format/style* should be considered because they affect student access to the content and have implications for the language used to present the content.

## LEP Partnership

Words, Phrases, Sentences: What is familiar to these students (written vocabulary, phrases, sentence structures, tenses [e.g., present, past], constructions [e.g., plural \_s, possessive \_'s], proper names [i.e., in reading])?

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

Format/Style: With what formats/styles are these students familiar (e.g., bulleted lists, text boxes, underlining for emphasis)? How is information typically presented to these students during instruction?

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_

### STEP III: APPLICATION AND EVALUATION OF LINGUISTIC MODIFICATION STRATEGIES

Once Steps I and II are complete, determine which content/item lends itself to linguistic modification and develop/evaluate each test item according to the following dimensions: context; graphics; vocabulary/wording; sentence structure; and format/style (see **Table 2b** for linguistic modification guidelines and strategies for each dimension).<sup>11</sup>

Note: It is recommended that the following individuals be involved in all aspects of Step III: content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and individuals with knowledge of the English language learner population.

#### Step A: Categorize Content/Items

Sort content/items into one of three categories (within each category, keep content standards/items grouped by strand—e.g., Measurement, Reading):

1. Definitely eligible for linguistic modification;
2. Definitely NOT eligible for linguistic modification; or
3. Possibly eligible for linguistic modification.

Note: Eligibility is determined by the degree of construct-irrelevant language reflected in the content standard (i.e., how the standard is typically assessed with a test item) or present in the item—the more construct-irrelevant language, the more likely the content/item is eligible for linguistic modification. That is, the more likely construct-irrelevant language can be modified without significantly changing the assessed construct.

#### Step B: Apply Linguistic Modification Guidelines and Strategies

For content/items that are eligible and possibly eligible for linguistic modification, systematically apply the relevant guidelines and strategies presented in **Table 2b** (i.e., context, graphics, vocabulary/wording, sentence structure, format/style). Examples of linguistically modified items also are available for reference (**Table 3**).

Note: Training is critical. It is important to ensure the guidelines and strategies are accurately and consistently applied to the development of the items and that there are no

---

<sup>11</sup> *Context, graphics and format/style* are included among the aspects of an item addressed by linguistic modification because they are methods for purposefully presenting items such that linguistic access is facilitated and the assessed construct is not altered (e.g., supporting meaning, segmenting elements of language in bulleted lists, emphasizing key words using ALL CAPS or **bold-faced** font).

changes to the intended construct, cognitive complexity, and language demand(s) or the content vocabulary/terminology specified in the grade-level standards.

### **Step C: Evaluate the Linguistically Modified Items**

For each item, verify:

- the construct has not changed;
- the cognitive complexity of the item is appropriate; and
- the linguistically modified version provides linguistic access to ELLs in terms of
  - context
  - graphics
  - vocabulary/wording
  - sentence structure
  - format/style

Methods for verification include:

- Expert review (e.g., Technical Advisory Committee, content and bias review committees, independent external reviewers)
- Analysis of Variance
- Differential Item Functioning
- Factor Analysis
- Cognitive Interviews



**References: Part II**

Abedi, J. (2001). *Assessment and accommodations for English Language Learners: Issues and recommendations*. CRESST Policy Brief 4. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

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

Abedi, J. (2007). *Language factors in the assessment of English language learners: The theory and principles underlying the language modification approach*. Paper for the U.S. Department of Education—LEP Partnership.

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English Language Learners in NAEP*. Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). *Language accommodations for English Language Learners in large-scale assessments: Bilingual dictionaries and linguistic modification*. Los Angeles: University of California, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J. & Dietel, R. (2004). Challenges in the No Child Left Behind Act for English Language Learners. *Phi Delta Kappan*, 85(10), 782-785.

## LEP Partnership

- Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, 74(1), 1-28.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16-26.
- Abedi, J., Lord, C., & Plummer, J. (1995). *Language background as a variable in NAEP mathematics performance: NAEP task 3D: Language background study*. CSE Technical Report 429. Los Angeles: University of California, Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.
- Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test*. NCEO LEP Project Report 1. Minneapolis: National Center on Educational Outcomes.
- Bailey, A., Butler, F., & Sato, E. (2007). Standards-to-standards linkage under Title III: Exploring common language demands in ELD and science standards. *Applied Measurement in Education*.
- Bielinski, J., Sheinker, A., Ysseldyke, J. (2003, April). *Varied opinions on how to report accommodated test scores*. NCEO Synthesis Report 49. Minneapolis: National Center on Educational Outcomes.
- Blackorby, J., Wagner, M., Cameto, R., Davies, E., Levine, P., Newman, L., Marder, C., & Sumi, C. (2005). *Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities*. Paper

