Using Standards and Empirical Evidence to Develop Academic English Proficiency Test Items in Reading

CSE Technical Report 664

Alison L. Bailey, Robin Stevens, Frances A. Butler,
Becky Huang, and Judy N. Miyoshi
CRESST/University of California, Los Angeles

December 2005

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

Project 4.1 Developing Measures of Academic English Language Proficiency
Alison Bailey, Project Director

Copyright © 2005 The Regents of the University of California

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002-02, as administered by the Institute of Education Sciences (IES), U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Institute of Education Sciences, or the U.S. Department of Education.

USING STANDARDS AND EMPIRICAL EVIDENCE TO DEVELOP ACADEMIC ENGLISH PROFICIENCY TEST ITEMS IN READING[1]

Alison L. Bailey, Robin Stevens, Frances A. Butler,

Becky Huang, and Judy N. Miyoshi

CRESST/University of California, Los Angeles

Abstract

The work we report focuses on utilizing linguistic profiles of mathematics, science and social studies textbook selections for the creation of reading test specifications. Once we determined that a text and associated tasks fit within the parameters established in Butler et al. (2004), they underwent both internal and external review by language experts and content-area teachers. The external review provided data based on background questionnaires, text and item reviews used to judge representative aspects of topics and linguistic characteristics, and group interviews. Based on this information, the texts were either retained or rejected and items were retained, rejected or reserved for future modification. In the future, retained texts and items can be further analyzed for fit with empirically established text profiles.

Part I: Introduction

As specified in the abstract, the purpose of this report is to apply the information acquired from comprehensive linguistic analyses of fifth-grade texts previously conducted (Butler, Bailey, Stevens, Huang, & Lord, 2004) to the development of standards-informed academic language items. The work described

[1] Acknowledgments: We would like to thank the following for their role in the preparation of this work: the teachers who took part in the review and discussion of the texts and reading items developed here, administrative assistance from Soo Dennison and Morgan Joeck at the early stages of the work, Joan Herman for valuable feedback on an earlier draft of this report, and Fred Moss and Wade Contreras for the final formatting.

in this report focuses on utilizing the linguistic profiles of mathematics, science and social studies textbook selections for creating test specifications including guidelines for text selection and task/item writing. Once we determined that a text and associated tasks fit into the parameters established in Butler et al. (2004), they underwent both internal and external review by language experts and content-area teachers. The external review process was multifaceted. It provided data based on background questionnaires, text and item review forms that addressed whether the texts and items were topically and linguistically representative of the types teachers typically use in their classrooms, and group interviews. Based on this feedback the texts were retained or rejected and items were retained, rejected or reserved for future modification. In the future, retained texts and items can be further analyzed for fit with empirically established text profiles.

Overview

This report is presented in five sections, which together provide the background context for the assessment effort and the procedures followed in drafting academic English proficiency test items. Part I, Introduction, briefly sketches the empirical sources that led to the current work and provides the motivation for adopting a standards-informed approach to test development. Part II, Text Selection, describes the method followed in identifying appropriate texts across content areas on which to base the items. Part III, Item Development, lays out the steps in drafting items based on the texts selected and described in the previous section. Part IV, External Teacher Review, discusses the method and results of the external teacher review of texts and items, and the importance of this information in overall process. Finally, Part V, Conclusion and Recommendations, provides commentary on the successes and shortcomings of the current work and suggests a further step in the test development process, as well as expansion of the standards-informed approach to additional language modalities.

Background Context

With few assessments of language explicitly designed to measure English learners' (EL) knowledge of the language demands of the school setting (Bailey & Butler, 2002/2003; 2004), the National Center for Research on Evaluation, Standards and Student Testing (CRESST) Academic English Language Proficiency (AELP) project has systematically investigated this linguistic construct from a number of

different perspectives over the last 7 years. These include: (a) performance differences between native English speakers and ELs on standardized content assessments, (b) language demands of standardized content assessments in English, (c) relationships between language assessed on a language proficiency reading subtest and language used on standardized content assessments, (d) observation of teacher classroom talk, (e) analysis of state and national content-area standards, and (f) analysis of the linguistic features of textbook selections (e.g., Bailey, 2000; Bailey & Butler, 2002/2003, Bailey, Butler, LaFramenta, & Ong, 2001/2004; Butler & Castellon-Wellington, 2000; Butler & Stevens, 1997; Butler, Lord, Stevens, Borrego, & Bailey, 2003/2004; Stevens, Butler, & Castellon-Wellington, 2000). The most recent efforts focused on describing the linguistic features of mathematics, science and social studies textbooks that students must be able to read in fifth-grade (Butler et al. 2004). The results of these analyses are important for being able to generate tasks and items that measure the language necessary for students to benefit from instruction in mainstream classrooms.

Collectively this work has provided sufficient detail with which to develop a systematic approach to selecting content-area texts on linguistic criteria. The texts differed in type by content-area. Mathematics texts were comprised of word problems, whereas science and social studies texts were expository. We restricted the text types we selected because these types were the basis for the linguistic profiles of texts devised in earlier analyses (Butler et al. 2004). Using other variations of text types (i.e., those reliant on graphics, directions for conducting science experiments, etc.) might require additional research in order to make empirically-based claims about representing an academic language construct in text selections used in assessment.

The work presented in the current report is just one discrete stage in an ideally comprehensive test development process that involves several stages. Most test development efforts begin either with a formal needs analysis of some kind or grow out of the recognition that there is a need to access a particular skill or ability. The next stage involves articulating the construct—in this case, broadly stated, the academic language used in fifth-grade mathematics, science, and social studies texts. Once the construct is articulated or as it is being articulated, test developers working with content specialists begin to determine how best to access the construct within operational parameters. Specifications are drafted. Potential items and tasks are produced and, at the earliest stages, should be tried out to determine the efficacy of

the formats. At every stage, the process is iterative (or should be) with the construct, specifications, and items and tasks revised and refined. Even after a test becomes operational, the process should be ongoing as data become available (see Bailey & Butler, 2002/2003, pp. 22-30, for a more complete discussion of the test development process).

The role that a principled text selection and item development approach as well as adequate reviews (internal and external) play is vital to many aspects of the test development process. These aspects include construct validity (i.e., that we measure the school language that students are actually intended to know and use), technical adequacy (i.e., development of items that discriminate between low and high abilities), reliability (i.e., internal consistency, range of variability, inter-rater reliability) and efficiency or economy of item development (i.e., forestall the creation of too many "bad" items at an early stage of the test development process before the items are subject to potentially costly field testing).

We have deliberately chosen to call our test development approach a *standards-informed* rather than a *standards-based* approach because it is not solely reliant on the use of standards descriptors to inform text selection and item development.[2] Rather, the *standards-informed* approach integrates empirical evidence of the language encountered in school contexts (e.g., linguistic characteristics of textbooks, language demands of classroom discourse, etc.) with the tasks students are expected to master as they are reflected in English Language Development (ELD) and content-area standards. In the current work, we focus on the reading modality and integrate the results of linguistic analyses of textbooks at the fifth-grade with fifth-grade ELD, mathematics, science and social studies standards. Future test development efforts for different modalities and grade levels can model the same approach, but include different types of relevant empirical evidence. For example, a test of listening and speaking should include empirical evidence of the linguistic characteristics of classroom input from teachers and the language demands or expectations made on students' oral production, (see Bailey et al., 2001/2004 for examples of this type of empirical evidence).

[2] We see no reason why a *standards-based* approach could not be adopted in the future once standards themselves are subjected to validity studies in terms of content coverage, difficulty levels, sequencing, etc., and revised accordingly.

The work reported here utilizes California state standards for ELD, mathematics, science and social studies.[3] These standards offer no inherent advantage in the test development process we adopted—merely they were the standards aligned to the textbooks we had previously analyzed (Butler et al., 2004). Consequently, the California standards serve as an example application of the standards-informed approach. The same approach can be adopted with other state and national standards. Indeed some state standards offer far more specificity than those of California in terms of well-developed performance indicators associated with their standards that may make the language demands across subject areas more transparent to test developers (Bailey & Butler, 2002/2003).

Part II: Text Selection

The first phase of test construction for academic reading tasks is to select appropriate texts on which to base item and task development. In this section, we first explain the procedures that were followed for selecting texts, outline the criteria used to make the selections, present the texts selected, and discuss issues that arose while using the standards and applying the text selection criteria.

Procedures

There were three steps in our text selection process: (a) standards and text passage selection, (b) internal review of text selections, and (c) descriptive analyses. The first step was the initial text selection phase, during which potential texts were identified for each California state content-area standard across the three content areas—mathematics, science, and social studies (California State Board of Education, 1998, 1999, 2000). The second step consisted of an internal review of the texts identified during the first stage which required verification that the texts selected were aligned with the standards, that the linguistic and content-area criteria were met, and that the texts did not have any bias issues. The third step consisted of establishing a fit with the basic linguistic profiles of texts generated by descriptive analyses of Butler et al. (2004). The procedures for the three steps are discussed next.

[3] English Language Arts (ELA) has not yet been a focus of the AELP project at CRESST and thus empirical analysis of the linguistic characteristics of ELA textbooks is not available, using the standards-informed test development approach at this time.

Step 1

During initial text selection, two researchers independently reviewed the California standards for each content area and selected one indicator per standard to use as a topic guide when choosing texts. Each content area of the California state content standards contains multiple standards; within each standard there are one or more of what we refer to in this report as *indicators.* These indicators are statements that provide detailed information about the content standards and specify what students should know or be able to do to demonstrate mastery of the standard. Examples for content standards and corresponding indicators from mathematics, science, and social studies are provided in Figure 1.

Many indicators are complex, often containing multiple ideas, concepts, and/or topics. This became an important factor during text selection, since rather than attempting to select a text that covers every idea set forth in the indicator, we decided to focus on one or two main ideas or concepts to simplify the text selection process. The indicator for the social studies standard in Figure 1 is an example of a complex indicator with multiple concepts/topics.

Additionally, some standards are primarily fact-oriented. We avoided using those standards because they often lend themselves to the creation of discrete items that focus on individual facts about the content and thus do not provide a rich context for assessing language proficiency. For example, in the following social studies standard, emphasis is placed on memorization of states and state capitals.

> Students know the location of the current 50 states and the names of their capitals (California State Board of Education, 1998, p. 20).

When using the standards and selecting indicators, we focused on those standards and indicators that would enable us to choose the types of texts desired. In the current effort, the text types to be selected were word problems for mathematics and expository passages for science and social studies. In the following science standard, the texts would most likely consist of a series of steps or procedures for conducting experiments. Although procedural texts are important in science, this particular type was not included in the current item development effort.

> Scientific progress is made by asking meaningful questions and conducting careful investigations. As a basis of understanding this concept and addressing the content in the other three strands, students should develop their own questions and perform investigations… (California State Board of Education, 2000, p. 16).

After selecting indicators for each standard, two researchers then used a set of guidelines and criteria to choose texts aligned to the standards. Each researcher selected two texts per standard from California-approved mathematics, science, and social studies textbooks. Table 1 shows the textbooks used for each content area.

Mathematics Standard:

2.0 Students perform calculations and solve problems involving addition, subtraction, and simple multiplication and division of fractions and decimals.

Sample of a Corresponding Indicator:

2.4 Understand the concept of multiplication and division of fractions.

Science Standard:

1.0 Elements and their combinations account for all the varied types of matter in the world.

Sample of a Corresponding Indicator:

b. *Students know* all matter is made of atoms, which may combine to form molecules.

Social Studies Standard:

5.2 Students trace the routes of early explorers and describe the early explorations of the Americas.

Sample of a Corresponding Indicator:

1. Describe the entrepreneurial characteristics of early explorers (e.g., Christopher Columbus, Francisco Vásquez de Coronado) and the technological developments that made sea exploration by latitude and longitude possible (e.g., compass, sextant, astrolabe, seaworthy ships, chronometers, gunpowder).

Figure 1. Samples of Fifth-Grade California Standards and Indicators for Mathematics, Science, and Social Studies.[4]

[4] Examples drawn from California State Board of Education (1998, 1999, 2000).

Table 1

Fifth-Grade Textbooks Used in the Text Selection Process

Publisher	Content Area		
	Mathematics	Science	Social Studies
Harcourt	Math (2002) National Edition	Science (2000) California Edition	Social Studies: Early United States (2002) National Edition
Houghton Mifflin	Mathematics (2002) California Edition	Science (2000) California Edition	Social Studies: America Will Be (1999) National Edition
McGraw Hill	Math Explorations and Applications (2003) National Edition	Science (2000) California Edition	United States: Adventure in Time and Place (2001) National Edition

Three textbooks per content area were used to ensure a broad range of texts from which to select for each standard and corresponding indicator. After the initial text selections were made, the texts were screened again during the Step 2 internal review described here.

Step 2

During this stage, three different researchers reviewed the texts selected during Step 1. First the texts were reviewed using the guidelines and checklist developed for use in Step 1 (see Appendix A for the Text Selection Checklist). If the selections were deemed acceptable, the researchers checked them again for bias (i.e., themes or language—racial, religious, or ethical—that may be considered offensive or emotionally upsetting to any group.) If a selection was deemed unacceptable for any reason, it was reserved as an alternate text or eliminated completely if there were serious issues with the text (e.g., a problem with bias or a text with an unusual number of specialized vocabulary words).

Step 3

After the texts were screened and approved, they were prepared as electronic files, which were then used for analysis of the basic descriptive characteristics of the texts (i.e., number of words, number of sentences, mean number of words per sentence, and number of paragraphs). These analyses helped ensure that the texts were typical according to parameters established in Butler et al., (2003/2004). Further, the analyses ensured that the texts are not only appropriate for language test development in terms of alignment with the standards, but also that they are free from bias and fit the basic linguistic profile of texts for each content area and grade level (Butler et al., 2004).

After these analyses were run, the results were checked against the content-area text profiles mentioned previously (see Appendix B for the text profiles). If the results fell into the established parameters, they then underwent the external teacher review described in Part IV. A description of the guidelines and criteria that were used for initial text selection, the internal review, and descriptive analyses follows.

General Guidelines for Text Selection

Selecting content-area texts for language test development can be a difficult endeavor. It is important to choose texts that do not require extensive background knowledge, as the goal is to assess language not content-area knowledge. Therefore, the texts must provide enough information about a topic so that language items and tasks can be designed around the information contained in the text. In any language test, students should not be tested on how much they already know about a topic, but rather on how well they are able to understand and manipulate language to achieve linguistic goals, such as recognizing a comparison between two elements or summarizing the content of a paragraph. Some students will have been exposed to topics and concepts that others may not have been exposed to, but by assuring that all the information needed to complete a task or item is contained in the text, dependence on prior knowledge is kept to a minimum. This is a type of test bias to be aware of while selecting texts, creating items, and reviewing the results of pilot tests in order to create a fair, reliable, and valid assessment.

Additionally, because the current goal is to assess general academic language proficiency (e.g., *analyze, hypothesis*), it is important to avoid texts that are heavily laden with specialized content-area vocabulary (e.g., *igneous*, *penal*), unless the

words are either defined in the text or can be understood from context. Complex content-area texts that contain many specialized academic words may be more appropriate for assessing content-specific language proficiency, rather than a student's ability with general academic uses of language that cut across a number of different disciplines. The distinction between specialized and general academic vocabulary has become a major hallmark of the academic language construct in the work of a number of different language researchers (e.g., Bailey & Butler, 2002/2003; Martin, 1976; Nation & Coxhead, 2001; Scarcella & Zimmerman, 1998; Stevens, Butler, & Castellon-Wellington, 2000).

To facilitate the systematic and consistent review of potential texts, a set of general guidelines for selecting texts was developed in conjunction with content-specific criteria. The rationale used when creating these guidelines and criteria included: basing guidelines on good testing practices (e.g., it is important to reduce linguistic and cultural bias when developing language assessments); establishing guidelines and criteria with a foundation in prior empirical research that provided information about the linguistic features of texts (i.e., the texts must be similar to the types of texts we used in the prior research in order to make claims about the "typicalness" of a text); and reducing need for extensive background knowledge of any particular subject or topic, again a type of "good testing practice" when selecting texts for any type of language assessment. The guidelines and criteria are listed in the two following sections.

There are seven guidelines researchers considered when selecting texts. Guidelines 1 and 2 were already presented earlier in this document within the description of the procedures for Step 1. The first two guidelines are:

Guideline 1

Select standards or indicators that are not excessively fact-oriented.

Using a standard that focuses on factual knowledge may make choosing a linguistically-rich text more difficult (e.g., a text that provides explanations vs. a text that states facts).

Guideline 2

Select standards and indicators that lend themselves to specific types of text.

As shown in the sample standards discussed previously, some standards and indicators lend themselves to, and in some cases even require, the use of specific types of texts (e.g., directions for experiments or procedures).

Guideline 3

Choose texts that do not require additional support for student understanding.

In this phase of test development, text selections should be able to "stand alone" and should not require the use of supplementary graphics or teacher assistance. However, brief references to experiments and/or other page numbers in the textbook did not disqualify a text. In the first following example, texts aligned with this indicator would probably involve the use of a map and would therefore be unacceptable. In the second example, the excerpt includes a reference to a previous text, which is acceptable because the selection does not require information from the prior text to understand the material.

> *Example 1:* Locate on maps of North and South America land claimed by Spain, France, England, Portugal, the Netherlands, Sweden, and Russia (California State Board of Education, 1998, p. 17).

> *Example 2:* As you may remember, 97 percent of all the water on Earth is salt water found in earth's oceans. Although all of Earth's water supports life, not all of it is safe for humans to drink… (Badders et al., 2000, pp. D12-D13)

Guideline 4

Choose texts that provide general and/or introductory information.

Texts of this type help reduce the potential for a lack of background knowledge interfering with assessment of language proficiency. Texts should contain enough information to develop a variety of language assessment questions, all of which can be answered by reading the text. In texts that introduce new material for example, there is a tendency to define or contextualize new vocabulary, which helps the reader. It also helps "level the playing field" for students who may not have studied a particular concept yet. In the following example provided, not only is the text aligned to the indicator, it also provides introductory information and defines new vocabulary in context.

Science Indicator: Students know the sequential steps of digestion and the roles of teeth and the mouth, esophagus, stomach, small intestine, large intestine, and colon in the function of the digestive system (California State Board of Education, 2000, p. 15).

Text excerpt aligned to the standard/indicator: …Two other organs have a role in digestion. The *liver* produces bile, which is stored in the *gallbladde*r until it's needed. Bile breaks down fats into smaller particles that can be more easily digested. The *pancreas* produces a fluid that neutralizes stomach acid and chemicals that help finish digestion (Frank et al., 2000, p. A19).

Guideline 5

Avoid texts that are conceptually dense.

Texts that discuss multiple contexts without giving an explanation of the concepts increase the potential need for supplemental materials or teacher assistance. In the following example, the text excerpt is aligned to the standard, but the content is too conceptually dense for language assessment purposes because it assumes knowledge of chemistry.

Science Indicator: Students know properties of solid, liquid, and gaseous substances, such as sugar ($C_6H_{12}O_6$), water (H_2O), helium (He), oxygen (O_2), nitrogen (N_2), and carbon dioxide (CO_2) (California State Board of Education, 2000, p. 14).

