An Approach to Operationalizing Academic Language for Language Test Development Purposes: Evidence From Fifth-Grade Science and Math

CSE Report 626

Frances A. Butler
CRESST / University of California, Los Angeles

Carol Lord
California State University, Long Beach

Robin Stevens, Malka Borrego, and
Alison L. Bailey
CRESST / University of California, Los Angeles

April 2004

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

Project 4.1 Developing Measures of Academic English Language Proficiency
Frances A. Butler and Alison L. Bailey, Project Co-Directors, CRESST/UCLA

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

The findings and opinions expressed in this report do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

AN APPROACH TO OPERATIONALIZING ACADEMIC LANGUAGE FOR LANGUAGE TEST DEVELOPMENT PURPOSES: EVIDENCE FROM FIFTH-GRADE SCIENCE AND MATH[1]

Frances A. Butler
CRESST/University of California, Los Angeles

Carol Lord
California State University, Long Beach

Robin Stevens, Malka Borrego, and Alison L. Bailey
CRESST/University of California, Los Angeles

Abstract

This report details an exploratory study that employs empirical methods to operationalize academic language (AL) for language test development purposes. First, an evidence-based approach to operationalizing AL is discussed, followed by a proposed methodology for systematically collecting and synthesizing AL research findings from two data sources—standards and textbooks at the fifth-grade level for science and math. In addition, we explore the potential for using extant classroom videos as an AL data source and present preliminary findings. We then begin to articulate a process for synthesizing AL data for test development purposes, specifically to inform specifications and prototype tasks. Findings include the use of language functions as an organizing structure for operationalizing AL. The research reported here will lead to a better understanding of the kinds of language students need for successful school performance, which will help in the development of English language tests for English language learners that are aligned to the language of the classroom.

Introduction

The No Child Left Behind Act of 2001 placed the assessment of the English language skills of students for whom English is a second language front and center in the educational arena in the United States. Under this law, English language learners (ELLs) must show measurable progress each year in English language

[1] The authors wish to thank Kathleen M. Bailey, Marguerite Ann Snow, and Joan Herman for their comments on earlier drafts of this report. In addition, the authors acknowledge the untiring contributions of research assistants Priya Abeywickrama and Francisco Herrera. Laquita Stewart, Danna Schacter, and Monica de Gyarfas provided project assistance at different stages of the research, and we thank them. Special thanks as well to Wade Contreras for final editing.

development (ELD). Consequently, states are scrambling to identify or develop tests of English language proficiency that can help them meet this federal mandate (Olson, 2002). One of the problems that has surfaced in the search for English language tests for K-12 ELLs is the inadequacy of existing instruments for assessing the type of English language skills students must have to be successful in the school setting. The content of existing language tests tends to be limited in that those tests generally tap more informal, social uses of English which, while important, are not sufficient for handling the full range of classroom language and the language of standardized content tests (Bailey & Butler, 2002; Bailey & Butler, in press; Bailey, Butler, LaFramenta, & Ong, 2004; Butler & Castellon-Wellington, 2000; Butler & Stevens, 2001; Stevens, Butler, & Castellon-Wellington, 2000). Thus tests are needed that more directly assess the language of school—academic language (AL), in this case, specifically, academic English.[2]

In order for test designers to produce AL tests, the construct must be described in sufficient detail to allow for the development of a framework and specifications that will yield appropriate tasks that reflect the language students must be able to handle in the classroom. The work described in this document is a continuation of a larger, ongoing effort at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) to articulate the construct of AL for broad educational application, taking an evidence-based approach (Bailey & Butler, 2002; Bailey et al., 2004). In addition to the development of AL tests, which are the focus of this report, CRESST envisions the application of the AL framework to curriculum development and professional development efforts as well. Indeed, the ideal scenario involves close linkage among the three applications—curriculum, testing, and professional development—with all students benefiting from an increased emphasis on, and attention to, AL across all content areas.

The work presented here is organized in the following way: first, an evidence-based approach to operationalizing AL is discussed, including the focus and goals for the current year, followed by an explanation of the empirical nature of our work. Next we present evidence from the two types of data sources studied in this segment of the research—standards and textbooks. We also present preliminary findings from a potential data source—extant classroom video. Within each section, we discuss the data sources and different methodologies used, and present results.