- presented at the Technology and Persons with Disabilities Conference at the Center on Disabilities, California State University, Northridge.
- Bowe, F.G. (2000). *Universal design in education: teaching non-traditional students*. Westport, CT: Bergen & Garvey.
- Butler, F.A., Lord, C., Stevens, R., Borrego, M., & Bailey, A.L. (2004). *An approach to operationalizing academic language for language test development purposes: Evidence from fifth-grade science and math* (CSE Report 626). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing. Retrieved March 21, 2007 from <http://www.cse.ucla.edu/products/Reports/R626.pdf>
- Butler, F.A. & Stevens, R. (2001). Standardized assessment of the content knowledge of English Language Learners K-12: Current trends and old dilemmas. *Language Testing 2001*, 18(4), 409-427.
- Castellon-Wellington, M. (2000). *The impact of preference for accommodations: The performance of ELLs on large-scale academic achievement tests*. CRESST Technical Report 524. Los Angeles: University of California, National Center for the Study of Evaluation, Standards, and Student Testing.
- Center for Equity and Excellence in Education (2005). *Recommendations for State ELLs Accommodation Policies*. Retrieved July 10, 2005, from the World Wide Web: [http://ceee.gwu.edu/AA/Accommodations\\_Recos.html](http://ceee.gwu.edu/AA/Accommodations_Recos.html)
- Cummins, J. (1981). Four misconceptions about language proficiency in bilingual education. *NABE Journal*, 5(3), 31-45.

- Cummins, J. (2005). Language proficiency, bilingualism, and academic achievement. In P.A. Richard-Amato and M.A. Snow (Eds.), *Academic success for English language learners* (pp. 76-86). White Plains, NY: Longman.
- Elliott, S., Kratochwill, T.R., & McKeivitt, B.C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39*(1), 3-24.
- Garcia, G.N. (2000). Lesson from research: What is the length of time it takes limited English proficient students to acquire English and succeed in an all-English classroom? *National Clearinghouse for Bilingual Education, 5*. Washington, DC.
- Goh, D.S. (2004). *Assessment accommodations for diverse learners*. Boston: Pearson.
- Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93*(2), 113-125.
- Holmes, D. & Duron, S. (2000). *LEP students and high stakes assessment*. Washington, DC: National Clearinghouse for Bilingual Education, US Department of Education.
- Johnstone, C. J. (2003). *Improving validity of large-scale tests: Universal Design and student performance*. NCEO Technical Report No. 37. Minneapolis, MN: National Center on Educational Outcomes.
- Johnstone, C. J., Thompson, S.J., Moen, R.E., Bolt, S., & Kato, K. (2005). *Analyzing results of large-scale assessments to ensure universal design*. NCEO Technical Report No. 41. Minneapolis, MN: National Center on Educational Outcomes.
- Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State School Officers.

- Kopriva, R., Samuelson, K., Wiley, D., & Winter, P. (2003). *Evidentiary logic in the assessment of diverse learners*. Paper presented at the annual conference of the National Council on Measurement in Education, Chicago, IL.
- LaCelle-Peterson, M.W. & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64, 55-75.
- Liu, K., Anderson, M., Swierzbin, B., & Thurlow, M. (1999). *Bilingual accommodations for Limited English Proficient students on statewide reading tests*. NCEO State Assessment Series, Minnesota Report 20.
- Messick, S (1993). Validity. In R .L. Linn (Ed.). *Educational Measurement* (pp. 13-104). Phoenix, AZ: Oryx.
- National Research Council. (2002). In J. Koenig, (Eds.), *Reporting test results for students with disabilities and English-language learners*. Washington, DC: National Academies.
- National Research Council. (2004). In J.Koenig & L. Bachman, (Eds.), *Keeping score for all: The effects of inclusion and accommodation policies on large-scale educational assessments*. Washington, DC: National Academies.
- Paulsen, C. A. & Levine, R. (1999). *The applicability of the cognitive laboratory method to the development of achievement test items*. Paper presented at the annual meeting of the American Educational Research Association, Montreal.
- Rabinowitz, S., Ananda, S., & Bell, A. (2005). Strategies to assess the core academic knowledge of English language learners. *Journal of Applied Testing Technology*.