Text excerpt aligned to standard/indicator: …When compounds react, they change, and form new products. For example, hydrochloric acid contains hydrogen and chlorine atoms. It reacts with sodium hydroxide (a base), which contains sodium, oxygen, and hydrogen atoms. The products are sodium chloride, which contains sodium and chlorine atoms, and water, which contains hydrogen and oxygen atoms. Water and salts are often the products of reactions between acids and bases (Frank et al., 2000, pp. C50-51).

Guideline 6

Select topics and texts that provide opportunity for reader engagement.

Some topics, such as food and sports, may be more interesting and engaging to students because they are more relevant to students' lives. The following text example discusses a topic familiar to students from many different cultures.

Ice cream is a mixture. There are many flavors of ice cream, and each contains different ingredients. Mixtures, including the various flavors of ice cream, don't have chemical formulas… (Badders et al., 2000, p. C49).

Guideline 7

Avoid topics and texts that are potentially offensive or upsetting to students (e.g., prejudicial statements, bias toward one gender or another, or ethical issues such as the death penalty).

Texts describing murder, capital punishment, or dissection may be inappropriate because of their controversial nature for some groups and should be avoided in language assessments since the focus should be on language and not content. In the following example indicator, slavery is the main topic. Although it is an important part of American history, it may also introduce some sensitivity issues and therefore the selection of texts aligned to this indicator was avoided.

> Describe the introduction of slavery into America, the responses of slave families to their condition, the ongoing struggle between proponents of slavery, and the gradual institutionalization of slavery in the South (California State Board of Education, 1998, p. 18).

In addition to the general guidelines described previously, we also used a set of content-specific criteria to judge whether a text might be appropriate or not for use in test development. These criteria are described in the following section.

Content-specific Selection Criteria

Mathematics

For mathematics we selected word problems of two or more sentences in length. All consist of a real-life scenario used to set up the mathematical problem. None contains references to visuals. We chose scenarios that are of a more general or everyday nature, such as topics related to food or eating. The following is an example of a word problem based on an everyday topic:

> You and your aunt are planning to take a train to visit City Museum. Your aunt will pick you up at 10:30 am. It takes 20 minutes to drive to the North Conway train station. You must be home by 5:00 pm. How long can you stay at the museum? (Maletsky et al., 2002, p. 486)

We attempted to avoid selections with unusual or unfamiliar topics and those that contain unusual or low-frequency vocabulary. Low-frequency vocabulary may be general or academic in nature, like the word *earthquake*, but may not be a part of

the lexicon for students from certain regions. Although knowledge of unusual or low-frequency vocabulary may not impede student ability to solve a mathematics problem, it may act as a distraction, especially to ELs. The following example consists of a scenario with an unusual topic and low-frequency vocabulary (e.g., stilt walking):

> In 1980, a man walked 3,008 mi on stilts from Los Angeles to Bowen, Kentucky. The trip took 158 days. In 1891, a stilt walker traveled from Paris, France, to Moscow, Russia, going 1,830 mi in about 54 days. Who traveled faster? (Greenes et al., 2002, p. 156)

Science

For science we selected multi-paragraph expository passages of approximately five to eight paragraphs in length that are of an introductory or general nature. They may be composed of an explanation or description of a process or concept with organizational features such as *exemplification*, *labeling*, and *definitions*. The following excerpt provides a good example of text that introduces a science concept:

> Scientists have classified plants into two main groups. Vascular plants, such as ferns and trees, have tubes. Because they have tubes to carry water and nutrients, vascular plants can grow quite tall.
>
> Nonvascular plants, such as mosses, do not have tubes. So water must move from cell to cell. These plants need to live in a moist place, and they do not grow to be very large. (Frank et al., 2000, p. A53)

Science texts may include many new and unfamiliar vocabulary words. To the extent possible, we chose texts that define, explain, or paraphrase specialized or technical vocabulary. For example, if the life cycle of a maple tree is used as a means of exemplifying sexual reproduction in plants, it must be clear that: (a) a maple is a tree and (b) that the maple tree is used in the text for the purpose of exemplifying. The following excerpt defines two new terms in context. However, it does assume some knowledge of content-area vocabulary, such as *forecast*, *station models*, and *weather systems*, which may be acceptable as long as items are not designed to measure background knowledge of these concepts.

> Scientists usually forecast the weather using a *synoptic weather map*. This type of map shows a summary of the weather using station models. By comparing maps made every six hours, scientists can tell how weather systems are moving. They then use this information to predict what the weather will look like hours later.

> If you look at weather records to see what happened in the past, you can find patterns. *Statistical forecasting* is based on finding weather patterns.
>
> For example, suppose you notice that the wind has just started blowing from the west. Past records show that 75 out of the last 100 times the wind blew from the west, your weather became clearer and colder. What weather prediction would you make? (Moyer et al., 2001, p. 156)

Social Studies

For social studies, we selected multi-paragraph expository texts of five to eight paragraphs in length. As with science, social studies texts may include many new and unfamiliar vocabulary words. We sought texts that define, explain, or paraphrase such words. Social studies texts also contain a large number of proper nouns. In general we avoided texts that use many decontextualized proper nouns. The following excerpt contains proper nouns, although for the most part they are contextualized in the selection or explained. This selection also contains a reference to text outside the passage (the example is italicized). A brief reference such as this was considered acceptable, since it is typical of texts in social studies.

> Early in 1847 another group of religious followers headed west. More than 14,000 travelers left from Nauvoo, Illinois, along a route called the Mormon Trail. The trail got its name from the people who traveled it. They were part of a religious group called the Church of Jesus Christ of Latter-day Saints, or Mormons. Because of mistreatment in Nauvoo, the Mormons decided to move west. *On the Infographic on pages 430-431, you can see that the Mormon Trail closely follows the Oregon Trail through the Great Plains.*
>
> The Mormon leader Brigham Young led the long march. The Mormons' wagons crossed ice-covered rivers as they followed the Oregon Trail. Near Fort Bridger the group left the trail and headed south into lands claimed by Mexico. Finally, in July 1847, the first group of Mormons reached a large lake now known as the Great Salt Lake. The region's inhabitants included the Ute (YOOT) and Shoshone peoples. The present-day state of Utah got its name from the Ute. (Banks et al., 2001, p. 429)

Although the use of primary sources is common in social studies, at this time, we avoided texts that contain extended primary source excerpts, such as poems, historical documents, or a paragraph from a biography. Quotations of one to two sentences in length, however, were considered appropriate, since they are also typical of social studies texts at this grade level. The following excerpt includes an example of an acceptable primary source quotation:

> In letters to her family, Narcissa Whitman described the rich farmland, dense woodlands, and mountains. "It is indeed, a lovely situation," Whitman wrote. Her letters were later published and helped encourage Americans to settle in Oregon. (Banks et al., 2001, p. 429)

Taken together, the general guidelines and content-specific criteria described previously provided a foundation upon which researchers could make informed text selections for the purpose of test development. We turn now to a discussion of the texts selected for external review.

Texts Selected

A total of 11 mathematics word problems, 6 science passages, and 5 social studies passages were approved for external review and item development purposes via the procedures described in the previous sections (e.g., Steps 1 and 2 of the text selection process). We present three tables (Tables 2-4), one per content area, and each with a brief description justifying the selections.

Mathematics

Table 2 provides information about the 11 word problems selected. Included in the table are selection titles, the standard and indicator aligned to each selection, descriptive statistics, and source information.

Table 2

Texts Selected from Grade 5 Mathematics Textbooks

Selection Topic	California Standard (Indicator)	Descriptive Statistics				Source and Publisher
		Total Word Count	No. of Sentences	Avg. No. of Words Per Sentence	No. of Paragraphs	
"Walking 70,000 miles"	Number Sense 1.0 (1.1)	33	2	16.50	1	*Math* © 2002, Harcourt, p. 205
"Carl went to the fair"	Number Sense 2.0 (2.1)	48	6	8.00	1	*Math* © 2002, Harcourt, p. 59
"Lemonade sales"	Number Sense 2.0 (2.1)	44	3	14.67	1	*Math* © 2002, Harcourt, p. 327
"Mary's lunch"	Number Sense 2.0 (2.1)	35	4	8.75	1	*Math* © 2002, Harcourt, p. 71
"Camping trip"	Number Sense 2.0 (2.3)	43	4	10.75	1	*Math* © 2002, Harcourt, p. 318
"Stiltwalker"	Number Sense 2.0 (2.3)	42	4	10.50	1	Mathematics © 2002, Houghton Mifflin, p. 156
"Puppet necklace"	Algebra and Functions 1.0 (1.2)	48	4	12.00	1	*Mathematics* © 2002, Houghton Mifflin, p. 425
"Birthday"	Algebra & Fractions 1.0 (1.3)	29	3	9.67	1	*Mathematics* © 2002, Houghton Mifflin, p. 113
"Traffic light"	Statistics, Data Analysis, and Probability 1.0 (1.1)	88	5	17.60	1	*Math Explorations and Applications* © 2003, McGraw Hill, p. 313
"Books"	Mathematical Reasoning 1.0 (1.1)	39	6	6.50	1	*Mathematics* © 2002, Houghton Mifflin, p. 561
"Hot dogs & soda"	Mathematical Reasoning 2.0 (2.5)	212	15	14.13	2	*Math Explorations and Applications* © 2003, McGraw Hill, p. 225

Selections include a range of topics, some more familiar to fifth graders than others. For example, Selections 2-5, 8, 10, and 11 are considered to be everyday topics for most children (e.g., lunch, birthdays, and books). Some selections are based on less-familiar topics, e.g., puppet necklace (7) and traffic light (9). However, we felt that it was important to include these word problems since they met the other criteria and would be reviewed by classroom teachers prior to being tried out in classrooms.

The selections range in length from 29-212 words, with an average of 60 words per word problem (see Appendix C for the descriptive statistics for each content area). Two selections are quite a bit longer than the others; number 9 (88 words) and number 11 (212 words). Excluding the longest word problem, number 11, drops the range to a more typical 29-88 words with an average length of 45 words. Overall, the average sentence length for this set of word problems is slightly longer than the norm for fifth-grade at 12 words (average is about 11), with a range of 7-18 words per sentence. Four of the 11 word problems contain more than 14 words per sentence, contributing to the longer average sentence length.

All of the selections except one (again, number 11) are composed of one paragraph, which is typical for word problems at this grade level. The average number of sentences per paragraph is slightly above normal at 4 sentences, as opposed to the typical number of 3. The paragraphs range in length from 2-15 sentences. Excluding selection 11, the range is only 2-6 sentences, which is a typical range.

Some of the word problems selected, then, are not considered typical in terms of sentence length, number of sentences per word problem, and number of paragraphs. However, as mentioned in the previous section, choosing "typical" word problems would mean restricting most selections to 2-3 sentences in length. Shorter word problems contain less content upon which to base items. Therefore, we felt it was important to include a few that were atypical in the attempt to include word problems that contain the full range of grammatical, lexical, and organizational features typical for the content area.

Science

Six passages were approved for external review across a range of topics, all of which introduce major concepts that are aligned to standards, such as the digestive system and the solar system. Information about the passages is provided in Table 3.

Table 3

Texts Selected from Grade 5 Science Textbooks

Selection Topic	California Standard (Indicator)	Descriptive Statistics				Source and Publisher
		Total Word Count	No. of Sentences	Avg. No. of Words Per Sentence	No. of Paragraphs	
"The Makeup of a Mixture"	5.1. Physical science (F)	407	30	13.57	7	*Science* © 2000, Houghton Mifflin, p. C49-C50
"Using Physical and Chemical Properties"	5.1. Physical science (F)	419	28	14.96	6	*Science* © 2000, Harcourt, p. C24-C25
"The Digestive System"	5.2. Life science (C)	259	20	12.95	5	*Science* © 2000, Harcourt, p. A19
"Water in the Air"	5.3. Earth science: Water (C)	335	20	16.75	6	*Science* © 2000, Harcourt, p. B15-B16
"Using Water Wisely"	5.3. Earth science: Water (D)	524	34	15.41	8	*Science* © 2000, Houghton Mifflin, p. D47-D49
"How is the Moon Different from Earth?"	5.5. Earth science (B)	246	22	11.18	5	*Science* © 2000, McGraw Hill, p. 410

The average length of the science selections is 365 words, which is shorter than the selections analyzed in Butler et al., (2004). However, these selections are of a more appropriate length for developing reading proficiency items than the longer selections used for research purposes, namely linguistic analysis. The variation in the lengths of the selections is intentional (246-524 word range), as a range of lengths is needed to address different levels of reading proficiency within one assessment. Therefore, the within paragraph statistics are more important in this case; they should reflect similarities to the paragraph-level statistics in the linguistic profile.

There is an average of 26 sentences per selection, with an average sentence length of 14 words, just slightly longer than typical (13). Due to the shorter total length, the average number of paragraphs per selection is also smaller (6), but the average number of sentences per paragraph is typical of the content area at 4 sentences each. Overall, these selections exhibit many of the features that are regarded as typical, e.g., organizational features such as *comparison* and *description*, despite the fact that they are shorter in length.

Social Studies

A total of 5 texts were selected from social studies textbooks, with a range of topics including native peoples of North America, early American life, and famous Americans. These are shown in Table 4.

Table 4

Texts Selected from Grade 5 Social Studies Textbooks

Selection Topic	California Standard (Indicator)[a]	Descriptive Statistics				Source and Publisher
		Total Word Count	No. of Sentences	Avg. No. of Words Per Sentence	No. of Paragraphs	
"The Tlingit"	5.1 (2)	543	41	13.24	8	*United States: Adventure in Time and Place* © 2001, McGraw Hill, p. 84-85
"The French in North America"	5.3 (2)	490	29	16.90	7	*Social Studies: Early United States* © 2002, Harcourt, p. 166-170
"New England Towns"	5.4 (5 and 7)	525	35	15.00	10	*Social Studies: Early United States* © 2002, Harcourt, p. 229-230
"The Life of a Leader"	5.5 (4)	442	32	13.81	7	*United States: Adventure in Time and Place* © 2001, McGraw Hill, p. 322-323
"Women and the War"	5.6 (3)	306	21	14.57	6	*Social Studies: Early United States* © 2002, Harcourt, p. 302-304

[a]The California Social Studies Standards are numbered but do not have subtitles like those of mathematics and science.

As with science, these selections are shorter in total word length, with an average of 461 words each, and also vary in length (306-543 words). This was necessary due to the purpose for which these texts were being selected, i.e., test development. Longer passages are inappropriate for reading assessment, especially at the fifth-grade level. They each take too long to read and require a greater attention span on the part of the students.

There are an average of 32 sentences per selection and 15 words per sentence, slightly higher than the typical sentence length in Butler et al., (2004) of 14 words per sentence. There are fewer paragraphs, due to the shorter length of the selections; however, the number of sentences per paragraph is typical of social studies at 4 sentences per paragraph. Thus, overall, the social studies selections are strong candidates for test development purposes, at least based on descriptive/basic linguistic characteristics.

Summary

On the whole, the selections in each content area exhibit some deviation from the profiles of typical texts established in Butler et al., (2004). However, this should not be considered problematic because the purpose for making the selections differed slightly from the earlier research. In Butler et al., texts were selected for research purposes and length was not restricted. In the current work, texts could not be too long because they are intended for use in tests of second language reading proficiency, as pointed out previously. A range of text lengths is needed to measure varying levels of reading proficiency within one assessment, especially at this grade level. Long texts are more time consuming and require longer attention spans and the ability to integrate material across larger amounts of read text. On the other hand, because background or general knowledge across students varies, shorter texts allows for a greater number of texts that can vary in topic to be more "fair."

In addition, in our earlier research we noted some topic-related variation; therefore, some of the variation we found in this set of texts may be in part due to topic-related differences. Despite this variation, the ranges for sentence length, number of paragraphs per selection, and number of sentences per paragraph for each content area are smaller than established in earlier research, indicating a greater amount of consistency within the selections as a group. Taken together with the alignment of the texts to the standards and the generality and introductory

nature of the selections, the selections hold real potential for use in test development.

Issues Confronted

In general, it was difficult to select content-area texts aligned to the standards because the standards necessarily are focused on content and not language. This selection process was further complicated by the interaction between content knowledge and the language ability of students. In light of research that suggests background knowledge can play a critical role in reading comprehension performance (e.g., Garcia, 1991), it was important to keep in mind that we were selecting texts linguistically typical of the content area that contained content aligned to the standards. Yet we tried to make selections that would not unfairly advantage students who had broader background knowledge. Our purpose in language assessment is to measure student understanding of the language of written instruction. Thus, once texts have been approved for use in item development, the key is to ensure that any associated language assessment items can be answered solely by reading the text or by student general (language) knowledge and not content-area knowledge.

Several specific issues were raised during the text selection process that warrant brief discussion here. Most of the issues revolved around the use of standards as a basis for making text selections, although there were a few issues with using the guidelines and criteria. These are discussed next, in turn.

Standards

In general, it was more difficult to select texts aligned to the mathematics standards than science and social studies. Mathematics performance standards and indicators tend to be very general and abstract, making it difficult to identify specific word problems that are aligned to the standards. Some mathematics standards predispose the use of graphics (e.g., histograms, circle graphs, etc.), so it is difficult to select texts that conform to the standards and the text format being used in this stage of research, i.e., word problems. In other cases, the standards and associated indicators obviate the possibility of texts, since they focus more on the output students must produce rather than input in the form of word problems. For example, one of the standards for measurement and geometry requires students to "identify, describe, and classify the properties of, and the relationships between,

plane and solid geometric figures" (California State Board of Education, 1999, p. 22). Two of the three indicators under this particular standard thus require students to measure or draw various geometric figures by using tools. This limited the choice of indicators for some standards.

Last, some of the mathematics standards and indicators either make it impossible to align texts to them or require more than a single word problem to address the standard/indicator, because they are too conceptually-based or call for mathematical reasoning skills. For example, one standard states "students use strategies, skills, and concepts in finding solutions" (California State Board of Education, 1999, p. 23). While this standard might be used to select certain types of texts from mathematics textbooks, it is difficult to identify word problems that specifically address it. In another standard, students "move beyond a particular problem by generalizing to other situations" (California State Board of Education, p. 23). This standard clearly would call for more than one word problem to address it adequately. Indeed, the three indicators for the standard actually represent a set of hierarchical concepts that move the student toward the development of generalization abilities. While this is an important skill with language ramifications, selecting standards-aligned word problems is difficult at best.

Regarding the use of science standards and indicators, these also present difficulties in terms of deriving the selection of texts for use in language assessment. Most science standards are focused on concepts, all of which tend to include many specialized content-area words, even in the standards and indicators. The extensive use of specialized vocabulary is important to avoid when making selections for the purpose of general language proficiency, since the use of such vocabulary assumes a greater breadth of content-area knowledge from the students.

Additionally, some science indicators, as with mathematics, predispose the types of texts used, e.g., standards related to investigation and experimentation indicate the use of procedures and steps in the texts. These types of standards and indicators may be better utilized when selecting texts for different types of assessments, such as performance-based tests.

There were two key problems with social studies standards; mainly that they include many potentially controversial topics and also many proper nouns, both of which we had tried to avoid per the general guidelines and content-specific criteria

we had established. The issues related to the guidelines and criteria are discussed in more detail here.

Guidelines and Criteria

In any effort to systematically select texts, there should be a uniform set of guidelines and criteria upon which to base the selections. As mentioned previously, we developed our guidelines and criteria on the basis of the research conducted in Butler et al., (2003/2004) and Butler et al. (2004). Overall they were useful in making the text selections we needed, however, some modifications would be recommended for future use.