[2] We use the term *academic language* (AL) throughout this document to refer specifically to *academic English*, the focus of our work. Clearly, similar efforts could be carried out for other languages.

We then synthesize the findings across types of evidence and finally apply the findings to test development.

Evidence-Based Approach to Operationalizing Academic Language Proficiency

The current climate of educational reform and the passage of the No Child Left Behind Act of 2001 have moved the field of education toward scientific models of evidence-based research (Feuer, Towne, & Shavelson, 2002). The National Research Council (2002) reports on the nature of scientific inquiry in education and how the U.S. federal government can support scientifically based research. The call for evidentiary bases to educational research provides impetus for assembling a variety of sources to operationalize AL. In response, Bailey and Butler (2002) present an evidence-based research framework for operationalizing AL. Their approach identifies six sources of evidence that can help specify language demands in academic settings. The evidence comprises:

1. Review of empirical studies of ELL/English-only student performance and language demands of content and ELD assessments.
2. Examination of the language prerequisites assumed in national content standards.
3. Examination of the language prerequisites assumed in state content standards.[3]
4. Examination of the language prerequisites assumed in English as a second language (ESL) standards.
5. Documentation of teacher expectations for language comprehension and production.
6. Analysis of classroom exposure to AL[4], including teacher talk and textbooks.

The rationale for this approach is that in order to adequately describe the construct of AL, there is a need to systematically document the kind of language students are expected to understand and use for successful school performance. The types of evidence identified above provide (a) contexts of language use, including classrooms, activities, and materials such as textbooks, and (b) the specific types of language features and functions used within and across those contexts. According to Bailey and Butler (2002):

[3] Henceforth in the document, content standards are referred to as *subject-matter standards* to avoid confusion with discussions of test content.

[4] Classroom exposure to AL includes all aspects of the curriculum.

> The strength of an evidence-based approach to ALP [academic language proficiency] is that it provides a mechanism for capturing not just the linguistic features of language—vocabulary, syntax, and discourse and the features of language use within and across content areas—but also the linguistic demands created and/or assumed by a broader array of stakeholders. That is, the broader educational community provides [some] evidence of academic language in national content standards (e.g., National Science Education Standards of the National Research Council, National Council for the Social Studies), state content standards (e.g., California Department of Education, Texas Education Agency), and standardized achievement tests (e.g., the Stanford Achievement Test, the Iowa Tests of Basic Skills) (p. 5).

While all of the types of evidence may not prove to be equally useful in articulation of the construct, each type should be considered, at least initially, because of its contribution to the current overall U.S. educational environment. Also, the nature and quality of the information gleaned from each type of evidence will vary, and thus the different types of information will contribute to our goal in different ways. (See Bailey & Butler, 2002, for a detailed discussion of each type of evidence.)

Prior Empirical Research on Academic Language in the Content Areas

This work builds on multiple studies that have been conducted at CRESST to begin to characterize AL for the purpose of developing AL tasks. The initial work on the larger effort began with a focus on the use of test accommodations with ELLs on large-scale content assessments and the articulation of subgroups of ELLs on the basis of factors such as academic language proficiency and years in U.S. schools (Butler & Stevens, 1997). It became evident from that initial work that there would be long-term benefits from attention to the teaching and assessment of academic English. So while one strand of research at CRESST has continued with test accommodations (Abedi, Courtney, & Leon, 2001; 2002), we shifted our efforts to operationalizing AL. As evidence documented a mismatch between the language assessed by a language proficiency test and the language used in large-scale content tests (Butler & Castellon-Wellington, 2000; Stevens et al., 2000), researchers at CRESST began characterizing AL for test development purposes (Bailey et al., 2004; Butler, Stevens, & Castellon-Wellington, 1999; Stevens et al., 2000). Additional research on the impact of factors thought to influence ELL content-area test performance has also been investigated at CRESST (Abedi & Lord, 2001; Abedi, Lord, & Hofstetter, 1998; Abedi, Lord, & Plummer, 1997; Aguirre-Munoz, 2000; Bailey, 2000). From these efforts, relevant empirical data from the following evidentiary sources have been compiled. The data come primarily from schools in California.