- Rabinowitz, S. & Sato, E. (2005). *The technical adequacy of assessments for alternate student populations*. San Francisco: WestEd.
- Rivera, C. & Collum, E. (2004). *An analysis of state assessment policies addressing the accommodation of English Language Learners*. Issue paper commissioned for the National Assessment Governing Board Conference on Increasing the Participation of SD and LEP Students in NAEP. Arlington, VA: George Washington University.
- Rivera, C. & Stansfield, C.W. (2001). *The effects of linguistic simplification of science test items on performance of Limited English Proficient and monolingual English-speaking students*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Rochester Institute of Technology. (2004). *Class act: Access for deaf and hard-of-hearing students*. Online source: <http://www.rit.edu/~classact/site/universaldesign.html>
- Sato, E., Lagunoff, R., Worth, P., Bailey, A. L., & Butler, F. A. (2005). *ELD Standards Linkage and Test Alignment Under Title III: A Pilot Study of the CELDT and the California ELD and Content Standards*. (Final Report to the California Department of Education). San Francisco: WestEd.
- Sato, E. & Rabinowitz, S. (forthcoming-a). *Assessing English Language Learners: Critical Considerations for Validity*. [working title].
- Sato, E. & Rabinowitz, S. (forthcoming-b). *Assessment Accommodations for English Language Learners: An Empirical Investigation of Strategies to Improve Access* [working title].
- Sato, E., Rabinowitz, S., & Gallagher, C. (forthcoming). *Access and Special Student Populations—the Similarities/Differences in the Needs of English Language*

- Learners and Students with Disabilities: Implications for Standards, Assessment, and Instruction* [working title].
- Schleppegrell, M.J. (2001). Linguistic features of the language of schooling. *Linguistics and Education*, 12 (4), 431-459.
- Scribner, A.P. (2002). Best assessment and intervention practices with second language learners. In A.Thomas & J.Grimes (Eds.), *Best practices in school psychology IV*. Washington, DC: National Association of School Psychologists.
- Sireci, S.G., Li, S., & Scarpati, S. (2002). *The effects of test accommodations on test performance: A review of the literature*. CEA Research Report 485. Amherst, MA: School of Education, University of Massachusetts.
- Solano-Flores, G. & Li, M. (2006). The use of generalizability (G) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice*, 25 (1), 13-22.
- Solano-Flores, G. & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English language learners. *Educational Researcher*, 32(2), 3-13.
- Thurlow, M.L., McGrew, K.S., Tindal, G., Thompson, S.J., Ysseldyke, J.E., & Elliot, J.L. (2000). *Assessment accommodations research: Considerations for design and analysis*. NCEO Technical Report 26. Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M. L., & Wiener, D. (2000). *Non-approved accommodations: Recommendations for use and reporting* (Policy Directions No. 11). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

LEP Partnership

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. (NISE Research Monograph No. 8). Madison: University of Wisconsin—Madison, National Institute for Science Education. Washington, DC: Council of Chief State School Officers.

Yansen, E. & Shulman, E. (1996). Language assessment: multicultural considerations. In L. Suzuki, P. Meller, & J. Ponterotto (Eds.), *The handbook of multicultural assessment* (pp. 353–393). San Francisco: Jossey-Bass



**Table 1a: Language Demands—Linguistic Skills<sup>12</sup>**

<b>Language Demand</b> <i>The item/task or standard requires students to demonstrate knowledge of:</i>	<b>Definition/Application of Linguistic Skill</b> <i>The ability to:</i>	<b>Notes</b>
Phonemes	Identify, manipulate, and produce the individual distinctive sounds that make up spoken words. (L, S)	Phonology
Syllables	Identify the division of words into the smallest units of sequential speech sounds, composed of a vowel sound or a vowel-consonant combination. (L, R)	Phonology
Morphemes	Identify and distinguish the smallest unit of meaning (i.e., words, roots, or affixes). (L, R)	Morphology
Vocabulary Words	Identify and determine meaning of spoken or written words or short phrases in context; produce spoken or written words relevant to a particular context. (L, S, R, W)	Lexicon
Phrases & Sentences	Determine meaning of spoken and written phrases and sentences; generate original phrases and sentences using grammatical forms. (L, S, R, W)	Syntax Semantics
Sound-Symbol Correspondences	Identify the relationship between letters of written language (graphemes) and the individual sounds (phonemes). (R, W)	Orthography
Written English Conventions	Recognize and apply written English conventions [i.e., punctuation, capitalization, spelling, paragraph structure, format (including text features)]. (R, W)	Writing conventions