First, the general guidelines apply somewhat differently across the content areas. This suggests that the general guidelines might be more useful if each guideline was adjusted to fit the different content areas more closely and then added to the content-area criteria, eliminating the use of general guidelines completely. For example, in science, many standards and indicators are focused on what students should "know," making them fact-oriented and violating our first general guideline. Perhaps the guideline should be less about whether the standard or indicator is fact-oriented and more dependent on simply finding a text that is aligned with the standard. Just because the standard or indicator focuses on factual knowledge does not mean that the texts will be unsuitable. This also applies to the second guideline, which was to select indicators that lent themselves to the type of text needed. The content of the standards and indicators does not necessarily preclude a text type. To keep with the goal of aligning text selection to standards, it is important to review texts aligned to the standards before deciding whether or not a standard will be able to "produce" the needed text type.

The guidelines stipulating the avoidance of texts that require additional support and the selection of texts with introductory information are less applicable to mathematics than to science and social studies. Word problems usually consist of self-contained scenarios that have little to do with mathematics concepts and so they are not conceptually dense or introductory in the way that science and social studies texts are (i.e., word problems are not focused on introducing concepts and ideas). Although word problems are associated with concepts being taught, since we are not assessing mathematics concepts this is not an issue unless the problem contains an undue number of specialized mathematics concept words.

Regarding the use of visuals with science and social studies texts, science texts contain many difficult concepts that are frequently illustrated as they are discussed, making it difficult to select texts without visuals. Future efforts may need to include texts with visuals, as these may be more typical for the content area at this grade level. With social studies texts, there are more serious issues with prior content-area knowledge, such as geography terms and proper names. Most social studies texts assume some basic knowledge and often use the names of people and places without providing context.

Regarding the content-area criteria, when selecting mathematics word problems we found that most word problems are only two sentences in length, which is appropriate according to our criteria. However, short word problems often do not contain enough language upon which to base the development of items and tasks. Therefore, we attempted to select longer word problems so that more items could be developed for each word problem. A better alternative though, for future efforts, may be to think of math word problems in a different way when developing tests and items. An example would be the development of vocabulary and cloze grammar items based on word problems instead of items assessing more complex language, or perhaps the use of problems with charts or graphs to expand the types of items that can be created.

Vocabulary was a major confounding issue when selecting texts. The content-area criteria were oriented toward avoiding texts that contain many specialized words, while acknowledging that it is normal for texts to contain some. In actuality, it was hard to avoid texts that made extensive use of these words. In science, vocabulary is inextricably linked to the concepts being taught, so the texts contain many challenging words that are difficult to grasp, even when defined in the text. In social studies, historical events and content-area concepts, such as economic terminology, are the content of study and thus also contain terms closely tied to the standards. While we do not suggest changing the way texts were selected in this regard, it is important to note that vocabulary will continue to be an issue no matter how the guidelines and criteria are written. Selecting vocabulary for assessment and designing items and tasks based on the content-area texts will require much caution, as content and language are so closely intertwined.

Overall, most of the problems we had associated with using the standards, general guidelines, and content-area criteria can be addressed through revisions to the guidelines and criteria and in some cases by simply adding more examples or

even modifying the types of texts being selected. There were a few other problems with text selection not mentioned previously, but they were minor. For example, although it was not the norm, the content of some indicators could not be identified in any of the textbooks. There were also a few differences between the textbooks, but this varied by content area. For example, one of the textbooks contained fewer word problems than the others, so a greater number of the selections were made from the other two textbooks. Some textbooks match their unit titles with the standards better than others, which makes it easier to select standards-based texts from those textbooks.

Taken together, these observations are useful for future efforts because they prepare the text selection team for problems they may encounter; however, many of the points made previously are typical of text selection issues with any test development effort and often impact the actual development of items and tasks as researchers move forward. A discussion of the item development efforts based on these texts follows.

Part III: Item Development

Developing an assessment is comprised of a series of iterative steps designed to produce the most reliable, meaningful assessment possible. Three components of test development are discussed in this report: texts used as the basis for item development, item formats, and test content. Texts are selected that represent the types of texts students must be able to read in English-only content-area classes (see Part II for a discussion of the text selection process). Research is conducted to determine the most appropriate item formats to use in the test being developed. Empirical research and standards are used to determine test content. Taken together, these three components are important parts of an assessment framework, which guides the development of specifications for the types of texts to use, the item formats to create, and the content of the assessment.

In this section of the report, we first describe the steps we took to determine the item formats we would use and what the content of the draft items would be. We present a sample assessment framework based on this research and sample specifications for three different item types.

Analysis of Item Formats

While we were in the process of selecting the texts, we were simultaneously analyzing what the best item formats would be to use for this test development effort. Ideally, we would review the full range of items and tasks used not only in content-area textbooks written in English, but also on classroom-based assessments and standardized tests. The purpose for doing this is to develop academic language proficiency items based on authentic item and task formats used in mainstream English-only classrooms. EL students must be provided opportunities to take tests that contain the same types of items, thus giving them the test-taking experience they will need to cope later with tests of content-area knowledge written for fluent English speakers. Naturally, there is going to be variation in the item formats depending on context: textbooks contain more projects and performance-related tasks, while standardized tests are dominated by multiple-choice items, because these allow for the rapid scoring of large numbers of tests at one time. In the current effort, we focused on the item formats used in textbooks, since textbook language has been the focus of our research over the last two years.

To systematically identify the item formats frequently used in each content area, we surveyed two fifth-grade California-approved textbooks (see Table 1) for science and social studies and identified the item formats that are used to measure student reading comprehension and conceptual understanding. We did not review the item formats in mathematics, since mathematics items and tasks generally fall into two categories: mathematical computation without a language component and word problems that end with either a statement or 'wh' question. Two chapters were randomly selected from each of the textbooks. We reviewed the chapters, compiling lists of the item formats and also noting the relative frequency with which each type occurred.

From these lists, we separated the item formats into categories based on their prompt (or stem) and response formats. There are many different types of items and tasks used in the textbooks, ranging from full-scale projects and experiments to simple 'wh' questions with multiple-choice response options. Frequently occurring types in both science and social studies include (a) open-ended 'wh' questions, which call for a word, phrase, fragment, or sentence-length response; (b) sentence completion items where students use words from the passage or a word bank to complete the sentence; (c) cloze items (a paragraph or passage with words omitted

intentionally) where students use a word bank to complete the blanks; (d) items that require students to complete graphic organizers such as a table or chart; and (e) 'wh' questions with multiple-choice response options. Multiple-choice stems are often statements that require students to complete the sentence with one of the multiple-choice options.

On the basis of our review we narrowed down to 10 the list of item formats that are common to both content areas, eliminating performance type items that depend on writing skills, since our current efforts are focused on the assessment of academic reading proficiency. These item formats are presented in Table 5.

Table 5

Frequently Used Item Formats in Science and Social Studies Textbooks

Item type no.	Prompt/stem format	Response format
1	'Wh' questions	Answer with a word, phrase, sentence fragment, or list
		Multiple-choice options that are sentence fragments
		Multiple-choice options that are complete sentences
2	Sentence completion/statement	Fill in blank with word from text
		Fill in blank with word/phrase from word/phrase bank
		Multiple-choice options that are single words
3	Cloze paragraph	Fill in blank with word from word bank
4	Sequencing (prompt is an imperative statement)	Fill in blanks next to each sentence with a number from 1-x or a letter from a-x
5	Matching (prompt is an imperative statement)	Fill in blank with letter of corresponding word or phrase
6	Graphic organizer (prompt is an imperative statement)	Fill in spaces in a diagram, chart, or table using words or phrases from a word bank or from a text

These item formats will become a part of the assessment framework discussed in the following sections. We turn now to a discussion of the test content.

Test Content

Test content is ideally based on a combination of sources relevant to the test taker's needs or goals. In this case, test material must reflect the types of reading tasks students are assigned in mainstream English classrooms at the fifth grade, and the test content must assess the student's ability to make meaning out of the reading materials. As discussed earlier, one important type of reading material students must understand in the classroom is textbook language, which is one reason we have focused on the use of content-area textbook selections here.

To determine which reading skills and abilities to assess, we have largely drawn from our empirical findings on the linguistic characteristics of texts in fifth-grade textbooks (Butler et al., 2003/2004; Butler et al., 2004). Predominant characteristics of texts identified across content areas are considered critical linguistic features that students must demonstrate mastery over and are considered candidates for a test of general academic language proficiency. In Butler et al., (2003/2004), the content for sample task specifications was divided into three main categories: reading skills, language functions with embedded grammatical features, and vocabulary (see Table 6).

The subcategories reflect the empirical findings in that they are some of the most frequently used language features at the fifth-grade level in the materials analyzed (e.g., the use of logical connectors was determined to be an important linguistic cue for students to understand an explanation).

In Butler et al., (2004), a sample content framework for developing academic language proficiency items was derived from the research. It is organized a little differently, presenting potential test content in three categories: vocabulary, grammar, and organization of text (see Table 7).

Table 6

Content for Task Specifications

Content Category
Reading skills
(1) Identify main idea
(2) Locate supporting details
Language functions (with embedded grammatical features)
(1) Comparison/contrast
(a) Adverbial comparatives
(b) Comparative adjective forms
(c) Logical connectors
(2) Description
(a) Logical connectors
(b) Nominal structures
(c) Passive voice
(d) Prepositions
(e) Simple present tense
(f) Subordinate clauses
(3) Explanation
(a) Logical connectors
Vocabulary
(1) Identify meaning in context
(2) Draw meaning from embedded definition(s)

Note. From An Approach to Operationalizing Academic Language for Language Test Development Purposes: Evidence from Fifth-Grade Science and Math, by F. A. Butler, C. Lord, R. Stevens, M. Borrego, and A. L. Bailey, 2003/2004, CSE report # 626. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Copyright 2004 by CRESST. Reprinted with permission.

Table 7

Content Framework for Developing an Assessment of Academic Language Proficiency

Content Category	Mathematics	Science	Social Studies
Vocabulary			
Clause connectors	√	√	√
Non-academic vocabulary			
Academic vocabulary (AV)			
General AV (high-frequency)	√	√	√
Specialized AV (defined in context)	--	√	√
Measurement words	√	√	--
Proper nouns	--	--	√
Grammar			
Nominalizations	--	√	√
Noun phrases	√	√	√
Participial modifiers	--	√	√
Passive forms	--	√	√
Prepositional phrases	√	√	√
Organization of Text			
Comparison	√	√	√
Definition	--	√	√
Description	√	√	√
Enumeration	√	√	√
Exemplification	--	√	√
Explanation	--	√	√
Labeling	--	√	√
Paraphrase	√	√	√
Scenario	√	--	--
Sequencing	√	√	√

Note. From *Academic English in Fifth-grade Mathematics Science, and Social Studies Textbooks* (p. 110), by F. A. Butler, A. L. Bailey, R. Stevens, B. Huang, and C. Lord, 2004, CSE report # 642. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Copyright 2004 by CRESST. Reprinted with permission.

Table 7 also shows which features are more predominant in each content area, as indicated by the check marks. No attention, however, is given to global reading skills in this table, such as skimming or scanning for supporting details.

In addition to examining empirical evidence from prior CRESST research, we also reviewed the California English Language Development Standards (1999) for reading (the Grades 3-5 cluster) to determine if the empirical evidence mentioned previously is similar to the ELD content described in the standards. The ELD standards for reading are divided into four main categories: (a) word analysis, (b) fluency and systematic vocabulary development, (c) reading comprehension, and (d) literary response and analysis. We found that the California ELD Standards are more general than the empirical research cited previously and also include frequent references to speaking and listening tasks, which we would not be able to apply when developing a general assessment of academic English *reading* proficiency. Reading standards that include aspects of speaking and listening are appropriate for integrated skills assessments (e.g., reading and speaking) and classroom-based assessments (e.g., teacher-created assessments). The standards in Table 8 are examples of those that are related to the types of potential test content contained in Tables 6 and 7.

Table 8

Examples of California English Language Development Standards (1999)[5]

CA ELD standard	Example
Word Analysis, Decoding and Word Recognition, Advanced ELD level	Apply knowledge of word relationships, such as roots and affixes to derive meaning from literature and texts in content areas (p. 34).
Fluency and Systematic Vocabulary Development, Vocabulary and Concept Development, Early advanced ELD level	Recognize that some words have multiple meanings (e.g., *present/gift*, *present/time*) in literature and texts in content areas (p. 43).
Reading Comprehension, Structural Features of Informational Materials, Advanced ELD level	Identify significant structural (organizational) patterns in text, such as compare and contrast, sequential and chronological order, and cause and effect (p. 57).
Literary Response and Analysis, Narrative Analysis of Grade-Level-Appropriate Text, Intermediate ELD level	Apply knowledge of language to derive meaning from literary texts and comprehend them (p. 63).

[5] We are including an example of a Literary Response and Analysis standard for the purpose of showing how empirical research and standards can be linked, even though the ELA subject area is not, as already mentioned, included in the current efforts.

Synthesizing evidence from these three sources, we constructed a more detailed assessment content framework that focuses on reading content area texts in mathematics, science, and social studies at the fifth-grade level. This new framework is provided in Table 9.

Using this framework, test developers can isolate different linguistic features and skills and use this information to develop item specifications. For example, we can create item specifications that target the assessment of general reading comprehension, such as: the ability to distinguish the central point of a text or to understand the problem statement in a word problem; the ability to understand the organizational features of texts (e.g., why a particular quote is used); or the ability to make meaning out of grammatical cues embedded within the organizational features of a text (e.g., know that logical connectors such as *because* often signal that a causal relationship is being established). Specifications can also be designed that focus on specific vocabulary knowledge and skills, such as general vocabulary knowledge (e.g., core general vocabulary needed to process grade-level texts) or the ability to use context to guess the meaning of unfamiliar words or idiomatic phrases.

After creating the framework, we then created item specifications for each item type shown in Table 5 and wrote items based on the item specifications and texts selected (see Tables 2-4). A total of 59 items was created for six of the selected mathematics texts (18 items), four science texts (23), and three social studies texts (18). Table 10 shows the distribution of each item type across the content areas.

Table 9

Content Framework for Developing Academic English Reading Proficiency Items

	Mathematics	Science	Social Studies
Overall comprehension of a text			
Understand central point/topic	--	√	√
Understand structure/organization	--	√	√
Understand the problem statement	√	--	--
Understand the author's purpose	--	√	√
Organization of Text & Grammatical Features			
Comparison/contrast	√	√	√
Understand supporting details			
Understand comparative constructions (e.g., comparative adjective forms, superlatives)			
Definition	--	√	√
Understand definitions stated via use of paraphrase			
Description	√	√	√
Scan for supporting details			
Exemplification	√	√	√
Understand enumerated examples			
Explanation	--	√	√
Understand cause and effect using grammatical cues (logical connectors)			
Make inferences			
Labeling	--	√	√
Understand descriptions of processes that end with the labeling of a term			
Understand supporting details and word forms			
Quotation	--	√	√
Understand the purpose for and use of quotes in a text			
Understand supporting details			
Sequencing	√	√	√
Understand a sequence of events or a process using grammatical cues (logical connectors)			
Vocabulary			
Know high-frequency general vocabulary (e.g., articles, copula verb, prepositions, conjunctions, adjectives)	√	√	√
Know high-frequency general academic vocabulary (e.g., verb forms, adverbs, comparative adverbial phrases)	√	√	√
Use contextual cues to guess the meaning of vocabulary	√	√	√
Understand definitions provided in the text (general vocabulary, general academic, and specialized academic)	√	√	√

Table 10

Distribution of Item Types by Content Area (Total no. of items = 59)[6]

Item type no.	Math	Science	Social studies
1a	7	3	3
1b	5	4	1
1c	1		3
2a		7	3
2b		4	
2c	1	1	2
3		1	
4	1	1	1
5			4
6	3	2	1
Total no. of items	18	23	18

The item types are fairly evenly distributed across the three content areas, with some differences attributable to the types of texts used in mathematics, science, and social studies. Mathematics texts are short, for example, so fewer items and item types can be created on the basis of each text.

One sample text per content area is provided in Appendix D, along with the items written to correspond with the text. Item specifications for three of the different item types included in these example texts and item sequences are presented here.

Sample Item Specification No. 1

The following sample specification shows which framework category the specification fits into (see Table 9), and also gives detailed information about the task format and stimulus and response attributes. This first specification is for a traditional multiple-choice item type.

[6] Please refer to Table 5 for the description of the 10 item types.

Sample Item Specification #1

Framework category: Overall Comprehension of a Text

General description and text type

Students will identify the problem statement in a mathematics word problem.

Task format

'Wh' question with multiple-choice sentence options.

Stimulus attributes

A mathematics word problem generally of 2-3 sentences in length with a problem question or imperative statement at the end.

Response attributes

Circle the correct multiple-choice option from the four options provided.

Item notes

The student must read the text, read the question and response options, and then circle the option that is the correct answer.

Sample Task #1–Overall comprehension of a text: Understand the problem statement

On a weekend camping trip, Ken, Eric, and their dad went for a walk on the Appalachian Trail. The first hour, they walked 3/8 mile. The second hour, they walked 4/5 mile. About how many miles did the boys and their dad walk? (Maletsky et al., 2001, p. 318)

What is this math problem asking about?

a) How long the family walked.
b) How far the family walked.
c) How many people walked.
d) How many hours the boys walked.

Key: B

Figure 2. Sample Item Specification #1.

Sample Item Specification No. 2

Again, the sample specification provides item writers detail about which framework category the specification fits into, the task format, and stimulus and response attributes. The following specification describes a sequencing task.

Sample Item Specification #2

Framework category: Organization of Text & Grammatical Features

General description and text type

Students will put a series of events described in a multi-paragraph expository text into the correct order.

Task format

Sentence sequencing

Stimulus attributes

A multi-paragraph expository text usually of 3-5 paragraphs in length.

Response attributes

The stimulus is followed by a list of sentences with blanks beside each. The first answer is given as an example. Students fill in the blanks with numbers, sequencing the events from first to last.

Item notes

The sentences are paraphrased statements drawn from the text. Students must read the text and then revisit the text to ensure they put the events in the order that they occurred. They number the list of events from one to six.

Sample Task #2 – Organization of text and grammatical features: Sequencing

[see Appendix D for the text]

Put the six sentences in the order in which the events occurred. The first one is done for you.

1	George Washington was born.
_____	His troops won an important battle.
_____	He became an elected official.
_____	He married his wife.
_____	He joined the military.
_____	He worked as a surveyor.

Key:

1	George Washington was born.
6	His troops won an important battle.
4	He became an elected official.
5	He married his wife.
3	He joined the military.
2	He worked as a surveyor.

Figure 3. Sample Item Specification #2.

Sample Item Specification No. 3

This last specification is for a type of sentence completion item. This type of sentence completion item provides test takers with a word bank from which to select the correct word. A more difficult version of this item type would not include a word bank.

Sample Item Specification #3

Framework category: Vocabulary

General description and text type

Students will complete sentences using words that are defined in a multi-paragraph expository text.

Task format

Sentence completion using words from a word bank.

Stimulus attributes

A multi-paragraph expository text usually consisting of 3-5 paragraphs.

Response attributes

The stimulus is followed by incomplete sentences and a word bank (list of words). Students complete each sentence by filling in the blank with the correct word from the word bank.