Related Language Modalities: Listening (L); Speaking (S); Reading (R); Writing (W)  
 [From: Sato, Lagunoff, Worth, Bailey, & Butler (2005); Bailey, Butler, & Sato (2007);  
 Sato, Rabinowitz, & Gallagher (forthcoming)]

<sup>12</sup> See also the description of linguistic features that may affect comprehension in Appendix B of Part I: *Language factors in the assessment of English language learners: The theory and principles underlying the linguistic modification approach* (Abedi, 2007).

**Table 1b: Language Demands—Academic Language Functions**

<b>Language Demand</b>  <i>The item/task or standard requires students to use language to:</i>	<b>Definition/Application of Academic Language Function</b>  <i>The ability to:</i>
Identify	Use a word or phrase to name an object, action, event, idea, fact, problem, need, or process.
Label	Use a word or phrase to name an object, action, event, or idea.
Enumerate	Use words or phrases to name distinct objects, actions, events, or ideas in a series, set, or in steps.
Classify	Use words, phrases, or sentences to assign/associate an object, action, event, or idea to the category or type to which it belongs.
Sequence	Use words, phrases, or sentences to express the order of information (e.g., a series of objects, actions, events, ideas). Discourse markers include adverbials such as <i>first, next, then, finally</i> .
Organize	Use words, phrases, or sentences to express relationships between/among objects, actions, events, or ideas, or the structure or arrangement of information. Discourse markers include coordinating conjunctions such as <i>and, but, yet, or</i> , and adverbials such as <i>first, next, then, finally</i> .
Compare/Contrast	Use words, phrases, or sentences to express similarities and/or differences between two or more objects, actions, events, or ideas. Discourse markers include coordinating conjunctions <i>and, but, yet, or</i> , and adverbials such as <i>similarly, likewise, in contrast, instead, despite this</i> .
Inquire	Use words, phrases, or sentences to solicit information (e.g., <i>yes-no</i> questions, <i>wh</i> -questions, statements used as questions).
Describe	Use a word, phrase, or sentence to express the attributes or properties of an object, action, event, or idea.
Define	Use a word, phrase, or sentence to express the meaning of a given word, phrase, or expression.
Explain	Use phrases or sentences to express the rationale, reasons, causes, or relationships related to one or more actions, events, ideas, or processes. Discourse markers include coordinating conjunctions <i>so, for</i> , and adverbials such as <i>therefore, as a result, for that</i>

<b>Language Demand</b>  <i>The item/task or standard requires students to use language to:</i>	<b>Definition/Application of Academic Language Function</b>  <i>The ability to:</i>
	<i>reason.</i>
Retell	Use phrases or sentences to relate or repeat information. Discourse markers include coordinating conjunctions such as <i>and, but</i> , and adverbials such as <i>first, next, then, finally</i> .
Summarize	Use phrases, sentences, and/or paragraphs to express important facts or ideas and relevant details about one or more objects, actions, events, ideas, or processes. Discourse structures include: beginning with an introductory sentence that specifies purpose or topic.
Interpret	Use phrases or sentences to express understanding of the intended or alternate meaning of information.
Analyze	Use phrases or sentences to indicate parts of a whole and/or the relationship between/among parts of an action, event, idea, or process. Relationship verbs such as <i>contain, entail, consist of</i> , partitives such as <i>a part of, a segment of</i> , and quantifiers such as <i>some, a good number of, almost all, a few, hardly any</i> often are used.
Generalize	Use phrases or sentences to express an opinion, principle, trend, or conclusion that is based on facts, statistics, or other information, and/or to extend that opinion/principle/etc. to other relevant situations/contexts/etc.
Infer	Use words, phrases, or sentences to express understanding of implied/implicit based on available information. Discourse markers include inferential logical connectors such as <i>although, while, thus, therefore</i> .
Predict	Use words, phrases, or sentences to express an idea or notion about a future action or event based on available information. Discourse markers include adverbials such as <i>maybe, perhaps, obviously, evidently</i> .
Hypothesize	Use phrases or sentences to express an idea/expectation or possible outcome based on available information. Discourse markers include adverbials such as <i>generally, typically, obviously, evidently</i> .
Argue	Use phrases or sentences to present a point of view with the intent of communicating or supporting a particular position or