Item notes

The sentences are paraphrased statements drawn from definitions in the text. Students must read the text and then revisit the text to ensure they choose the correct word to complete each sentence.

Sample Task #3 – Vocabulary: Understand definitions provided in the text

[see Appendix D for the text]

Complete each sentence with one of the words in the list. Each word can be used only once.

saliva gastric juice bile pancreas

The _____ is an organ that helps the body complete digestion.

Your mouth makes a fluid called _____.

A fluid that helps the body break down fats is _____.

Your stomach makes _____ to help you digest proteins.

Key: pancreas, saliva, bile, gastric juice

Figure 4. Sample Item Specification #3.

Response Options in the Academic Language Testing Context

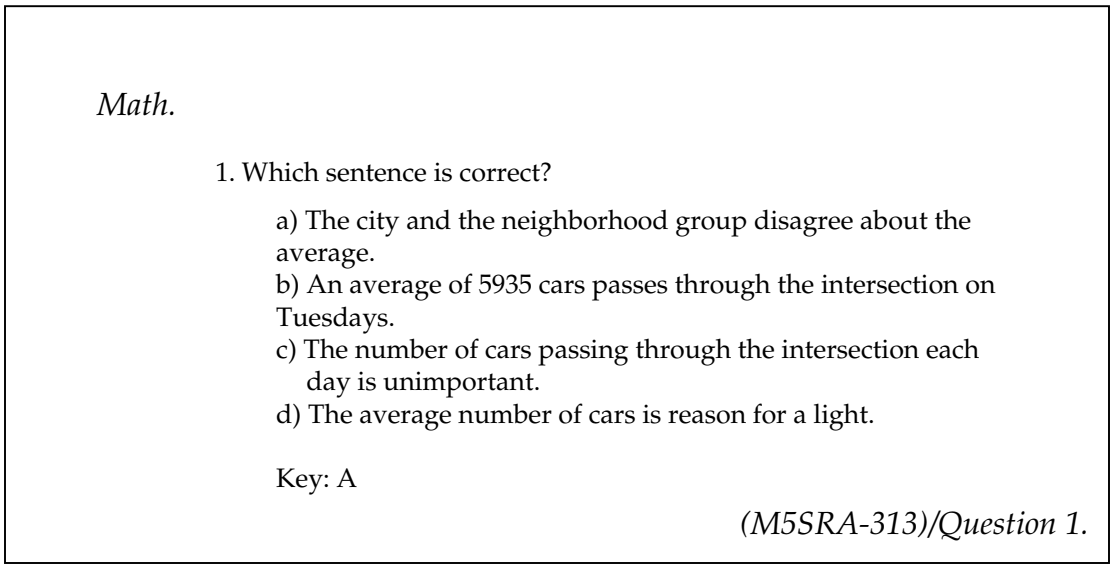
In this section, we pay closer attention to the linguistic characteristics and manipulations we made to the correct and distractor options in the reading tasks. In particular, we argue why we believe the response options we have created provide useful examples for future test development in the area of academic language proficiency assessment.

We focus primarily on task types that involve the selection of the correct response option from among several options because these most clearly demonstrate the deliberate linguistic manipulations that are possible in tasks aimed at assessing ELD in the academic context. Previous research suggests that the optimal number of response options in the case of multiple-choice questions (MCQ) is three or four (e.g., Kehoe, 1995). We devised at least four options for every MCQ in a task sequence. General guidelines found in the testing literature include, making certain that the intended answer is correct or clearly best, making all alternatives grammatically consistent with the stem of the item and parallel in form, avoiding verbal clues that might enable students to select the correct answer or to eliminate an incorrect alternative, making the distractors plausible and attractive to the uninformed, varying the length of the correct answer to eliminate length as a clue, and avoiding the use of general alternatives such as "all of the above," and "none of the above." (e.g., Haladyna, 1994; Gronlund, 1998).

Following these guidelines, the options we created were intended to be clear and unambiguous, and our distractor options were created to be plausible to students who are unsure of the answer. We also tried to make distractors comparable in length, grammatical form, and level of difficulty to the answer key to avoid any biased cueing or extraneous clues. Despite the ease of construction they offered, we refrained from using "all of the above," "none of the above," or options of this kind because test-takers might perceive those options as "fillers" used by uninformed test makers (Osterlind, 1998), and because they posed missed opportunities to manipulate language in order to determine which linguistic features students relied too much or too little upon in distractors with more substance.

Given that the tasks are designed to assess students' reading ability in English rather than content area knowledge, in the following discussions of example items for each subject matter, we first describe the linguistic knowledge required of

students to identify the correct option in a test item. We then demonstrate the linguistic characteristics and cognitive demands of each of the distractors in the test item.



Math.

1. Which sentence is correct?

a) The city and the neighborhood group disagree about the average.
b) An average of 5935 cars passes through the intersection on Tuesdays.
c) The number of cars passing through the intersection each day is unimportant.
d) The average number of cars is reason for a light.

Key: A

(M5SRA-313)/Question 1.

Figure 5. Math stem question.

The stem question asks students to pick out the correct sentence from the four options. It is a single clearly formulated question and stated without ambiguity (Gronlund, 1998). Identification of the correct response option requires the vocabulary knowledge of "disagree," which serves as the key word in the sentence. Students must not only be able to understand the lexical meaning of "disagree," but also process multiple sentences in the text passage that specifically describe the conflicting views between the city and the neighborhood group. The processing of those sentences requires multiple levels of linguistic knowledge, including lexical (e.g., "claim"), syntactical (e.g., the linguistic structure that depicts the relationship between two clauses), and functional (e.g., contrast and comparison between two sentences). To arrive at the correct answer, students must recognize the correspondence between the word "disagree" and the meaning conveyed by multiple sentences.

On the other hand, distractor options, need to be ruled out for students to ideally confirm their selection of the correct answer. Distractor *b* is designed to attract unsure students who use the strategy of matching phrases in the option and those available in the text. Specifically, the verbal phrase "pass through the

intersections" is used in the beginning of the text passage as well. In addition, students need to understand the meaning of "average" to differentiate between this option and the statement in the text passage that contains similar information ("…the following numbers of vehicles were counted…Tuesday, 5935…"). Distractor *c* assesses less the lexical knowledge but, like the answer key, requires the ability to process multiple sentences in the text passage to identify the critical issue under debate. Option *d* is a plausible distractor because it is a partially correct factual statement. However, it is incomplete and lacks specificity. That is, the statement is only correct for one of the two opposing camps described in the passage. Students need to carefully read all information on the opposing groups in the text passage to confidently disqualify this distractor.

Science.

2. Read this sentence from paragraph two of the passage:

The only weathering and erosion is <u>due to</u> the impact of rocks from space hitting the Moon's surface.

Which words below have the same meaning as the words underlined in the sentence? Circle the best answer.

a) caused by
b) a part of
c) similar to
d) in order to

Key: A

(S5MH-410)/Question 2

Figure 6. Science stem question.

This question taps directly into students' linguistic knowledge, and specifically, the knowledge of synonymous phrases. The stem question is clearly stated and the target point of assessment is well defined (i.e., "Which words below have *the same meaning* as the words underlined in the sentence?"). To answer this question correctly, students must first of all understand that the phrase "due to" signals

causal-effect relationship. Likewise, students must know that the answer "caused by" conveys the same meaning.

Ruling out the distractors might not be as easy a task as identifying the answer using linguistic knowledge of synonyms. In order to dismiss the competing options, students not only need to have grammatical knowledge but the ability to refer back to the text passage and check the option against the information therein. Specifically, distractors *b* and *d* are not grammatical options when replacing the phrase "due to" in the stem question. Option *d* is explicitly ungrammatical, whereas option *b* requires more sophisticated linguistic knowledge, i.e., the count-mass syntactic distinction. The fact that weathering and erosion are both mass nouns disqualifies option *b*, which contains a syntactically incompatible determiner "a" in the phrase. On the other hand, option *c* is not only grammatically correct, but also semantically plausible, which makes it a strong candidate for uninformed students. Rejection of this distractor relies on students' ability to understand the meaning of the passage and check the option against the correct information.

Social Studies.

3. How were Patriot women different from Loyalist women? Circle the best answer.

a) Patriot women wanted independence.
b) Loyalist women brought colonists food.
c) Patriot women and Loyalist women were the same.
d) Both types of women lived in the colonies.

Key: A

(SS5HB-302)/Question 3

Figure 7. Social studies stem question.

Answering the question requires the linguistic knowledge of synonyms and antonyms, as well as the ability to process the text to correctly identify the answer. The stem question itself uses a compare and contrast grammatical construction. Specifically, students must understand what the key phrase "different from" means

in order to move on to the next step—searching for the right answer out of four options. Distractors *c* and *d* can easily be ruled out if students understand the meanings of the critical vocabulary/phrases, that is, "the same" in option *c* and "both" in option *d*. The phrase "the same" in option *c* points out explicitly what this sentence is about: the similarity between Patriot women and Loyalist women, which is opposite to what the stem question asks about. Option *c*, on the other hand, uses the word "both," which signals the similarity between the two groups as well, but in a more implicit way compared to option *d*. To reject distractor *b* and confirm *a* as the correct answer, students must be able to comprehend the content of the passage.

Part IV: External Teacher Review

The overarching purpose of the external teacher review effort reported here was to determine whether our process and instruments for eliciting teacher feedback are sufficiently clear and useful for providing critical input for the text selection and task/item development stages of the test development process. Specifically, the goal was threefold: (a) to determine what teacher participants thought about each text and item sequence and to have their average responses and variation in responses compiled in one document, (b) to use the reviews to eliminate texts in principled ways from test development, and (c) to use the specific comments and suggestions provided by the teachers as feedback for rejecting or modifying items and item sequences.

The findings described in this section are based on four types of data generated during the teacher review sessions. These are: (a) questionnaire data on teacher and classroom demographics, as well as teacher responses to questions about the frequency of use of textbooks and assessments and typical teaching practices involving those materials, (b) teacher review of texts primarily in terms of language appropriateness for native English-speaking and EL students, (c) teacher review of tasks and items primarily in terms of language appropriateness for native English-speaking and EL students, and (d) group interviews using a focus group approach with teachers to elicit further feedback and extended discussion of the texts and associated tasks/items.

Method

Participants

Ten educators from five different Los Angeles county public school districts were invited to participate in the teacher review. These teachers were contacted based on prior nominations by district administrators who felt they would contribute to discussions of content area language and English language development. Of the 10 participants, 3 were Grade 4 teachers, 3 were Grade 5 teachers, 3 were Grade 6 teachers, and one was an ELD specialist, all with at least 5 years of teaching experience. Two of the invited teacher participants were male and the remaining 8 were female. Two of the sixth-grade teachers taught mathematics and science and one taught social studies and English language arts. One fourth-grade teacher did not teach science at her school. The remaining fourth- and fifth-grade teachers taught all content areas. The main criteria for teacher selection were that the majority of the teacher's students be native English speakers, Redesignated Fluent English Proficient (RFEP) and/or initially identified as Fluent English Proficient (FEP) students. This was done in order to gather information from teachers on what English speaking students are typically exposed to in the classroom, with an emphasis on gaining a clearer understanding of what students performing at grade level are expected to know and be able to handle. The ELD specialist was invited to participate to provide general feedback on specific issues concerning EL students.

Procedures

Participant teachers first completed a mail-in basic demographic questionnaire (see AELP Teacher Questionnaire Part I, Appendix E). They were then invited to CRESST/Center for the Study of Evaluation (CSE) on the University of California, Los Angeles (UCLA) campus to participate in a day-long task rating and focus group discussion. First, they were provided with texts and tasks from the fifth-grade level and, as applicable, from their content area to review. CRESST researchers provided the criteria for reviewing the texts and items (see Appendix F for copies of text and item review forms). After reading through and reviewing the tasks, and time permitting, the review facilitators facilitated informal small group discussions in which participants had the option to share aloud their review of the tasks. Finally, to capture grade and content level similarities and differences, a whole-group

discussion followed the small-group discussion, and included questions surrounding such topics as difficulty level of the tasks, grade appropriateness, and content relevancy.

Since our purpose is to explore teachers' response to the language of texts and discover how they react to language demands of textbooks and test items based on the texts, we decided to use focus group methodology for group discussion/interviews. As Morgan (1998) states, focus groups draw on three of the fundamental strengths that are shared by all qualitative methods: (a) exploration and discovery (i.e., focus groups are used to learn about topics or people who are poorly understood), (b) context and depth (or a way to better understand the background behind people's thoughts and experiences), and (c) interpretation (i.e., a way of understanding why things are the way they are and how they get to be that way). Therefore, this was the understanding and approach that we used for our focus group study. That is, through the focus group, we wanted to gain a better understanding of the context and depth in which teachers use textbooks and to learn about teacher practice in the classroom (e.g., testing) by content area and to a lesser degree, grade level. Finally, we viewed interpretation as a way to explore and to better understand the similarities and differences in textbook use and teacher practice across content areas and to some extent grade levels.

Specifically, the teacher reviews began with an introduction to the project (i.e., definition of academic language and issues in assessing academic language proficiency). This was followed by providing procedures and directions for selecting standards-based texts for task development to the whole group. Detailed directions for completing the text review forms were discussed prior to assignment to mathematics, science, and social studies working groups for the text review. The directions on the review forms read "*The purpose of this review is to determine if the texts selected are linguistically representative [typical] of texts (or tasks/items as applicable) used in content area classrooms with native English-speaking students.*" Teachers then participated in small-group discussions within their working groups. This was followed by whole-group discussion between the teachers and facilitators.

A similar process was planned for the task and item review session, culminating in a whole group discussion. However, due to the underestimation of the time required to complete the text review and task and item review, small group discussions were either very brief or not conducted at all in order for participants to have time to complete the review forms and participate in a full-group discussion.

Upon the completion of the task/item review sessions teachers were asked to complete the AELP Teacher Questionnaire Part II (see Appendix E).

Text and Item Review and Focus Group Discussions

Each mathematics, science, and social studies working group included a Grade 4 teacher, a Grade 5 teacher, and a Grade 6 teacher. Grade 6 teachers were placed in the working group for the content area(s) that they typically taught. The number of text selections each group reviewed varied across content area and review type (i.e., text or item). For the text review, the mathematics group reviewed 11 text selections, the science group reviewed six text selections, and the social studies group reviewed five text selections. The ELD specialist was given a subset of material from each of the three content areas and switched from group to group, as time permitted.

For the small-group focus group discussions that took place after the text review, review facilitators took notes after asking participants *"Is the text representative of texts you use in class?"* For the whole-group focus group discussions that concluded the text review, content-area groups were asked to share their reactions to the text selections for each respective content area.

For the item review, the mathematics group reviewed items for 6 of the 11 math problem selections (which contained 18 items), the science group reviewed items for four of the six text selections that they reviewed during the text review session (23 items), and the social studies group reviewed items from three of the text review selections (18 items). Again, the ELD specialist was given a subset of material from each of the three content areas and moved from group to group, as time permitted. However, due to the additional time required to complete the item review forms, small-group focus group discussions were not conducted. For the whole group discussion, content-area groups were asked to share their reaction to the items for their content area.

Instrumentation

AELP Teacher Questionnaire, Part I. The AELP Teacher Questionnaire, Part I consisted of three main sections. The first section included teacher name, school, and grade questions. The second section included classroom enrollment questions (i.e., number of students and numbers of Limited English Proficient [LEP], RFEP, and FEP students). The third section included informational questions about textbook

and supplemental material used in the classroom. Due to the nature of these questions, the AELP Teacher Questionnaire, Part I was mailed to participants 2 weeks prior to the AELP review sessions in order to give participants an opportunity to verify enrollment, textbook and supplemental material information that might not be readily accessible in venues other than in a participant's classroom or at a participant's school site.

AELP Teacher Questionnaire, Part II. The AELP Teacher Questionnaire, Part II consisted of three sections. The first section included questions about the grade level and content area(s) taught, and teaching experience of participants. The second section focused on frequency, grouping and format questions about textbooks, supplementary materials, reading and classroom assignments, and tasks for mathematics, science, social studies, and language arts.[7] The third section included questions about the frequency and types of items and tasks participants use for classroom instruction-based assessment in mathematics, science, social studies, and language arts. The AELP Teacher Questionnaire, Part II was administered on site during the text and task/item review sessions.

AELP Text Review Form. The AELP Text Review form was developed by CRESST to determine whether the texts selected by CRESST researchers are linguistically representative of texts typically used in content area classrooms with native English speaking students. One form was completed per text reviewed (Appendix G). The AELP Text Review form consisted of a series of questions about the texts selected by CRESST researchers. The first section included teacher name, school, and grade questions. The next section included a series of questions about the text under review. For example, the first two questions asked if the text was representative of textbooks and other materials used in the participant's classroom. Participants were then asked to rate the difficulty of the text for native English speakers and ELs. A 5-point Likert scale (*easy* to *difficult*) was used for this question. Participants were asked to list difficult and grammatical features deemed challenging to students directly onto the text for the range of English knowledge and abilities of each teacher's students. These lists were created separately for EL students and for both native English speakers and EL students. Questions asked participants to provide an explanation if they thought there were any sensitivity

[7] Language arts is included for completeness and to inform future test development efforts. Due to the focus of prior CRESST work, in the current study, we only had teachers review texts and items for mathematics, science, and social studies.

issues regarding the text, and the last item provided space for additional comments/questions.

AELP Item Review Form. The AELP Item Review Form was developed by CRESST to determine whether the items are representative of the types of items teachers use with native English-speaking students and ELs. One form was completed per item. Additionally, the item review form asked teachers to rate the difficulty of the task item using a 5-point Likert scale (*easy* to *difficult)*. Participants were asked to list vocabulary and grammatical features deemed challenging to their students for each item. The last question required participants to evaluate the items in terms of potential sensitivity issues (e.g., cultural bias, gender bias). Participants were also prompted for additional comments/questions.

Analytic Plan

Quantitative data from the questionnaires were entered into SPSS for analysis using descriptive statistics procedures. The open-ended responses to the questionnaire items were entered into Excel spreadsheets verbatim for use in illuminating the quantitative findings. Where applicable, responses to the text and item review forms were also quantified. Open-ended responses and those vocabulary and grammatical features in the text selections and items that teachers identified as challenging for native English-speaking students, for ELs, or for both, were typed verbatim into Excel spreadsheets. Notes from the small group and full-group discussions were typed up by the three review facilitators and open-coded for key themes that emerged from the teachers' comments. Those themes were used in illustrating or further elaborating the teachers' responses to the text and task/item review forms.

Results

Questionnaire Data

Teacher responses to the questionnaire Parts I and II are summarized in the following paragraphs. These responses are important for two reasons. First we will be able to describe the professional characteristics of the participating teachers. This provides us with some indication of how representative the teachers are in terms of teaching experience and in terms of the language status of students in their mainstream classrooms. Second, the teachers' responses will give us a more concrete

sense of the frequency and nature of reading assignments and content-area assessments typically given to mainstream students at the upper elementary/lower middle-school levels. This information can be used in the test development process to ensure that the types of reading passages and the types of tasks/items used in assessments are familiar to both teachers and students. This will allow for as authentic an assessment of the English language as it is used in academic contexts as possible.