<b>Language Demand</b>  <i>The item/task or standard requires students to use language to:</i>	<b>Definition/Application of Academic Language Function</b>  <i>The ability to:</i>
Persuade  Negotiate	<p>conviction. Discourse structures include expressions such as <i>in my opinion, it seems to me</i>, and adverbials such as <i>since, because, although, however</i>.</p> <p>Use phrases or sentences to present ideas, opinions, and/or principles with the intent of creating agreement around or convincing others of a position or conviction. Discourse markers include expressions such as <i>in my opinion, it seems to me</i>, and adverbials such as <i>since, because, although, however</i>.</p> <p>Use phrases or sentences to engage in a discussion with the purpose of creating mutual agreement from two or more different points of view.</p>
Synthesize	<p>Use phrases or sentences to express, describe, or explain relationships among two or more ideas. Relationship verbs such as <i>contain, entail, consist of</i>, partitives such as <i>a part of, a segment of</i>, and quantifiers such as <i>some, a good number of, almost all, a few, hardly any</i> often are used.</p>
Critique	<p>Use phrases or sentences to express a focused review or analysis of an object, action, event, idea, or text.</p>
Evaluate	<p>Use phrases or sentences to express a judgment about the meaning, importance, or significance of an action, event, idea, or text.</p>

Note: Some academic language functions are grouped; this occurs when the language demands are very similar, or represent multiple levels of essentially the same demand. [From: Sato, Lagunoff, Worth, Bailey, & Butler (2005); Bailey, Butler, & Sato (2007); Sato, Rabinowitz, & Gallagher (forthcoming)]

**Table 2a: Linguistic Modification Guidelines and Strategies: What Current Research Suggests<sup>13</sup>**

<p><b>Item Context</b></p> <ul style="list-style-type: none"> <li>▪ Familiar to students; no cultural or linguistic bias</li> </ul>	<p><b>Item Graphics</b></p> <ul style="list-style-type: none"> <li>▪ Familiar to students; no cultural or linguistic bias</li> <li>▪ Symbols, legends, and key vocabulary are relevant to the construct and familiar to ELLs</li> <li>▪ Graphic and labeling/naming conventions are consistent</li> <li>▪ Graphics support ELLs' understanding of assessed content</li> </ul>
<p><b>Item Vocabulary/Wording</b></p> <ul style="list-style-type: none"> <li>▪ High frequency words</li> <li>▪ Common and familiar words</li> <li>▪ Relevant technical terms; technical terms defined as appropriate</li> <li>▪ No ambiguous words or unnecessary words with multiple meanings</li> <li>▪ No irregularly spelled words</li> <li>▪ Use of formal proper names is relevant and appropriate</li> <li>▪ Naming conventions are consistent with graphics/stimuli</li> </ul>	<p><b>Item Sentence Structure</b></p> <ul style="list-style-type: none"> <li>▪ Referents are clear; noun-pronoun relationships are clear; antecedent references are made clear</li> <li>▪ Grammatical structures are clear Typically, <ul style="list-style-type: none"> <li>▪ past or future tense verb forms are changed to present</li> <li>▪ passive verb forms are changed to active verb forms</li> <li>▪ complex structures are changed to subject-verb-object structure</li> <li>▪ long nominals/names/phrases are shortened (e.g., "last year's class vice-president" becomes "a student leader")</li> <li>▪ compound sentences are replaced with two separate sentences, especially in comparative structures</li> <li>▪ long prepositional phrases are reduced or removed</li> <li>▪ conditional clauses are replaced with separate sentences or the ordering of a clause within a sentence is changed for clarity</li> <li>▪ relative clauses are removed or rephrased for clarity</li> </ul> </li> <li>▪ Questions framed in negative terms are rephrased</li> </ul>
<p><b>Item Format/Style</b></p> <ul style="list-style-type: none"> <li>▪ Parts of the item/question are clear; distinctions are relevant and appropriate</li> <li>▪ Long problem statements are segmented or shortened</li> <li>▪ The order of operations is explicit</li> </ul>	

[From: Sato & Rabinowitz (forthcoming-b)]

<sup>13</sup> *Context, graphics and format/style* are included among the aspects of an item addressed by linguistic modification because they are methods for purposefully presenting items such that linguistic access is facilitated and the assessed construct is not altered (e.g., supporting meaning, segmenting elements of language in bulleted lists, emphasizing key words using ALL CAPS or **bold-faced** font).