Teacher, student, and classroom demographics. The teachers had 10 years of teaching experience with their current grade level on average (SD = 5.5). The variation is quite considerable with years teaching at current grade ranging from 3 to 16 years. The ELD specialist had been teaching for a total of 42 years but taught classes across the full spectrum of elementary grades. The average class size taught was 32 (SD = 4) and ranged in size from just 27 students in a fourth-grade class to 36 students in a sixth-grade class.

As mentioned, the teachers had been deliberately selected because they taught primarily native speakers, RFEP and/or FEP students. We found that, on average, classes contained just one EL student (SD = 2), but the range included as many as 5 EL students in a fifth-grade class to none in 4 other classes (across all three grade levels).[8] There was an average of 2 RFEP (SD = 2) and 6 FEP (SD = 10) students per class. There was large variation in the latter with one fifth-grade class having 29 LEP students and 2 classes (fourth and sixth grades) having no LEP students at all.

Teachers rated student reading ability separately for native speakers and any students for whom English was a second language (i.e., any English as a Second Language [ESL] background student, namely EL, RFEP and FEP students).[9] The response options for this question included (a) "below grade level," (b) "at grade level," (c) "above grade level," and (d) "Not Applicable." These ratings were conducted by teachers in their content area review working groups. On average, teachers of mathematics rated native speaking students as 2.3 (SD = 0.58) and ESL background students as 1.3 (0.58). All the teachers of science rated native speaking students as 2 and ESL students as 1.67 (SD = 0.58) on average. The teachers of social studies rated native speaking students as 1.67 (SD = 0.58) on average and all the ESL

[8] One sixth-grade teacher did not know the English language proficiency status of students in the class.

[9] These ratings were actually part of the AELP Text Review Form rather than the teacher questionnaire but provide student demographic information pertinent to this section of the report.

background students as 1. The differences across content-area working groups may reflect the greater demands on student reading ability and teacher expectations of reading ability in social studies compared with mathematics. These results however, suggest that most teachers rate students as only at grade level or below in reading ability, with the ESL background students between about a third of a point and a full point behind their native speaking peers.

Summary of teacher textbook and classroom assessment use. The teacher questionnaire also provided information on teacher use of textbooks and classroom assessment. However, this information can only be suggestive given the small number of teachers in the focus group. The purpose of including these questions on the questionnaire was primarily to pilot their accessibility and utility rather than collect meaningful data about textbook and assessment practices at this stage in our research. However, for completeness, Appendix F provides full details of the teachers' responses by content area. A brief summary of responses to the questionnaire is provided here.

The key similarities across content areas are:

- *Format of reading from textbooks*: Nearly all teachers teaching in any content area report primarily reading to the whole class.

- *Explaining vocabulary from textbooks:* No matter what content area they teach, nearly all teachers report providing an explanation of new vocabulary on a daily basis.

- *Frequency of assigning graphics-based tasks from textbooks:* Across all content areas, most teachers rarely assign graphics-based tasks from textbooks.

- *Assigning individual and pair/group essay or report writing*: (Applicable to science, social studies and language arts only). Most teachers teaching these content areas assign report or essay writing on a monthly basis.

The key differences across content areas are:

- *Frequency of reading from textbooks*: With the exception of science teaching, most teachers use textbooks daily. Most teachers report using science textbooks just once or twice a week.

- *Size of repertoire of reading formats*: In all content areas but mathematics, most teachers report using all four reading formats (teacher to whole group,

student to whole group, student-to-student and silent reading). For mathematics, most teachers report using just two reading formats (teacher to whole group and silent reading).

- *Frequency of assigning reading from textbooks for homework*: Teacher responses are quite variable by content area. Most teachers never assign reading for homework in mathematics. In science and social studies this is primarily one or two times a week. Most language arts teachers assign reading homework daily.

- *Frequency of assigning outside projects:* This is extremely rare for teachers of mathematics, whereas most teachers in the remaining three content areas report assigning projects to students (both individually and in pairs/groups) to be conducted outside the classroom context.

- *Frequency of assigning short answer tasks from textbooks:* Short answer task types are given on a daily basis by half the teachers teaching mathematics, and most remaining teachers give them once or twice a week, whereas such tasks were given by most teachers in the other content areas only once or twice a week.

- *Frequency of in-class assessment:* Most teachers in mathematics, science and social studies report assessing students once or twice a month. However while half of the language arts teachers follow this schedule, half reported assessing students far more frequently (a quarter daily and a quarter at least once a week).

- *Preferred types of in-class assessment:* Science, social studies and language arts rely most on short answer format items, whereas mathematics teachers rely most on multiple-choice items. However, a number of additional item types closely followed these main preferences, with mathematics and language arts teachers using fill-in responses as well, and science teachers also relying on multiple choice. Teachers teaching social studies also report use of graphic organizers during assessment, and these teachers as well as language arts teachers use essay writing as another common form of assessment.

Text Review

Findings from the text review session will be discussed by content matter in the following paragraphs. Teacher responses to the AELP Text Review Forms for each of the text selections and their verbal comments during the focus group discussions, where relevant, are also provided.

Review of mathematics text selections. The three teachers in the math working group reviewed and commented on 11 text selections and the ELD specialist reviewed and commented on five of these selections. Beginning with texts that teachers felt were unrepresentative of what they typically use in their classes, in the case of two texts reviewed by the three math teachers, all agreed that the items were not representative of their textbook texts and only one of these texts was considered by just one teacher as representative of other types of material used in the classroom. A further text was also identified as atypical by three of the teachers and as only somewhat representative by the fourth teacher.

In the case of two of these texts, the issue was mathematical complexity. For example, the sixth-grade teacher in the mathematics group noted that profit and loss would be an economics concept still unfamiliar to many of his students. This teacher stated, "The whole idea of profit and cost. That problem is pretty challenging. You have to sort of teach the concepts that you know. Just because you sell things for $1.25 doesn't mean you make $1.25. It's subtracting the cost from the money taken in to figure out the profit and that is a whole separate lesson" (Task and Item Review Whole Group Discussion Notes, June 24, 2004). The third text was considered unnecessarily challenging because it contained several references to different cities, states and countries that were mathematically irrelevant. This text also contained vocabulary that was considered challenging for native English-speaking students. For example, the use of the abbreviation *mi* for *miles* was highlighted by one teacher. Other vocabulary in this text was considered a likely source of difficulty for both native English-speaking and EL students alike (i.e., *stiltwalker*). All three texts received among the highest ratings given for degree of difficulty for native English-speaking students (mean ratings of between 3.25 and 4.0) and EL students (mean ratings of between 3.75 and 4.67).[10] None of these items were considered lacking due to issues of insensitivity to culture, race, gender, etc.

At the opposite extreme to the reviews of these three text selections are six texts that at least half the teachers who reviewed them rated as representative of the types of text their students would encounter in their mathematics textbooks. These texts were also mostly rated as being representative of other print materials by most of the teachers. Difficulty ratings for native English-speaking students ranged from just 2 to 3, and for EL students from 3 to 3.67. One of these texts was the only text in the

[10] The overall average of the mean difficulty ratings of the individual texts for EL students was 3.74, and the average of the mean text difficulty ratings for native English-speaking students was 2.96.

mathematics selections that was rated as insensitive because the problem includes a reference to old-age, potentially raising the subject of death to which some students may be sensitive. Just two remaining texts were rated by only one teacher in each case as representative of the textbooks and of other print material in the classroom. One of these texts was rated as very difficult for both native English-speaking and EL students (4 and 4.5, respectively); the other was rated as difficult for native English-speaking students only.

Many of the vocabulary items that teachers selected as challenging were the same for both EL and native English-speaking students. For example, one of the fifth-grade teachers stated, "The word 'atlases' in the plural form is an unusual word even for native speakers. ELs could probably get stuck on that word and not see everything else." Overall, teachers felt that using more generic language would help reduce students' processing load in solving mathematics problems (Text Review Small Group Discussion Notes, June 24, 2004).

Vocabulary items that were identified as challenging only for EL students include such words as *borrowed, earn, worth, supplies* and *local.* At the grammatical level most structures that the teachers identified were considered to be challenging for EL students only. These grammatical structures include *"more/fewer X than Y," "passing through the intersection,"* and *"X is twice the cost of Y."* Structures that the teachers felt both EL and native English-speaking students would have difficulty with include *"explain how you found...,"*and *"twice as many."* Teachers' responses to the open-ended comments section of the AELP Text Review Form revealed specific details and reasons for some of the teacher judgments about vocabulary and grammatical difficulties. For example, one teacher suggested substituting the word *color* for a general concept word such as *type* because *color* directly names the property that the students need to manipulate.

In summary, it appears that 6 of the 11 texts were adequately representative of texts the teachers expect students to encounter in their mathematics classes and moreover these texts were surprisingly homogenous in their relatively favorable rating of difficulty for both native English-speaking students and EL students.

Review of science text selections. The three teachers in the science working group reviewed and commented on 6 text selections and the ELD specialist reviewed and commented on 2 of these selections. Just one of the 6 text selections was thought to be unrepresentative by at least half (2) of the teachers. These teachers

felt the text is poorly written or "choppy." Two of the remaining 5 texts were unanimously voted as typical, 2 were rated typical by the majority of the group, and one was rated typical by only one teacher and somewhat typical by the remaining teachers. The latter one was less liked because teachers felt it was "conceptually simple" and contained "less content specific vocabulary." This text also received the lowest difficulty ratings for both native English-speaking (2) and EL (3) students. Every text was rated as either representative or somewhat representative of the supplementary material these teachers typically use in their classrooms. None of the texts were deemed to have issues related to race, gender, etc. The level of difficulty for EL and native English-speaking students was comparable, with the average ratings for native English-speaking at 3.04 (range 2 to 3.67) and at 3.68 for EL students (range 3 to 4.67).

Much of the same vocabulary that was identified as challenging was specialized academic vocabulary specific to the science content area (e.g., *atmospheric, erosion, condensing*), and thought to be difficult for both native English-speaking and EL students. Vocabulary that was identified as difficult for EL students only included *moistens, mix, continual,* and phonetic spellings of multisyllabic words such as *kan sev va'shen* for *conservation.* Many grammatical structures were identified as challenging for both native English-speaking and EL students. These include use of the existential "*There is…"* and the locative *There* as in *"There, water and minerals diffuse into the blood and wastes are removed from the body."* Challenging grammatical structures identified for ELs only included comparative constructions such as *less X than Y,* and constructions with verbs separated from their subjects by prepositional phrases in the noun phrase (e.g., "*Fresh water for a given area is determined…."). In* sum, overall we can consider the selection of these science texts to be quite successful. Most of these text selections were thought to be typical of the science textbooks and supplementary materials from which the teachers typically use.

Review of social studies text selections. The three teachers in the social studies working group reviewed and commented on 5 text selections and the ELD specialist reviewed and commented on 2 of these selections. Three of the texts were rated as representative of social studies textbooks by at least 2 of the teachers. One of the two atypical texts was thought to use "easier" language and shorter sentence structures. The amount of detail in the content in this text selection was also thought to be greater than what is typically found in one teacher's textbook (e.g., *"More info used to*

give a biography of one person. Our text, America Will Be, does not do this"). The remaining atypical text was judged so by one teacher because it used more examples and more descriptive language than her own textbook. Neither atypical text appears to be more or less difficult for native English-speaking and/or EL students than any other social studies text. These two texts and most of the other texts were at least representative of supplementary materials for nearly all the teachers. In terms of difficulty ratings overall, there was little variation across texts for either native English-speaking or EL ratings. The average rating for native English-speaking student difficulty was 2.77 (range 2.75 to 3) and 4.27 (range 4 to 4.67) for EL student difficulty. One teacher felt that the text that dealt with life in a colonial New England town may have an issue of bias against women, but did not elaborate on this further.

Most vocabulary identified as challenging was deemed so for both native English-speaking and EL students, consisting primarily of specialized content vocabulary such as *cartographer* and *surveyor*. It is interesting to note given the history text selected, that much of this vocabulary is considered "old-fashioned" and unlikely to be found in the children's lexicons (e.g., *lame, cobbler, herder, gristmill).* Vocabulary that was thought to be difficult just for EL students alone include *fuel, narrow, generosity* and *determining.* Teachers identified many phrases that they felt would be grammatically challenging for EL students, but that they did not think would trouble native English-speaking students. These include somewhat idiomatic uses of English such as "*to map the place," "this side of the grave"* and "*self-sufficient communities."* Many grammatical structures were deemed to be challenging for both native English-speaking and EL students, for example, "*…hold ourselves bound to obey any laws,* and *"…a surprise attack on the close to 140 Hessian troops".* Comparative constructions such as "*as X as they were Y"* (e.g., *"They were often as beautiful as they were useful"*) were identified as difficult for EL students only.

In summary, the selection of texts for social studies was generally quite successful in terms of representing the type of textbooks and materials teachers typically use in class. However, the difference in the degree of difficulty these texts were rated for with respect to native English-speaking and EL students is quite large. On average, the texts are rated 1.5 points more difficult for EL students than for native English-speaking students.

Comparisons across the three content areas. We were most successful at selecting representative science texts and, to just a slightly lesser degree, social studies texts. With an average rating of 4.27, social studies texts were rated the

hardest for EL students of all content areas, and science was rated the lowest at 3.68, but closely followed by mathematics at 3.74. Moreover, the difference in degree of difficultly for native English-speaking and EL students was the greatest in social studies which may be largely a function of all the "difficult" vocabulary and idiomatic phrasing.

Task/Item Review

Teacher responses to the questions on the AELP Task/Item Review Form for each of the items within a text selection, and summary information on items taken in aggregate for each content area, will be provided, including teacher comments made during the focus group discussions where relevant. Tables 11 and 12 present information discussed and summarized in each subsection. Table 11 shows the number and proportion of items by item type category that the teachers rated as representative of items they typically used for evaluation or practice with Native English-Speaking Students (EO) and EL students in classroom settings. Table 12 shows the mean ratings for perceived difficulty of each item type category for EO and EL students, according to the teachers.

Table 11

Teacher Rating of Item Typicality For Both Native English-Speaking Students (EO) and English Learners (EL)

Item Type	Math		Science		Social Studies	
	EO	EL	EO	EL	EO	EL
'Wh' questions	26/48	26/48	21/26	21/26	21/25	15/25
	(54%)	(54%)	(81%)	(81%)	(84%)	(60%)
	[13]	[13]	[7]	[7]	[7]	[7]
Sentence completion/statement	2/3	2/3	25/40	25/40	13/20	11/20
	(67%)	(67%)	(63%)	(63%)	(65%)	(55%)
	[1]	[1]	[12]	[12]	[5]	[5]
Cloze paragraph	N/A	N/A	0/3	0/3	N/A	N/A
			0	0		
			[1]	[1]		
Sequencing	2/3	2/3	2/3	2/3	4/4	3/4
(prompt is an imperative statement)	(67%)	(67%)	(67%)	(67%)	(100%)	(75%)
	[1]	[1]	[1]	[1]	[1]	[1]
Matching	N/A	N/A	N/A	N/A	14/14	9/14
(prompt is an imperative statement)					(100%)	(64%)
					[4]	[4]
Graphic organizer	5/9	5/9	6/8	6/8	2/4	1/4
(prompt is an imperative statement)	(56%)	(56%)	(75%)	(75%)	(50%)	(25%)
	[3]	[3]	[2]	[2]	[1]	[1]

Note. The denominator indicates the total number of teacher responses obtained for a particular item type, and the numerator refers to the number of teacher responses that rated the item as representative of the test item used with English-only speaking or with EL students. Percentage in the parenthesis is derived from the fraction. Numbers in brackets refer to the number of item types in each subject matter. N/A denotes that an item was not developed for a particular content area.

Table 12

Teacher Rating of Item Difficulty For Both EO and EL

Item Type	Math		Science		Social Studies	
	EO	EL	EO	EL	EO	EL
'Wh' questions	2.40	3.15	3.15	4.19	3.38	4.56
	(1-5)	(1-5)	(2-4)	(3-5)	(3-5)	(3-5)
	[13]	[13]	[7]	[7]	[7]	[7]
Sentence completion/statement	2.33	3.33	2.81	3.76	3.95	4.55
	(1-3)	(3-4)	(1-5)	(3-5)	(3-5)	(3-5)
	[1]	[1]	[12]	[12]	[5]	[5]
Cloze paragraph	N/A	N/A	4	4.67	N/A	N/A
			(2-5)	(4-5)		
			[1]	[1]		
Sequencing	1.33	2	3	4.33	3	4.25
(prompt is an imperative statement)	(1-2)	(1-4)	(2-4)	(4-5)	(0)	(3-5)
	[1]	[1]	[1]	[1]	[1]	[1]
Matching	N/A	N/A	N/A	N/A	2.94	4.38
(prompt is an imperative statement)					(2-3)	(3-5)
					[4]	[4]
Graphic organizer	2.44	3.33	2.88	3.88	4.25	5
(prompt is an imperative statement)	(1-4)	(2-5)	(1-4)	(3-5)	(4-5)	(0)
	[3]	[3]	[2]	[2]	[1]	[1]

Note. The mean teacher rating on item difficulty for English-speaking and EL students. The range of ratings is included in parentheses. Numbers in brackets refer to the number of item types in each subject matter. N/A denotes that an item was not developed for a particular content area.

Review of mathematics tasks/items. The three teachers in the mathematics working group reviewed and commented on 18 tasks/items and the ELD specialist reviewed and commented on 9 of these items. Ten of the 18 mathematics items were of a type used by at least two of the teachers with both EO and EL students (i.e., 50% or more of items in item type categories were rated typical; see Table 11 for comparable ratings for item use with EO and EL students). The average difficulty rating of these items for native English-speaking students was 2.38 (range 1.33 to 4.00) and 3.17 for EL students (range 1.67 to 4.67; see Table 12). The difference in difficulty for native English-speaking students and EL students is relatively small.

Indeed, one teacher felt some items were *"difficult for both regular and EL students."* There were item types that were thought to assume too much background knowledge (e.g., about days of the week and synonyms for *weekend).* All three item types for one text selection were judged to be typical of items used with EOs and ELs by at least two of the teachers. Indeed, one teacher wrote: *"This type of questioning is excellent! It helps EL students become more analytical and helps syntax etc."* Finally, some item types were deemed as "*not [a] math issue,"* although it had been stated that the task and items were designed to assess academic uses of English not content knowledge.

Much of the language difficulty identified in the mathematics items occurs at the lexical level rather than at the grammatical level. Vocabulary that was identified as challenging for both native English-speaking and EL students tends to be general academic vocabulary (i.e., words not specific to any given content area). These words include *summarize, organize,* and *approximately*. Vocabulary identified as only difficult for EL students includes some everyday words (i.e., non-academic vocabulary) such as *weekend* and *vehicle.* Teachers identified uses of prepositions, especially in an unfamiliar grammatical construction to be challenging for all students (e.g., *"What is the word problem asking about?"* and *"Over which days did the camping trip occur?"*).

To summarize, 10 of the 18 mathematics items were of a type used by at least half of the four teachers with both native English-speaking and EL students. The fact that the remaining eight items were not commonly used by these teachers is not entirely surprising given they are more typically focused on assigning mathematics problems rather than providing a variety of language-rich tasks. The difficulty of these items for native English-speaking students was rated low at just over 2 on a 5-point scale and that for EL students at about one point higher. No mathematics items contained any material that was thought to be sensitive to issues of race, gender, etc.

Review of science tasks/items. The three teachers in the science working group reviewed and commented on 23 tasks/items and the ELD specialist reviewed and commented on 11 of these items. Sixteen of the 23 science items were rated as a type used with all students by at least two and often more of teachers (see Table 11 for comparable ratings for use with EO and EL students). The average difficulty rating for native English-speaking students at 2.96 (range 2 to 4) is one point lower than that for EL students at 3.97 (range 3 to 5; see Table 12). As with mathematics, one

item type that was rated as atypical was deemed *"not a question about science,"* although again we had stressed that the task and items were designed to assess academic English language not content knowledge. Some item types were rated to be very difficult for all students with teachers commenting that *"[I] never use cloze exercise with verbs"* and *"…too difficult to create verb 'evaporate' from 'evaporation'."* At least one item type was reported to be used by no teacher at all.

Vocabulary in science identified as challenging only for EL students includes specialized academic vocabulary such as *evaporate, organ, conservation*. Other challenging vocabulary for EL students includes everyday words such as *moves* and *similar*. There were only three instances of vocabulary that was thought to be challenging to native English-speaking students as well as EL students (*factor, diffuse, invisible*)—a mix of general and specialized academic vocabulary. There is no grammatical structure considered challenging to EL students alone. A few grammatical structures were identified as challenging to all students, but these were few and include complex question formations such as *"What is the third factor used in describing the weather?"*

In summary, 16 of the 23 science items were rated as of a type used with all students by at least half of the four teachers. The difficulty of items for native English-speaking students was in the middle of the scale, with that for EL students one point higher. One science item was identified by one teacher as having a sensitivity bias.

Review of social studies tasks/items. The three teachers in the social studies working group reviewed and commented on 18 tasks/items and the ELD specialist reviewed and commented on 13 of these items. All the items were of a type that at least two (and often all) of the teachers reported using with native English-speaking students, whereas fewer items were rated as typically used with EL students by these teachers (see Table 11 for disparity in the ratings for use with EO and EL students). The items had an average difficulty rating of 3.45 (range 2.75 to 4.33) for native English-speaking students and 4.53 (range 4.45 to 5) for EL students (see Table 12). Teachers commented that some items were *"tricky!", "too hard [for EL students],"* had *"too many choices,"* and *"too much text to skim for word… very difficult for ELs."* For some items teachers made the practical suggestion of a "wordbank" from which students would be able to select the correct target words in order to answer the items. Two teachers rated two different items on colonial New England town-life as insensitive to women. An additional teacher rated one item as

insensitive to the linguistic limitation of EL students who may misinterpret directions to include information from visuals in their responses.

Many of the same vocabulary and grammatical challenges were identified for both native English-speaking and EL students. Vocabulary challenges include general academic vocabulary such as *explain* and *describe*, and specialized academic vocabulary such as *independence* and *rebellion*. Idiomatic phrases such as "*voice her opinion*," were also thought to be problematic for all students. Grammatical structures that were identified as a challenge for all include comparative constructions (e.g., *"…had higher ranks than…"),* gerunds (e.g., *"using the vocabulary"*), passive voice constructions (e.g., *"is quoted in the passage"*), and relative clauses (e.g., *in the order in which…"*).

To conclude, all the items were of a type used with native English-speaking students by at least half of the four teachers and 15 of these were of a type used with EL students. However most of these items were rated as fairly difficult for all students and for EL students in particular. Three items were rated as having some kind of bias.

Comparisons across the three content areas. We were most successful, according to teachers, at writing items that matched the format that teachers use with native English-speaking students in science and social studies and to a lesser degree with EL students in these subjects. This is not surprising given that the text selections and items were meant to closely mirror the types of reading material and level of language demand that students in mainstream classrooms encounter. In contrast, the item types developed for the mathematics texts were more often considered atypical of items used by the teachers in their mathematics classes. Indeed, several of these items and some science items were considered unrepresentative of the tasks teachers typically give to students, because they *are* language tasks. This suggests that a small number of teachers misunderstood the purpose of the items and tasks, which was to measure students' academic English skills. The disparity of item types thought typical of use with EO and EL students was greatest in social studies. Mathematics items were rated as simpler than either science or social studies items for both native English-speaking and EL students. Social studies items were the most difficult for all students. The differences between the difficulty ratings for native English-speaking and EL students are very similar across three content areas, ranging from 0.79 for mathematics, 1.01 for science, to 1.08 for social studies.

Part V: Conclusions and Recommendations

The purpose of conducting external reviews of the texts and items was threefold: (a) to have classroom teachers evaluate the selected texts and item sequences and formats, (b) to use the evaluations to eliminate texts in principled ways from test development, and (c) to use the specific comments and suggestions provided by the teachers as feedback for modifying or eliminating item types. Compilations of the teachers' reviews were given in Part IV. In this section we will concentrate on the latter two points.

Principled Elimination of Texts and Tasks/Items

A major caveat of this work is the weight we lend to the cross-grade teacher reviews of texts and items at this stage in the test development process. Specifically, fourth- through sixth-grade teachers were asked to evaluate fifth-grade texts and we asked content-area teachers to look at language assessment items (which, with the exception of the ELD specialist, they do not usually administer). Fourth-grade teachers might have thought the texts are too difficult because they teach younger students; sixth-grade teachers may have found the texts too simple or redundant. To that end, there are some review comments that may be more suitable and informative than others. For example, knowing that an item is "not worded at elementary school level" is the type of feedback we were looking for, whereas comments such as "not a math issue" or "too much information for students to sift through to get the answers" are less useful. We in fact want students to sift through some measure of language to answer certain types of language questions. This is important because it can give us an idea of how well students can gather meaning from different types of texts.

We therefore adopted a set of criteria for eliminating texts and items that involve taking all aspects of a text or item into account including: (a) representativeness of content in the textbooks used with native English-speaking and/or EL students, (b) representativeness of the content of supplementary material used with native English-speaking and/or EL students, (c) language difficulty level for native English-speaking and EL students, and (d) identifiable bias due to insensitivity to gender, race, etc. concerns. If half or more of the teachers reviewing the text or item identified it as lacking in one of these areas, then the text or item was

classified as weak in that area.[11] All four criteria were weighted equally in choosing to retain texts or items; that is, all criteria had to be met for half or more of the teachers in order for a text or item to be retained, with one exception—we did not give ratings of representativeness and difficulty level for EL students the same weight if they differed from the ratings of native-speaking students. This decision was made based on the main objective of creating assessments that will measure students' abilities to succeed in mainstream classroom environments. However, information on which texts and items are representative of the print material EL students encounter, or are only difficult for EL students, can be saved and possibly later used in the creation of "intermediate-difficulty" test items. With a sufficient range in difficulty, such test items may prove diagnostic for EL programming and instruction.

Modification of Items and Item Formats

With poorly-rated items or item formats, unlike with poorly-rated text selections constrained by the fact that they are from authentic textbooks, we have the opportunity to modify the language and content to make the items more acceptable. If an item is deemed insensitive due to gender, race, etc. we can attempt to rework the item avoiding the issues that led to its perceived bias during review. Similarly, if an item is deemed linguistically too simple, we can rewrite the item to better approximate the linguistic demands found in the text profiles of the different content areas. An item may require more academic language or longer, more complex sentence structures to make it sufficiently difficult for native-speakers. For example, the item shown in Figure 8 was written to be part of an item sequence for a mathematics word problem. The item received an average language difficulty rating for native speakers of only 1.33 (one of two items to be rated this low) by the three teachers who rated it. The average mathematics rating overall for language difficulty for native speakers was 2.38.[12]

---

[11] In future research, we recommend increasing the number of teachers making the evaluations of texts and items within a grade level. We expect then that the percentage of those who agree on the suitability of a text or item will increase because they will be more likely to share a common notion of what is representative and what is difficult for the same grade of students.

[12] This item was also rated the easiest of the mathematics items for EL students.

Read the following passage and then answer the questions.

On a weekend camping trip, Ken, Eric, and their dad went for a walk on the Appalachian Trail. The first hour, they walked 3/8 mile. The second hour, they walked 4/5 mile. About how many miles did the boys and their dad walk? (Maletsky et al., 2001, p. 318)

4. Which two days are weekend days? Circle the correct answer.

Sunday-Monday Friday-Saturday Saturday-Sunday

Figure 8. Math item M5H-318, number 4.

Teachers felt this item designed to assess student vocabulary knowledge, was too easy because, as one teacher wrote in the review, a student by "5th grade knows what days make a weekend". This item can therefore be modified to assess other more complex synonyms or paraphrasing in the text of the word problem to address intermediate and advanced language proficiency. For example, we can focus on the general academic vocabulary that is synonymous with the everyday language expressed by the phrase "About how many" to produce the reworked item shown in Figure 9.

4.a. Which three words in the word problem mean the same as "calculate approximately"? Circle the correct phrase.

The second hour Walked 3/8 mile About how many

Figure 9. Reworked version of Math item M5H-318, number 4.

Recommendations

A next step in the test development process with these items will be a thorough analysis of the retained texts and items in terms of their vocabulary, grammatical structures, and the organization of discourse in order to assure that they meet the linguistic criteria established in the linguistics profiles (see Appendix B). We then recommend that the viable items from this step be then subjected to small scale tryouts with both native English-speaking and EL students in fourth- through sixth-grade classrooms. From this step in the test development process prototype items

can then be evaluated on a number of criteria (e.g., abilities to discriminate between students with different levels of ELD proficiency; different grade levels, etc.).

Once prototyping has been completed for each item type, test developers can begin development of items for piloting followed by the eventual development and field testing of test forms. A rigorous development process such as this will help establish higher levels of reliability and validity in addition to paving the way for stronger curriculum and teacher training. That is, by approaching the assessment development endeavor in such a comprehensive and rigorous way this process will hopefully lead to improvements not only in the assessment of ELD, but also in the teaching of EL students and the development of appropriate language curriculum and teacher preparation. It may also importantly reveal in what ways native English-speaking students have language development challenges that need to be addressed more adequately by educators.

Finally, as noted in at the start of this report, future test development efforts for modalities other than reading (i.e., speaking, listening, and writing) could replicate this standards-informed approach using relevant empirical evidence from classroom discourse, writing assignment practices, etc. Descriptions of the linguistic characteristics of the academic uses of additional modalities can be integrated with ELD standards in speaking, listening, and writing, as well as the content-area standards that explicitly rely on students' oral language abilities and student production of printed text in the content-areas. This would supplement the approach we adopted with reading that relied both on standards and the prior textbook analyses we had conducted (Butler et al., 2004).

References

Armento, B. J., Cordova, J. M., Klor de Alva, J. J., Nash, G. B., Ng, F., Salter, C. L., Wilson, L.E., & Wixson, K.K. (1999). *Houghton Mifflin social studies: America will be* (21st Century ed., grade 5). Boston: Houghton Mifflin.

Badders, W., Bethel, L. J., Fu, V., Peck, D., Sumners, C., & Valentino, C. (2000). *Houghton Mifflin science: Discovery works* (California ed., grade 5). Boston: Houghton Mifflin.

Bailey, A. L. (2000). Language analysis of standardized achievement tests: Considerations in the assessment of English language learners. In The Validity of administering large-scale content assessments to English language learners: An investigation from three perspectives (Final Deliverable to OERI/OEMLA, Contract No. R305B960002; pp. 85-106). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bailey, A. L., & Butler, F. A. (2002/2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document*. (Final Deliverable to OERI/OBEMLA Contract No. R305B960002) (Currently available as CSE Tech. Rep. No. 611). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bailey, A. L. & Butler, F. A. (2004). Ethical Considerations in the Assessment of the Language and Content Knowledge of English Language Learners K-12. *Language Assessment Quarterly*, 1, 177-193.

Bailey, A. L., Butler, F. A., LaFramenta, C., & Ong, C. (2001/2004). *Towards the characterization of academic language in upper elementary science classrooms* (Final Deliverable to OERI Contract No. R305B960002) (Currently available as CSE Tech. Rep. No. 621). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bailey, A. L., Stevens, R., Butler, F. A., Huang, B., & Miyoshi, J. (2004). [Using Standards and Empirical Evidence to Develop Academic English Proficiency Test Items in Reading] Unpublished raw data.

Banks, J. A., Beyer, B. K., Contreras, G., Craven, J., Ladson-Billings, G., McFarland, M. A., & Parker, W.C. (2001). *United States adventures in time and place (grade 5)*. New York: McGraw-Hill.

Boehm, R. G., Hoone, C., McGowan, T. M., McKinney-Browning, M. C., Miramontes, O.B., Porter, P.H. (2002). *Harcourt Brace social studies: Early United States*. Orlando, FL: Harcourt Brace & Company.

Butler, F. A., Bailey A. L., Stevens, R., Huang, B. & Lord, C. (2004). *Academic English in Fifth-grade Mathematics, Science, and Social Studies Textbooks*. (Final Deliverable to IES, Contract No. R305B960002). (Currently available as CSE Tech. Rep. No. 642). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., & Castellon-Wellington, M. (2000). Students' concurrent performance on tests of English language proficiency and academic achievement. In J. Abedi. A. Bailey, & F. Butler. *The Validity of administering large-scale content assessments to English language learners: An investigation from three perspectives* (Final Deliverable to OERI/OBEMLA, Contract No. R305B960002; pp. 51-83). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., Lord, C., Stevens, R., Borrego, M., & Bailey, A. L. (2003/2004). An approach to operationalizing academic language for language test development purposes: Evidence from fifth-grade science and math (Final Deliverable to IES, Contract No. R305B960002) (Currently available as CSE Tech. Rep. No. 626). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations* (CSE Tech. Rep. No. 448). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

California State Board of Education. (1998). Grade five. In *History-social science content standards for California public schools: Kindergarten through grade twelve* (pp. 16-22). Sacramento, CA: California Department of Education. Retrieved April 21, 2004, from http://www.cde.ca.gov/re/pn/fd/documents/histsocsci-stnd.pdf

California State Board of Education. (1999). *English Language Development Standards for California Public Schools : Kindergarten through grade twelve.* Sacramento, CA: California Department of Education. Retrieved April 21, 2004, from http://www.cde.ca.gov/re/pn/fd/documents/englangdev-stnd.pdf

California State Board of Education. (1999). Grade five. In *Mathematics content standards for California public schools: Kindergarten through grade twelve* (pp. 20-23). Sacramento, CA: California Department of Education. Retrieved April 21, 2004, from http://www.cde.ca.gov/re/pn/fd/documents/math-stnd.pdf.

California State Board of Education. (2000). Grade five. In *Science content standards for California public schools: Kindergarten through grade twelve* (pp. 14-17). Sacramento, CA: California Department of Education. Retrieved April 21, 2004, from http://www.cde.ca.gov/re/pn/fd/documents/sci-stnd.pdf.

Frank, M. S., Jones, R. M., Krockover, G. H., Lang, M. P., McLeod, J. C., Valenta, C. J., & Van Deman, B.A. (2000). *Harcourt science (California ed., grade 5).* Orlando, FL: Harcourt.

Garcia, G. R. (1991). Factors influencing the English reading test performances of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26*, 371-391.

Greenes, C., Leiva, M. A., Vogeli, B. R., Larson, M., Shaw, J. M., & Stiff, L. (2002). *Houghton Mifflin mathematics (California ed., grade 5).* Boston MA: Houghton Mifflin.

Gronlund, N. E. (1998). Assessment of student achievement. Needham Heights, MA: Allyn and Bacon.

Haladyna, T. M. (1994). Developing and validating multiple-choice test items. Hillsdale, NJ: Lawrence Erbaum Associates, Inc., Publishers.

Kehoe, J. (1995) Writing Multiple-Choice Test Items. Retrieved January 3, 2005, from http://www.ericdigests.org/1997-1/test.html

Maletsky, E. M., Andrews, A. G., Bennett, J. M., Burton, G. M., Johnson, H. C., Luckie, L. A., & McLeod, J.C., Newman, V., Scheer, J.K., & Schultz, K.A. (2002). *Harcourt math (grade 5).* Orlando, Fl: Harcourt.

Martin, A. V. (1976). Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly, 10*(1), 91-97.

Morgan, D. L. (1998). *The focus group guidebook*. Thousand Oaks, CA: Sage Publications.

Moyer, R., Daniel, L., Hackett, J., Baptiste, P., Stryker, P., & Vasquez, J. (2001). *McGraw-Hill science (California ed., grade 5).* New York: McGraw-Hill.

Nation, I. S. P. & Coxhead, A. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew and M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge: Cambridge University Press.

Osterlind, S. J. (1998). Constructing test items: Multiple-choice, constructed-response, performance, and other formats. Norwell, MA: Kluwer Academic Publishers.

Scarcella, R. & Zimmerman, C. (1998). Academic words and gender: ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition*, *20*(1), 27-49.

Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs* (CSE Tech. Rep. No. 552). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Willoughby, S. S., Bereiter, C., Hilton, P., & Rubinstein, J. H. (2003). *SRA: Math explorations and applications (grade 5)*. Chicago, Il: SRA/McGraw-Hill.

Appendix A:

Text Selection Checklist

Directions: Please complete this form for each text selection. Space for comments is provided at the end. Be sure to name the number of the guideline that you are addressing when making comments.

Textbook: _____

Publisher name & ISBN: _____

Page No(s): _____

Standard and Performance Indicator Addressed: _____

Guideline	Yes/No	Description/topic
1		Is the standard or indicator too fact oriented?
2		Does the standard or indicator lend itself to texts of the type needed?
3		Does the text require additional support for student understanding?
4		Does the text provide enough general or introductory information to reduce the potential for lack of background knowledge interfering with assessment of language proficiency?
5		Is the text too conceptually dense?
6		Does the topic/text provide opportunity for reader engagement?
7		Is the topic/text potentially offensive or upsetting to students for any reason?

Please check one: Recommend text ☐
Do not recommend text ☐

Comments: Please note any comments or questions you may have about specific guidelines, texts, standards, or indicators in the space below, or provide an explanation if you do not recommend a text.