**Table 2b: Linguistic Modification Guidelines and Strategies: Refinements Based on Recent Research Findings<sup>14</sup>**

<b>Original</b>	<b>Refinement</b>
<p><b>Item Context</b></p> <ul style="list-style-type: none"> <li>▪ Familiar to students; no cultural or linguistic bias</li> </ul>	<p><b>Context</b>=situation; situates the problem (includes drawing a relationship between things)</p> <ul style="list-style-type: none"> <li>▪ Sometimes context is added to           <ul style="list-style-type: none"> <li>▪ remove passive voice construction in original item</li> <li>▪ remove past tense and conditionals in original item</li> <li>▪ break stem into shorter, less complex sentences (sometimes creates a story line)</li> </ul> </li> <li>▪ Context helps make language that is reflective of abstract/highly-generalized situations more concrete and relevant to ground the content being tested</li> <li>▪ Context for ELLs that facilitates access includes:           <ul style="list-style-type: none"> <li>▪ Concrete language</li> <li>▪ Illustrative language</li> <li>▪ Illustration/graphic               <ul style="list-style-type: none"> <li>▪ Sometimes adding a graphic reduces the need for language and/or the language load (see below for discussion of graphics)</li> </ul> </li> </ul> </li> </ul>
<p><b>Item Graphics</b></p> <ul style="list-style-type: none"> <li>▪ Familiar to students; no cultural or linguistic bias</li> <li>▪ Symbols, legends, and key vocabulary are relevant to the construct and familiar to ELLs</li> <li>▪ Graphic and labeling/naming conventions are consistent</li> <li>▪ Graphics support ELLs' understanding of assessed content</li> </ul>	<p><b>Graphics</b>=Includes diagrams, tables, charts, drawings, graphs, pictures, maps  <u>Note:</u> When a graphic is referred to/used in an item, it is referred to by its specific name (e.g., diagram, map, chart)</p> <ul style="list-style-type: none"> <li>▪ Graphics clarify (1) key aspects of the content/construct assessed and/or (2) what the student is expected to do—graphics use should be purposeful.</li> <li>▪ Graphics support context without requiring additional language (and may reinforce what</li> </ul>

<sup>14</sup> The author would like to thank the following individuals for informing the refinement of the linguistic modification guidelines and strategies presented in Table 2b: Patricia Armstrong, Joanne DaLuz, Paula Diamanti, Carole Gallagher, Cassandra Hawley, Loretta Kelley, Holly McKeag, Rachel Lagunoff, Claudia Long, Susan Porter, Stanley Rabinowitz.

Original	Refinement
	<p>is in the text of the item).</p> <ul style="list-style-type: none"> <li>▪ Graphics help students shift from one context to another within an assessment (e.g., from one type of item to another).</li> <li>▪ Allows student to verify understanding of key elements of the text of the item</li> <li>▪ Allows representation of key elements of the problem (necessary information; construct-relevant) so that this information does not need to be presented in words.</li> <li>▪ <u>Consideration</u>: How central is the information in the graphic to the construct? For example, if the graphic helps clarify construct-irrelevant information, then it may not be necessary—perhaps better to change the construct-irrelevant information. But, if the graphic helps to clarify the context or content that is construct-relevant or an operation related to the construct, then it may be necessary; otherwise, the graphic may be misleading or distracting. <u>Note</u>: Certain graphics are required/assessed in mathematics.</li> <li>▪ <u>Consideration</u>: Can the graphic accurately represent the complexity of the problem in its totality? If not, then the graphic may be misleading.</li> <li>▪ If the problem has a number of operations/steps, then it is important to simplify structures of the item (e.g., bulleted list with context or graphic, diagram that accurately reflects the problem in its totality).</li> <li>▪ Graphics allow for reduction of language and/or complexity of language.</li> <li>▪ Graphic needs to be consistent with the key elements of the item.</li> </ul>
<p><b>Item Vocabulary/Wording</b></p> <ul style="list-style-type: none"> <li>▪ High frequency words</li> <li>▪ Common and familiar words</li> <li>▪ Relevant technical terms; technical terms defined as appropriate</li> <li>▪ No ambiguous words or unnecessary words with multiple meanings</li> </ul>	<ul style="list-style-type: none"> <li>▪ Should be consistent with expectations/intention of the content standards—technical words/phrases okay as specified in standards</li> <li>▪ <u>Consideration</u>: Be mindful of precision of language—appropriate language does not simply mean “common” or “familiar”—must consider standards and Academic English Language (AEL)</li> </ul>

<b>Original</b>	<b>Refinement</b>
<ul style="list-style-type: none"> <li>▪ No irregularly spelled words</li> <li>▪ Use of formal proper names are relevant and appropriate</li> <li>▪ Naming conventions are consistent with graphics/stimuli</li> </ul>	<ul style="list-style-type: none"> <li>▪ Careful selection of vocabulary/phrases can result in simplification of sentence structures</li> <li>▪ Repeat key words/phrases needed by student to understand item and respond to item—providing synonyms for a key word may not always be helpful, given length and/or context of item; sometimes repeating the same key word is more appropriate (keep in mind the difference between instructional and assessment settings)</li> <li>▪ Consistent use of words/phrases within the context of the item—(also consider consistency within a strand—e.g., Reading, Measurement)</li> <li>▪ Support with context-familiar content-based abbreviations; make explicit connections between terms/abbreviations</li> <li>▪ Avoid words that are both nouns and verbs (e.g., carpet, value, cost); however, if a choice needs to be made, then the tendency is to use the word as a noun</li> <li>▪ Avoid hyphenated and compound words</li> <li>▪ Avoid gerunds</li> <li>▪ <u>Consideration</u>: Balance the amount and complexity of language with the amount of information necessary for student to understand/access the item (economy of language with meaning—purposeful use of language)</li> <li>▪ Relative pronouns (e.g., which) should have a referent (e.g., which expression, which adjective)</li> <li>▪ Construct-irrelevant vocabulary/phrases are at or below grade level</li> </ul>
<p><b>Item Sentence Structure</b></p> <ul style="list-style-type: none"> <li>▪ Referents are clear; noun-pronoun relationships are clear; antecedent references are made clear</li> <li>▪ Grammatical structures are clear Typically, <ul style="list-style-type: none"> <li>▪ past or future tense verb forms are changed to present</li> <li>▪ passive verb forms are</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>▪ Changing tense may help remove passive voice construction</li> <li>▪ Identifying the agent (e.g., proper noun) helps remove passive voice constructions</li> <li>▪ Verb should follow the subject (subject and verb should be adjacent to each other)—use familiar, common construction</li> <li>▪ One sentence per idea for complex items helps reduce inappropriate complexity of sentence structure (e.g., could use bulleted lists)</li> </ul>



Original	Refinement
<p>changed to active verb forms</p> <ul style="list-style-type: none"> <li>▪ complex structures are changed to subject-verb-object structure</li> <li>▪ long nominals/names/phrases are shortened (e.g., "last year's class vice-president" becomes "a student leader")</li> <li>▪ compound sentences are replaced with two separate sentences, especially in comparative structures</li> <li>▪ long prepositional phrases are reduced or removed</li> <li>▪ conditional clauses are replaced with separate sentences or the ordering of a clause within a sentence is changed for clarity</li> <li>▪ relative clauses are removed or rephrased for clarity</li> <li>▪ Questions framed in negative terms are rephrased</li> </ul>	<ul style="list-style-type: none"> <li>▪ Introductory phrases are removed (e.g., last week)—unless necessary for the item</li> <li>▪ Key information is presented upfront (first/early in item) and typically in simple sentence structure</li> <li>▪ Proper nouns should be ones that students are familiar with <i>reading</i></li> <li>▪ Complexity of sentence structure should be at or below grade level</li> <li>▪ Use traditional constructions—e.g., _'s for possessive; _s or _es for plural</li> </ul>
<p><b>Item Format/Style</b></p> <ul style="list-style-type: none"> <li>▪ Parts of the item/question are clear; distinctions are relevant and appropriate</li> <li>▪ Long problem statements are segmented or shortened</li> <li>▪ The order of operations is explicit</li> </ul>	<ul style="list-style-type: none"> <li>▪ Recommend first presenting text that introduces the graphic, followed by the graphic, and finally the stem</li> <li>▪ Formatting for emphasis of key words/terms (e.g., <b>bold</b>, ALL CAPS, <u>underline</u>)—focus on highly construct-relevant words/terms to call attention of ELLs to the word/term</li> <li>▪ <u>Consideration</u>: Is a block of text (a paragraph) appropriate for the context, construct? Blocks of text may be appropriate, depending on the construct assessed, the complexity of the information needed by the student to respond to the item, and the centrality of the context to the construct. May consider helping the language learner process the text with strategies that include: <ul style="list-style-type: none"> <li>▪ Bulleted lists</li> <li>▪ Off-setting key information</li> </ul> </li> </ul>