Appendix B:

Text Profiles

Table B1

Linguistic Profiles of Fifth-Grade Mathematics, Science, and Social Studies Text Selections[a]

	Math	Science	Social Studies
Mean no. of sentences per word problem or paragraph (range)	3 (2-7)	4 (1-8)	4 (1-9)
Mean no. of words per sentence (range)	11 (1-39)	13 (1-37)	14 (3-43)
Lexical diversity ratio	.43	.41	.49
Percentage of all categories of academic vocabulary words[b]	10% (14%)	21% (27%)	24% (24%)
General academic words only	3% (5%)	6% (11%)	3% (7%)
Specialized academic words only	4% (7%)	14% (14%)	9% (11%)
Measurement words only	3% (2%)	1% (1%)	<1%
Proper nouns only (specialized)	<1%	<1%	5% (7%)
Colloquialisms only	<1%	<1%	<1%
Vocabulary features			
Low-frequency words	8% (12%)	8% (12%)	8% (12%)
3-or-more-syllable words	6% (9%)	10% (15%)	12% (16%)
Derived words	2% (4%)	6% (11%)	8% (12%)
No. of unique clause connectors in each subject area	11	7	21
Avg. percentage of nominalizations per selection	<1%	2% (3%)	2% (3%)
Avg. percentage of each sentence type per selection			
Simple sentences	81%	61%	63%
Complex sentences	17%	36%	33%
Other sentence types	2%	3%	4%
Avg. percentage of dependent clauses per selection	6%	29%	28%
Mean no. of passive voice verb forms per sentence	.04	.24	.16
Mean no. of prepositional phrases per sentence	1	1	1
Mean no. of words per prepositional phrase (range)	4 (2-14)	4 (2-17)	4 (2-20)
Mean no. of noun phrases per sentence	.03	.16	.17
Mean no. of words per noun phrase (range)	2 (1-16)	3 (1-23)	3 (1-19)
Mean no. of participial modifiers per sentence	.03	.17	.17
Dominant organizational features			
Classification	0%	17%	0%

(table continues)

Table B1 (*continued*)

	Math	Science	Social Studies
Description	50%	100%	100%
Explanation	0%	42%	33%
Scenario	100%	0%	0%
Sequencing	0%	17%	25%
Supporting organizational features[c]			
Comparison	67%	83%	50%
Definition	0%	83%	75%
Enumeration	92%	100%	100%
Exemplification	0%	75%	83%
Labeling	0%	100%	100%
Paraphrase	17%	58%	67%
Provide instruction or guidance	25%	25%	0%
Quotation	0%	0%	92%
Reference to text or visual	0%	83%	58%
Sequencing	75%	42%	58%

[a]Numbers in this table have been rounded to the nearest whole number for percentages and the nearest one hundredth for decimals. [b]Percentages shown are token (type). [c]The five most frequently occurring supporting features in each subject area are listed here, although there is some overlap, resulting in a total number of 10 supporting features in the list. The percentages represent the percentage of selections in which a particular feature was identified in the passages analyzed in the Butler et al. (2004) report from which this table was excerpted.

Appendix C:

Descriptive Statistics for the Text Selections by Content Area

Table C1

Summary of Descriptive Statistics for the Text Selections by Content Area

Descriptive Statistics	Mathematics	Science	Social Studies
Total word count			
Range	29-212	246-524	306-543
Average (SD)	60.09 (52.73)	365 (106.07)	461.20 (94.93)
No. of sentences per selection			
Range	2-15	20-34	21-41
Average (SD)	5.09 (3.51)	25.67 (5.85)	31.60 (7.40)
Avg. no. of words per sentence			
Range	6.5-17.6	11.18-16.75	13.24-16.90
Average (SD)	11.73 (3.59)	14.14 (1.98)	14.70 (1.40)
No. of paragraphs per selection			
Range	1-2	5-8	6-10
Average (SD)	1.09 (0.30)	6.17 (1.17)	7.60 (1.52)
Avg. no. of sentences per paragraph			
Range	2-7.5	3.33-4.67	3.50-5.13
Average (SD)	4.41 (1.59)	4.16 (0.46)	4.17 (0.70)

Appendix D:

Sample Texts

Passage M5H-318 (Math)

Read the following passage and then answer the questions.

On a weekend camping trip, Ken, Eric, and their dad went for a walk on the Appalachian Trail. The first hour, they walked 3/8 mile. The second hour, they walked 4/5 mile. About how many miles did the boys and their dad walk? (Maletsky et al., 2002, p. 318)

1. What is this math problem asking about?

a) How long the family walked.
b) How far the family walked.
c) How many people walked.
d) How many hours the boys walked.

2. Over which days did the camping trip occur?

_____

Alternate Items

3. What two days of the week are weekend days?

_____

4. Which two days are weekend days? Circle the correct answer.

Sunday-Monday Friday-Saturday Saturday-Sunday

Passage S5H-A19 (Science)

Read the following passage and then answer the questions.

Your digestive system provides the nutrients your cells need to produce energy. To provide nutrients, the digestive system performs two functions. The first is to break food into nutrients. The second is to get the nutrients into the blood. Then the circulatory system transports them to your cells.

Digestion begins as you chew food, breaking it into smaller pieces so that you can swallow it. Glands in your mouth produce saliva. Saliva moistens food and begins to break down starchy foods, such as pasta, into sugars. (If you chew an unsalted cracker for a while, it will begin to taste sweet.)

When you swallow, food passes through the esophagus, a long tube that leads to the stomach. Gastric juice, produced by the stomach, contains acid and chemicals that break down proteins.

After several hours in the stomach, partly digested food moves into the small intestine. Digestion of food into nutrients is completed by chemicals produced in the small intestine. Nutrients diffuse through the villi, projections sticking out of the walls of the small intestine, into the blood. From the small intestine, undigested food passes into the large intestine. There, water and minerals diffuse into the blood, and wastes are removed from the body.

Two other organs have a role in digestion. The liver produces bile, which is stored in the gallbladder until it's needed. Bile breaks down fats into smaller particles that can be more easily digested. The pancreas produces a fluid that neutralizes stomach acid and chemicals that help finish digestion. (Frank et al., 2000, p. A19)

Complete each sentence with one of the words in the list. Each word can be used only once.

saliva	chew	gastric juice
bile	pancreas	

1. The _____ is an organ that helps the body complete digestion.

2. Your mouth makes a fluid called _____.

3. A fluid that helps the body break down fats is _____.

4. Your stomach makes _____ to help you digest proteins.

5. Put the following five sentences in the correct order. The first one is done for you.

__1__ You chew food into small pieces.
_____ Undigested food passes from the small intestine to the large intestine.
_____ The villi diffuse nutrients into the blood.
_____ Food passes through a long tube to the stomach.
_____ Your glands produce saliva.

6. Why is it important to digest food properly? Explain.

_____

_____

Read the following passage and then answer the questions.

George Washington was born in 1732 in Westmoreland County, Virginia. Although his parents were landowners, they were not one of Virginia's wealthiest families. Washington was good at mathematics, but never went to college.

Washington's first job, at the age of 16, was as a surveyor. A surveyor is a person who measures land. In the middle of the 1700s many colonists were moving west and needed his services. His work paid well, and he was able to use his money to buy land.

In 1752 the young Washington joined the Virginia militia. Washington hoped a military career would bring him honor. He became angry when he learned that soldiers from the colonies were paid less to fight for the British than soldiers in the regular British army. Then, during the French and Indian War, the British lowered Colonel Washington's rank because they did not want colonists to rise above captain. Washington left the militia in protest. He later returned when the governor of Virginia restored his original rank.

In 1758, while still in the military, Washington was elected to the Virginia House of Burgesses. There he met Thomas Jefferson and Patrick Henry, and later joined colonial protests against the British.

In 1759 Washington retired from military life to manage his lands. By then he had become the most famous American in the military. That same year he married a wealthy widow named Martha Custis. George and Martha Washington moved to Mount Vernon, the plantation he owned on the Potomac River in Virginia. Martha Washington also supported the patriots. During the American Revolution, she helped her husband with his paperwork. She also sewed socks and cooked soup for the soldiers.

Martha Washington often joined George Washington in the field, where things were going badly for the Continental Army at the end of 1776. Washington was discouraged. He wrote, "Such is my situation that if I were to wish the bitterest curse to an enemy on this side of the grave, I should put him in my [place] with my feelings."

Certain of future victories, General Howe decided to rest for the winter in New York City. Washington knew that the British would not try to advance again until the spring. So he planned a surprise attack on the close to 1,400 Hessian troops in Trenton, New Jersey. The password Washington gave his soldiers was "Victory or Death!" After nightfall on Christmas Day, December 25, 1776, Washington and his troops crossed the Delaware River into New Jersey. The next morning, they surprised the Hessians, who quickly surrendered. "This is a glorious day for our country," said Washington. (Banks et al., 2001, p. 322-323)

1. Put the six sentences in the order in which the events occurred.

_____ George Washington was born.
_____ His troops won an important battle.
_____ He became an elected official.
_____ He married his wife.
_____ He joined the military.
_____ He worked as a surveyor.

2. According to the passage, why did George Washington do well at his first job?

_____

_____

3. How were the colonial soldiers and British soldiers treated differently? Circle the best answer.

a) The British were paid more than the colonists.
b) The colonists had higher ranks than the British.
c) The British and colonists were treated the same.
d) Both types of soldier had socks and soup.

4. The passage says: "Such is my situation that if I were to wish the bitterest curse to an enemy on this side of the grave, I should put him in my [place] with my feelings." Which of the following statements is true? The quote is used to:

a) Describe George Washington's temperament.
b) Show how George Washington felt at the time.
c) Explain why George Washington was happy.
d) prove that George Washington was a good soldier.

Fill in the blanks using vocabulary words from the passage.

5. George Washington and his friends _____ against the British.

6. George Washington's wife was _____ at home and on the battlefield.

7. The Hessian troops were _____ by George Washington.

Appendix E:

AELP Teacher Questionnaire Parts I and II

Developing Standard-based Measures of Academic English Language Proficiency

Teacher Questionnaire: Part I

Thank you for agreeing to participate in the Academic English Language Proficiency (AELP) focus group for teachers. This is Part I of a two-part teacher questionnaire. It will take approximately 10 minutes to complete this part of the questionnaire. We will ask you to complete Part II at the AELP focus group meeting. This form should be completed prior to the focus group meeting on June 24 and then returned to Judy at the meeting. If you have any questions or need clarification, please contact Judy at (310) 794-9139. Thank you again for your invaluable assistance with this research project.

Name: _____ Grade: _____

School Name: _____

For Questions 1 and 2, please use data from the current academic year/semester.

1. a. Elementary School Teachers ONLY: How many students do you have in your class? _____

   b. Middle School Teachers ONLY: How many students do you have in each period? Zero Per. ____

   Per. 1 ____ Per. 2 ____ Per. 3 ____ Per. 4 ____ Per. 5 ____ Per. 6 ____ Per. 7 ____

2. a. Elementary School Teachers ONLY: How many of your students are classified as LEP, RFEP, and FEP?

   LEP _____ RFEP _____ FEP _____

   b. Middle School Teachers ONLY: How many of your students are classified as LEP, RFEP, and FEP?

Classification	Zero Per.	Per. 1	Per. 2	Per. 3	Per. 4	Per. 5	Per. 6	Per. 7
LEP	_____	_____	_____	_____	_____	_____	_____	_____
RFEP	_____	_____	_____	_____	_____	_____	_____	_____
FEP	_____	_____	_____	_____	_____	_____	_____	_____

For Questions 3 and 4, please answer each question as it applies to the subject area(s) that <u>you</u> currently teach. Middle School Teachers: If your list of textbook(s) or materials differs across class periods, please indicate the class period(s) in parentheses.

3. What textbooks do you currently use for the subjects below? Please include (1) textbook title, (2) publisher, and (3) copyright year (if available):

   Mathematics: _____

   Science: _____

   Social Studies: _____

   Language Arts: _____

4. What types of materials other than textbooks (e.g., textbook supplements (indicate), internet materials, trade books, realia, etc.) do you currently use for the subjects below?

   Mathematics: _____

   Science: _____

   Social Studies: _____

   Language Arts: _____

Developing Standard-based Measures of Academic English Language Proficiency

Teacher Questionnaire: Part II

Thank you for agreeing to participate in this survey. This is Part II of a two-part teacher questionnaire. The purpose of the questionnaire is to determine (a) how textbooks are used in and outside of the classroom, and (b) what types of tasks and test items are assigned to students in and outside of class. The information you provide will be used to guide text selection for the development of assessments of academic language proficiency for English learners. To assure confidentiality, your name and school site will be identified only by a code number. Any information that is obtained in connection with this study will remain confidential and will be disclosed only with your permission or as required by law. It will take approximately 15 minutes to complete this part of the questionnaire. Answer each question as best you can, and be sure to note any questions or comments you may have in the comments section at the end.

Name: _____

School Name: _____

5. What grade level(s) do you currently teach? _____
6. How long have you taught at this grade level? _____

7a. What subject area(s) do you currently teach?

Elementary School Teachers: ❒ "All" ❒ Other (Explain) _____

Middle School Teachers: ❒ Mathematics ❒ Science ❒ Social Studies

Question 7b is for Elementary School Teachers ONLY:

7b. On average, how much instructional time (hours) during the week is devoted to each of the following subject areas?

	Monday	Tuesday	Wednesday	Thursday	Friday
Mathematics:					
Science:					
Social Studies:					
Language Arts:					

For Questions 8 through 18, please answer each question as it applies to the subject area(s) that <u>you</u> curre teach.

8. How often do you explain or provide definitions for vocabulary used in textbooks during class?

Mathematics: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

9. How often do you provide explanations about the language used in textbooks during class?

Mathematics: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

10. How often do students have the opportunity to read textbooks **in class**?

Mathematics: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

11. What grouping and format do you use for reading textbooks (check all that apply)?

Mathematics: ❒ Teacher to whole class ❒ Student to whole class ❒ Student to student ❒ Silen reading

Science: ❒ Teacher to whole class ❒ Student to whole class ❒ Student to student ❒ Silen reading

Social Studies: ❒ Teacher to whole class ❒ Student to whole class ❒ Student to student ❒ Silen reading

Language Arts: ❒ Teacher to whole class ❒ Student to whole class ❒ Student to student ❒ Silen reading

12. How often do you give reading assignments from textbooks for homework?

Mathematics: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

13. How often do you give reading assignments from supplementary materials for homework?

Mathematics: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: ❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

14. Please tell us about your other uses of textbooks, if any (e.g., end of chapter review questions).

_____

_____

15. How do you assign problems, items, and/or tasks from textbooks for the following? Check all that apply:

Mathematics: ❒ in class, pair work ❒ in class, groups ❒ in class, individually
❒ at home, individually ❒ outside class, pairs or groups (e.g., projects)

Science: ❒ in class, pair work ❒ in class, groups ❒ in class, individually
❒ at home, individually ❒ outside class, pairs or groups (e.g., projects)

Social Studies: ❒ in class, pair work ❒ in class, groups ❒ in class, individually
❒ at home, individually ❒ outside class, pairs or groups (e.g., projects)

Language Arts: ❒ in class, pair work ❒ in class, groups ❒ in class, individually
❒ at home, individually ❒ outside class, pairs or groups (e.g., projects)

For questions 16 and 17, question 15 asks about tasks/items students do ***individually*** and question 16 asks about tasks/items students do in ***pairs or groups***.

16. How often do you assign the following types of tasks/items to your students to do ***individually***?

Mathematics: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

17. How often do you assign the following types of tasks/items to your students to do in ***pairs* or *groups***?

Mathematics: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Science: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Social Studies: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Language Arts: *Short answer questions*

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Complete graphics (tables, graphs, etc.)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Projects/reports/essays

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

Other (please describe: _____)

❒ Daily ❒ Once or twice a week ❒ Once or twice a month ❒ Never

18. How often do you give tests based on classroom instruction (i.e., not state or direct mandated standardized assessments)?

Mathematics:	❒ Daily	❒ Once or twice a week	❒ Once or twice a month	❒ Never
Science:	❒ Daily	❒ Once or twice a week	❒ Once or twice a month	❒ Never
Social Studies:	❒ Daily	❒ Once or twice a week	❒ Once or twice a month	❒ Never
Language Arts:	❒ Daily	❒ Once or twice a week	❒ Once or twice a month	❒ Never

19. What types of items or tasks do you use on your classroom-instruction-based tests (check all that apply)?

Mathematics:	Science:	Social Studies:	Language Arts:
❒ Essay	❒ Essay	❒ Essay	❒ Essay
❒ Fill-in	❒ Fill-in	❒ Fill-in	❒ Fill-in
❒ Graphic Organizers	❒ Graphic Organizers	❒ Graphic Organizers	❒ Graphic Organizers
❒ Multiple-choice	❒ Multiple-choice	❒ Multiple-choice	❒ Multiple-choice
❒ Short Answer	❒ Short Answer	❒ Short Answer	❒ Short Answer
❒ True/False	❒ True/False	❒ True/False	❒ True/False
❒ Other _____	❒ Other _____	❒ Other _____	❒ Other _____
❒ Other _____	❒ Other _____	❒ Other _____	❒ Other _____
❒ Other _____	❒ Other _____	❒ Other _____	❒ Other _____

20. Additional comments/Questions: _____

_____

Thank you!

Appendix F:

Textbooks and Assessments in the Classroom

Textbooks and supplementary materials. All the teachers reported using textbooks and supplementary materials in their classrooms. A number of the same textbooks appeared in more than one class. These include *Everyday Mathematics* (Everyday Learning, 2002), *Harcourt Math California Edition* (Harcourt, 2002), *Focus on Earth Science* (Prentice Hall 2001), *Harcourt Science California Edition* (Harcourt, 2000), and *Discovery Works* (Houghton Mifflin, 2000). Teachers used various aspects of the textbooks for different purposes. For example, teachers assigned the end of section and chapter questions to students to review topics and used the experiments suggested in the textbooks. Glossaries were used for "investigations." The textbooks were used to develop paragraph length answers to "chapter focus" questions, to prepare for tests, and for remedial tutoring. Supplementary materials included additional enrichment materials supplied by textbook publishers such as reteach and practice workbooks, and student reference books or study guides. Other supplementary materials included additional curricula (e.g., Bernstein's *Artful Learning* ELD model; *Write Traits*; Houghton Mifflin's *Into English*), internet searches, experiment kits, videos, games, field trips, skits, performances (e.g., musicals), maps, realia, manipulatives, picture books, flashcards, overheads, and projects.

Use of mathematics textbooks and assessments in the classroom. Five of the eight teachers who teach mathematics use textbooks on a daily basis. A further two use them at least once or twice a week. The majority (7/8) of the teachers read textbooks to the whole class and have students read to the whole class. Only three have students read to other students. Most teachers report engaging their students in silent reading (6/8). Use of just two different reading formats is the most popular category reported—the fewest number of different reading formats of all the content areas. In response to questions about whether they explain or provide definitions of vocabulary used in textbooks, most teachers teaching mathematics (5/8) report doing so on a daily basis and just half explain textbook language on a daily basis.

Turning to assignments, most teachers of mathematics (5/8) never gave reading assignments from textbooks for homework and half never assigned reading from supplementary materials for homework. All teachers reported assigning tasks (e.g., word problems) from textbooks in class with pairs and groups of students. The majority (6/8) did so with students working individually. Rarely (just one teacher) did teachers indicate assigning tasks outside of class to pairs or groups of students

(e.g., projects). The frequency with which teachers assign different types of tasks to individual and groups of students was also reported. Half of teachers assigned short answer mathematics items to students individually on a daily basis. Of the remaining teachers, most did so once or twice a week. Completing graphics (e.g., tables, diagrams etc.) are also used by most teachers just once or twice a week. Teachers who chose to respond to the open-ended prompt for information about any additional forms of tasks they assign mentioned using fluency passages and algorithm practice daily, and manipulatives and lengthy problem solving tasks once or twice a week. Assigning task types to pairs or groups changes these results somewhat, with most teachers assigning short answer responses to groupings of students only once or twice a week and graphics-based tasks even less often (once or twice a month). One teacher responded with additional information about assigning in-depth lengthy problem solving tasks to pairs or groups on at least a weekly basis.

In terms of frequency of testing student knowledge based on classroom instruction (i.e., not a state or district mandated standardized test), half of the teachers reported assessing their students once or twice a month in mathematics, closely followed by 38% who reported assessing students once or twice a week. The item type most often used is multiple choice as Figure F1 illustrates. This is closely followed by fill-in responses.