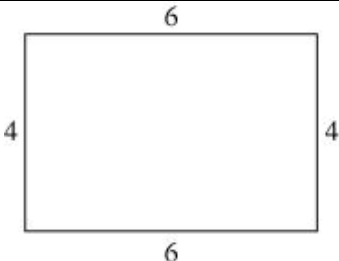
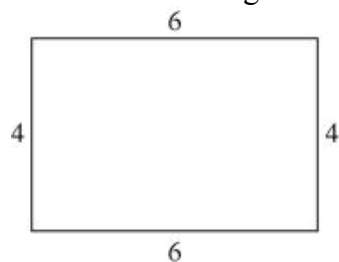
<b>Original</b>	<b>Refinement</b>
	<ul style="list-style-type: none"><li>▪ Emphasis of key words/terms</li><li>▪ Graphics</li><li>▪ Simple sentence construction; familiar context; familiar construction of words/verbs; etc. (refer to strategies presented previously)</li></ul>

[From: Sato & Rabinowitz (forthcoming-b)]

**Table 3: Examples of Original and Linguistically Modified Items**

<b>Strand</b>	<b>Construct</b>	<b>Original Item</b>	<b>Linguistically Modified Item<sup>15</sup></b>	<b>Notes</b>
Numbers/ Number Operations	Rounding estimation	Alba needed to know about how much the sum of 19.6, 23.8, and 38.4 is. She correctly rounded each of these numbers to the nearest whole number. What three numbers did she use?  A 19, 23, 38 B 19, 24, 38 C 20, 24, 38 D 20, 24, 39	Look at the numbers below. 19.6 23.8 38.4 Which list shows each number rounded to the nearest whole number?  A 19 23 38 B 19 24 38 C 20 24 38 D 20 24 39	Remove extraneous, potentially misleading information Change verb tense (present) Less complex sentence structure Off-set key information
Numbers/ Number Operations	Calculating cost	Kate bought a game for \$14.95, a book for \$5.85, and a hat for \$9.70. If the sales tax on these items is 6 percent and all 3 items are taxable, what is the total amount she must pay for the 3 items, including tax?  A \$32.33 B \$32.06 C \$30.56 D \$30.50	Kate buys the three items listed below. Game: \$14.95 Book: \$5.85 Hat: \$9.70 Kate pays 6% sales tax on the items. What is the total cost of the three items, including tax?  A \$32.33 B \$32.06 C \$30.56 D \$30.50	Change verb tense (present) Less complex sentence structure Off-set key information

<sup>15</sup> The appropriateness of the linguistically modified item should be judged vis-à-vis the assessed construct and the specific standard assessed. See *User's Guide*, Step III-C in particular.

Strand	Construct	Original Item	Linguistically Modified Item <sup>12</sup>	Notes
Measurement	Conversion	<p>If Jill is driving at 65 miles per hour, what is her approximate speed in kilometers per hour? (1 mile <math>\approx</math> 1.6 kilometers)</p> <p>A 16 B 41 C 104 D 173</p>	<p>65 miles per hour <math>\approx</math> _____ kilometers per hour (1 mile <math>\approx</math> 1.6 kilometers)</p> <p>A 16 B 41 C 104 D 173</p>	<p>Remove complex sentence structure Less complex structure of stem Off-set key information</p>
Measurement	Finding area	 <p>Which of the following numerical expressions gives the area of the rectangle above?</p> <p>A <math>4 \times 6</math> B <math>4 + 6</math> C <math>2(4 \times 6)</math> D <math>2(4 + 6)</math></p>	<p>Look at the rectangle below.</p>  <p>Which expression describes the area of the rectangle?</p> <p>A <math>4 \times 6</math> B <math>4 + 6</math> C <math>2(4 \times 6)</math> D <math>2(4 + 6)</math></p>	<p>Explicit directions/orientation to problem Less complex sentence construction</p>

[From: Sato & Rabinowitz (forthcoming-b)]

<sup>12</sup> The appropriateness of the linguistically modified item should be judged vis-à-vis the assessed construct and the specific standard assessed. See *User's Guide*, Step III-C in particular.