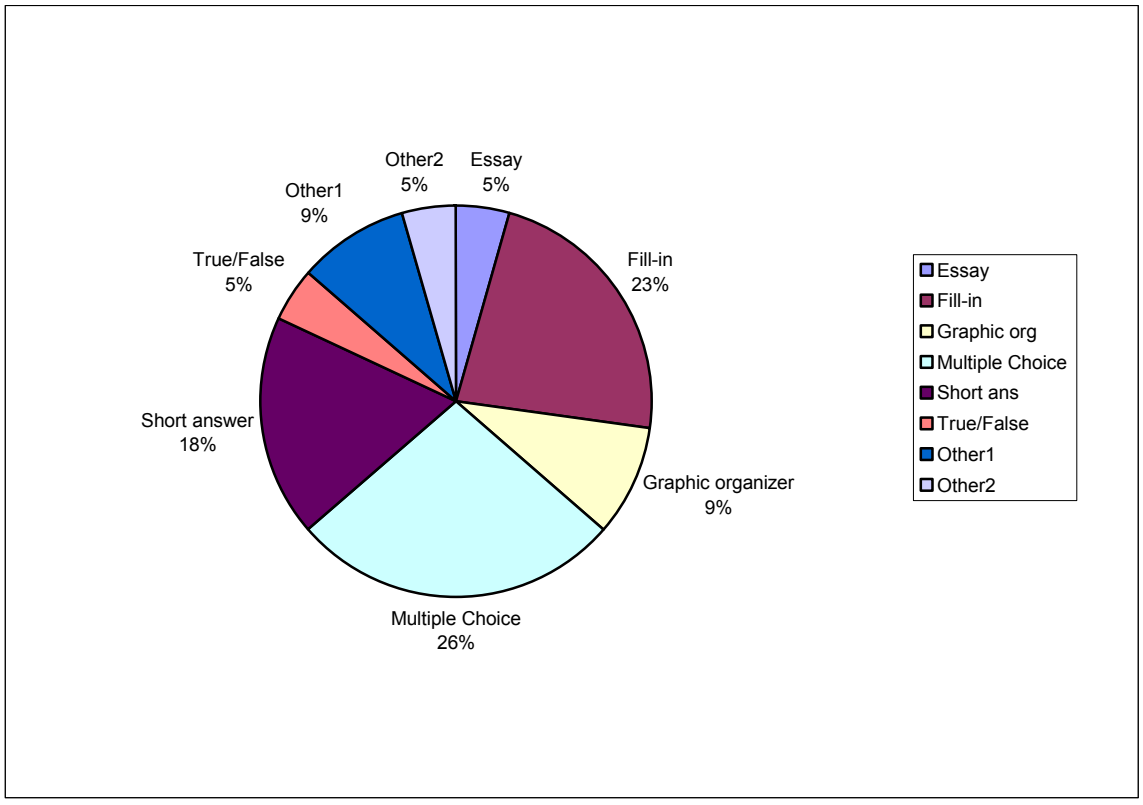


Figure F1. Mathematics assessment items by type. These percentages refer to the number of teachers reporting the type of assessment. Additional "other" types that teachers volunteered include solving equations/computation and word problems that we had intended to be covered by the response options to assessment items such as short answer, fill-in and multi choice.

Use of science textbooks and assessments in the classroom. Five of the eight teachers who teach science use textbooks once or twice a week. Only three use them daily. Reading to the whole group was favored by 7/8 of teachers. All teachers who teach science have students read to the whole group. Just five have students read to each other and six have students engage in silent reading. Use of all four reading formats in the classrooms received the most teacher ratings (3). Teachers were split fifty-fifty between explaining textbook vocabulary on a daily or once/twice a week basis. The majority (6/8) explain textbook language on a daily basis.

Five teachers who teach science assign science textbook reading for homework at least once or twice a week. Once or twice a week students are also assigned reading from supplementary materials for home work by three of the teachers; the remaining teachers were divided between never (2), once or twice a month (2) or daily (1). Most teachers assign tasks to work in class in pairs and individually. All assign in class group work and most (5/8) also assign outside projects. Most teachers assign short answer questions once or twice a week to both individual and groupings of students and tend to assign graphics-based tasks just once or twice a month to both individuals and groups. All assign a report, project or essay to individuals or groupings of students once or twice a month as well.

Turning to in-class assessment practices, 6/8 of the teachers report that they give assessments once or twice a month. Figure F2 shows that short answer is the most common type of science assessment these teachers utilize, closely followed by multiple choice.

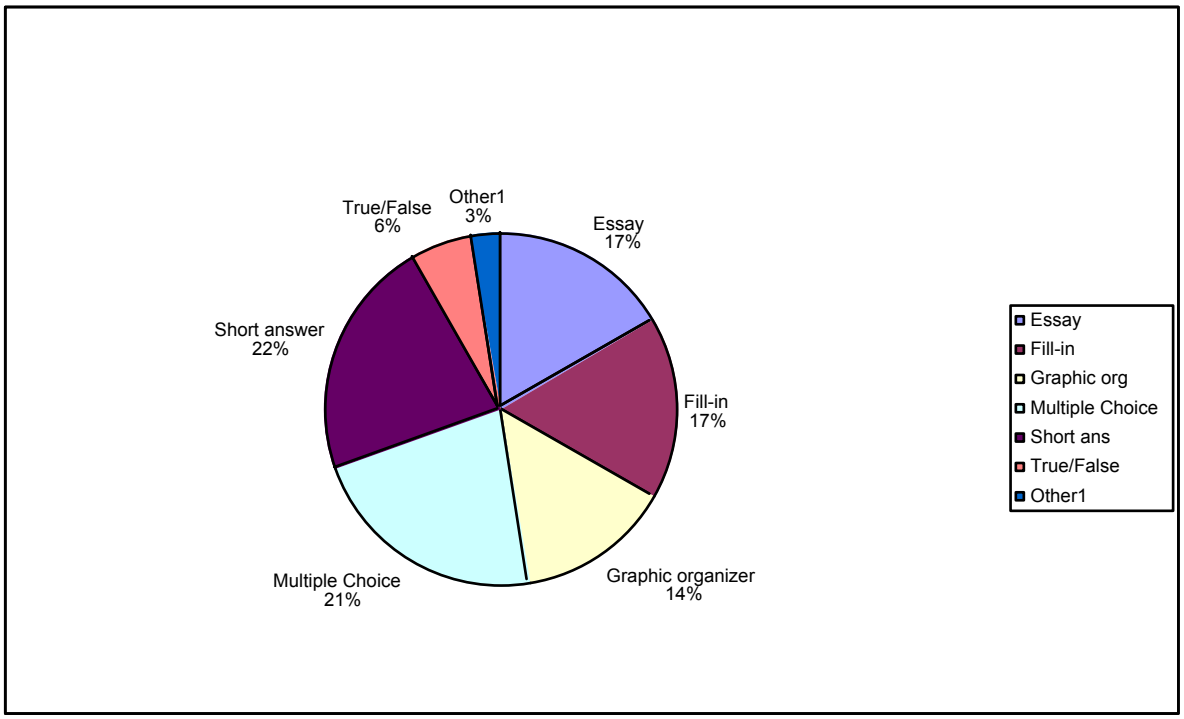


Figure F2. Science assessment items by type. Just one teacher volunteered information on an additional task type: she also uses drawing to assess student science knowledge.

Use of social studies textbooks and assessments in the classroom. Of the seven teachers who teach social studies, the majority of them (5) use textbooks daily. One hundred percent of these teachers read to the whole class and the majority (6) require students to read to the whole group, student-to-student and engage in silent reading. Five of these teachers use all four reading formats. The majority of teachers (5) explain textbook vocabulary and all explain textbook language on a daily basis.

Assigning reading from textbooks for homework once or twice a week was the most frequent response from teachers teaching social studies (3). Most teachers (5) assign homework involving the reading of supplementary materials just once or twice a month. The remaining teachers all give such an assignment daily. The majority of teachers (6) assign tasks in class to pairs and groups, and all teachers assign tasks to individuals in class. Five assign outside projects to pairs or groups of students. Short answer response tasks are assigned by most teachers (5) to individuals or groupings of students one or two times each week and graphics-based tasks are rarer (equal number of teachers [3] assign them to individual students once or twice a week, or once or twice a month—the remaining teacher, never). The majority of teachers assign these tasks to pairs or groups just once or twice a month. Most (6/7) assign a report, project, or essay just once or twice a month to both individual and groupings of students. Three teachers offered

information on additional tasks that they assign. One assigns individual note-taking daily and two others report that they have groups of students enact plays on at least a monthly basis.

Finally, five of the seven teachers teaching social studies report giving instruction-based assessments to their students at most once or twice a month. Figure F3 shows that short answers and graphic organizers are the most common type of assessments they utilize, closely followed by essays.

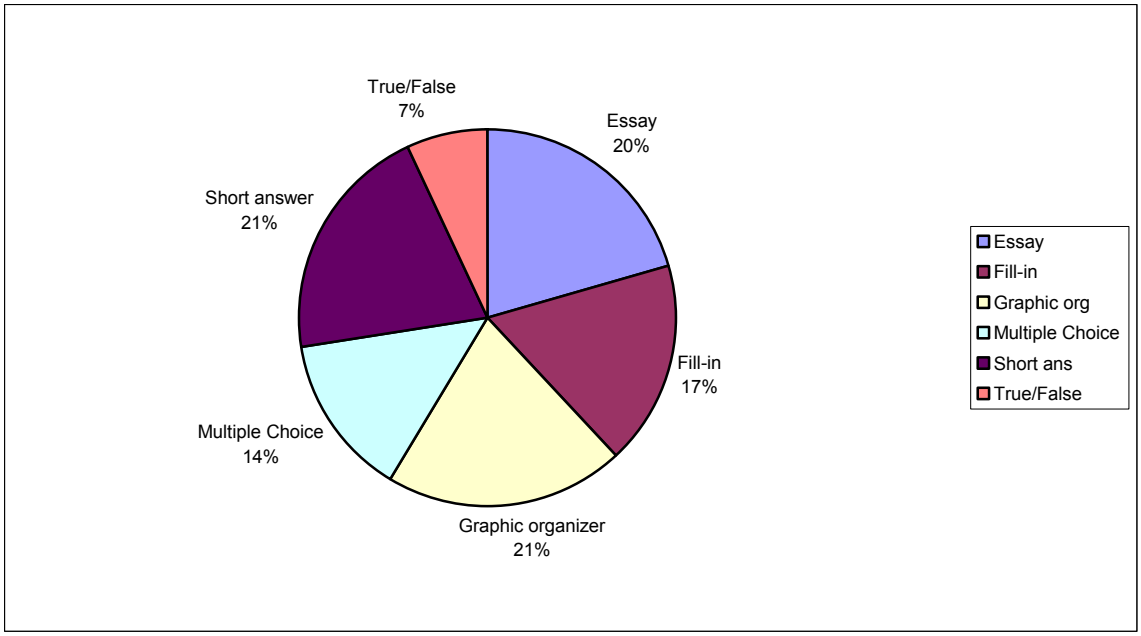


Figure F3. Social science assessment items by type.

Use of language arts textbooks and assessments in the classroom. By far the most widespread use of textbooks is in language arts, where seven of the eight teachers who teach language arts use textbooks on a daily basis. The majority of the teachers also read to the whole class, have students read to the whole class and to one another, and all require students to engage in silent reading. Five of the eight use all reading formats. All teachers report explaining both vocabulary and textbook language daily.

Half of the language arts teachers assign reading from textbooks for homework on a daily basis. Most of the remaining teachers give such assignments once or twice a week. Half of teachers also assign reading from supplementary materials on a once or twice a week basis with the remaining teachers split between daily assignment and just once or twice a month. Most teachers (7/8) assign in-class pair and group work and all assign individuals in class tasks. Six assign outside projects to pairs or groups of students. Most teachers assign short answer tasks to individuals and to

groups of students once or twice a week. Graphic organizers are more variably assigned with a small majority (3) assigning them once or twice a month only, other teachers split between never and weekly (2) and just one teacher daily—whether to individuals or to pairs/groups of students. Five teachers assign a report, project or essay just once or twice a month to individual students (6 assign such tasks to pairs/groups at this frequency). Far fewer teachers assign these assignments daily or even weekly. Two teachers offered individual daily writing as an additional assignment. Another teacher added daily note taking and one other mentioned conducting literature circles once or twice a week.

In terms of in-class assessment, half of the teachers teaching language arts report only giving assessments once or twice a month at the most. The remaining half is equally split between once or twice a week and daily assessment. Figure F4 shows that short answers and essays are the most common types of assessment, followed by fill-in responses.

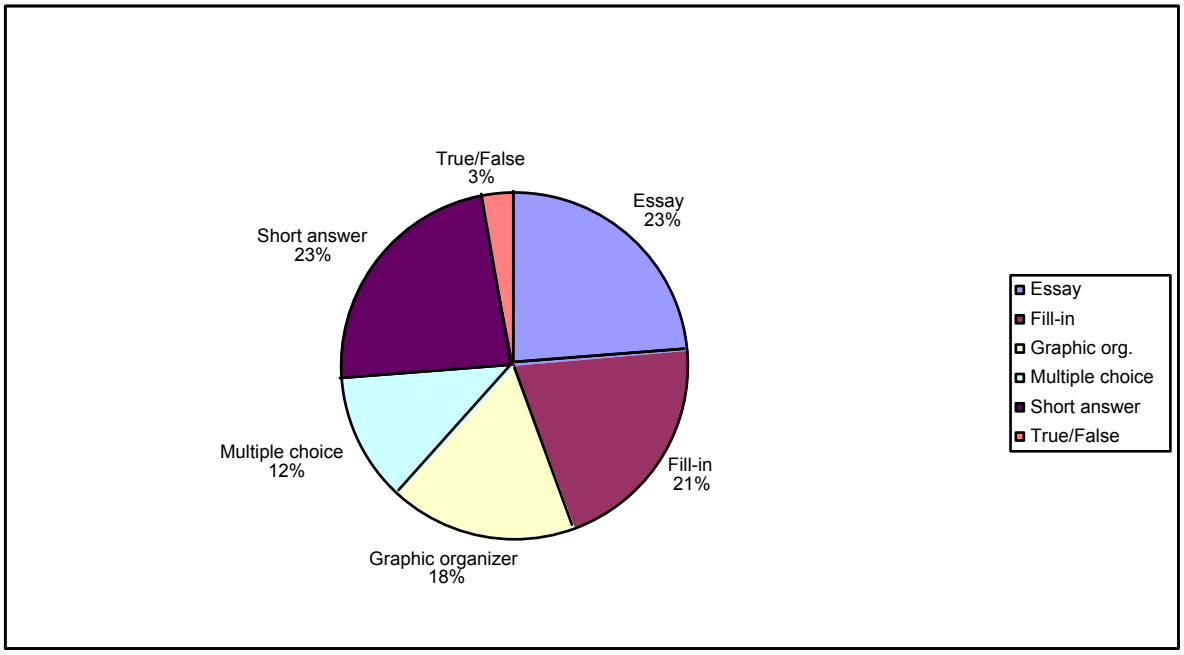


Figure F4. Language arts assessment items by type.

Summary of textbook and assessment use in the classroom. Overall the teachers across the four content areas showed a high degree of use and engagement with textbooks in their classes. Anecdotally, during the text review sessions, one of the teachers was so familiar with the style of her science textbook that she was able to accurately identify the title and publisher of one of the text selections. There were similarities and differences in patterns of textbook use, and in assignment and assessment practices across the content areas. These similarities and differences are

important to note so that the test development process can more precisely take account of both teacher and student familiarity with text types and tasks/item formats across content areas. Differences generally tend to distinguish mathematics from the other three content areas but occasional important differences also characterize science and language arts from the other content areas as well.

Appendix G:

Text and Item Review Forms

Developing Standard-based Measures of Academic English Language Proficiency

Focus Group Text Review Form
(2003-2004 School Year)

The purpose of this review is to determine if the texts selected are linguistically representative [typical] of texts used in subject area classrooms with native English-speaking students.

Reviewer information:

Name: _____ Grade: _____

School Name: _____

Text selection to be reviewed:

Subject Area: ❒ Math ❒ Science ❒ Social Studies

Selection Code:_____

Selection Title (if applicable):

_____

1. Is the selection representative of the textbooks you use in class? ❒ Yes ❒ No ❒ Somewhat

   Please provide the rationale for your response:

   _____

   _____

   _____

   _____

   _____

   _____

2. Is the selection representative of other materials you use in class? ❒ Yes ❒ No ❒ Somewhat

   Please provide the rationale for your response:

   _____

   _____

   _____

   _____

   _____

   _____

3. In reading ability, most of the native English-speaking students in my class are:

Below grade level	❒
At grade level	❒
Above grade level	❒
Not Applicable	❒

4. In reading ability, most of the English learners (i.e., students for whom English is a second language) in my class are:

Below grade level	❒
At grade level	❒
Above grade level	❒
Not Applicable	❒

5. Please rate this text in terms of difficulty for native English speaking students (Circle the appropriate number).

Difficult	1
Somewhat difficult	2
Appropriate	3
Somewhat easy	4
Easy	5

6. Please rate this text in terms of difficulty for English learners (Circle the appropriate number).

Difficult	1
Somewhat difficult	2
Appropriate	3
Somewhat easy	4
Easy	5

7. Please note any vocabulary items that you feel would present difficulties for:

   Native English speaking students

   _____

   _____

   _____

   _____

   English learners

   _____

   _____

   _____

   _____

8. Please note any grammatical features that you feel would require explanation for native English speaking students and/or English learners **directly on the text**.

   Directions: Use the YELLOW highlighter for native English speaking students, the BLUE highlighter for English learners, and the GREEN highlighter for both groups to indicate grammatical features requiring explanation.

9. Do you feel there are any sensitivity issues regarding this text (e.g., cultural or gender bias)?

   ❒ Yes ❒ No

   If your answer to the above question is yes, please explain.

   _____

   _____

   _____

   _____

   _____

10. Additional comments/questions:

    _____

    _____

    _____

    _____

    _____

Please staple the text selection you have just reviewed to this form after it is completed.

Developing Standard-based Measures of Academic English Language Proficiency

Focus Group Item Review Form
(2003-2004 School Year)

Reviewer information:

Name: _____ Grade: _____

School Name: _____

Item to be reviewed:

Subject Area: ❒ Math ❒ Science ❒ Social Studies

Passage: _____ Item number: _____

1. Have you ever asked your native English-speaking students to complete this type of item/task? ❒ Yes ❒ No

   If your answer to the above question is no, please explain why. _____

   _____

2. Have you ever asked your English learners to complete this type of item/task? ❒ Yes ❒ No

   If your answer to the above question is no, please explain why. _____

   _____

3. Please rate this item in terms of difficulty for native English-speaking students.

Difficult	Somewhat difficult	Appropriate	Somewhat easy	Easy
1	2	3	4	5

4. Please rate this item in terms of difficulty for English learners.

Difficult	Somewhat difficult	Appropriate	Somewhat easy	Easy
1	2	3	4	5

5. Please note any vocabulary words that you feel would require explanation directly on the item. Directions: Use YELLOW highlighter for native English-speaking students and BLUE for English learners, or GREEN for both.

6. Please note any grammatical features that you feel would require explanation directly on the item. Directions: Use YELLOW highlighter for native English-speaking students and BLUE for English learners, or GREEN for both.

7. Do you feel there are any sensitivity issues regarding this item (e.g., cultural or gender bias)? ❒ Yes ❒ No

   If your answer to the above question is yes, please explain. _____

   _____

8. Additional comments/questions: _____

   _____

   _____