Textbooks and Materials

Fourth- and fifth-grade science (Bailey et al., 2004)

Fifth- through eighth-grade social sciences (Butler et al., 1999)

Content Assessments

Seventh-grade social sciences (Stevens et al., 2000)

Eighth-grade math (Abedi et al., 1998)

Classroom Language: Observations and Interviews

Fourth- and fifth-grade science oral language observations (Bailey et al., 2004)

Across these research projects, an empirical foundation is being built upon which the AL construct can be operationalized.

Other research studies were also reviewed for their potential in contributing to the empirical evidence upon which the AL construct can be based. Few studies, however, use or describe empirical methods and approaches taken with enough specificity to meet the more vigorous standards implied by the current demand for research with a clear evidentiary basis. Specificity is needed to inform test development; without it, at the minimum, distinctions cannot be articulated between grade levels and content areas.

Some studies, however, provide a rich source of information with which we can compare our results, for example: Short's (1993) analysis of the language demands of middle school social studies texts and to a lesser degree classroom discourse; Coelho's (1982) analysis of secondary language demands in the classroom and description of the language features of geography, history, and science texts; and Dale & Cuevas' (1992) description of the general features of math vocabulary, syntax, semantics, and discourse. When possible, results of studies such as these will be included in the synthesis of available data in Section 6. We turn now to our current research focus and goals.

Current Research Focus and Goals

The work reported here builds on the previous CRESST research referenced above and continues the evidence-based approach with examination of (a) standards—national and state subject-matter and national ESL standards, and (b) science and math textbooks. In addition, a sample of classroom discourse from an extant video is examined to explore the feasibility of this type of data source for future research. The current research is concentrated at the fifth grade (or grade clusters which include fifth grade, e.g., 3-5, 4-6) in science and math. Fifth grade was decided upon as the starting point because instruction beginning in the third grade focuses more intensively on academic content, with reading as the medium of instruction, rather than focusing on the development of reading skills per se. Along these lines, Reppen (2001) has indicated that at fifth grade, students begin to control spoken and written language and are expected to use reading and writing to aid in their learning. Also, during previous grades, students are introduced to a variety of texts for different purposes, "so that by fifth grade, students [should] have some practice with different registers" (Reppen, p. 188). In the upper elementary grades, content becomes more specialized and the classroom discourse more cognitively demanding and linguistically complex.

We focused on standards and textbooks in science in order to complement existing oral language data (Bailey et al., 2004) and thereby be able to combine the language data from a range of sources to feed into language test development efforts. The goal is to provide a procedure for systematically collecting and synthesizing AL evidence from authentic sources of English language use across all content areas and grade levels to lead to a framework for developing test specifications. This document illustrates the procedure at one grade level in two subjects—science and math. The data are more extensive for science (see Figure 2 in Section 8 below), but the inclusion of math allows us to illustrate how similarities and differences across subject matter can be addressed in test development procedures. Later work will include language arts and social studies. The three questions below guided our effort. Since the work reported in this document will be continued in more depth, the questions are broad and will not be answered conclusively here. Rather they were intended to help shape the research and provide an organizing structure for looking at the data.

1. Will the proposed methodological approach provide a means for systematically collecting and synthesizing AL evidence for test development purposes?
2. What are the salient features of AL in fifth-grade science and math standards and textbooks?
3. Which of these language features are common to the two content areas of science and math and possibly other content areas?

In the long term, the information that answers these questions will provide the material necessary for the development of test specifications and AL tasks for assessing ELLs' academic English language proficiency. To the extent that we are able to identify and describe the language students need to reach academic goals across content areas, we will be able to evaluate their progress in acquiring AL. We turn now to the approach we have taken in this endeavor.

Empirical Approach

Through our empirical approach to characterizing academic language, we are looking at what students need to do with language to achieve a range of goals in the classroom. The most direct way to describe the language associated with these academic goals is to analyze the tasks and activities that students are expected to participate in and the associated language students need to complete the tasks. This approach has led to a focus on language functions. The term *language function* is used here to refer to the language students must understand and use to complete educational tasks. Typically such a task might be describing a scientific phenomenon, explaining an experimental result, summarizing a piece of text, or justifying one's opinion either orally or in writing. See Appendix A for a glossary of functions identified in this research.

The use of the term *function* is complicated by interdisciplinary differences in definition and application, as well as the ongoing tension between language and cognition. Bailey and Butler (2002), in presenting the evidenced-based research framework for operationalizing AL discussed above, state:

> While our focus is specifically on language, language is used for a purpose often to achieve a function, to explain, to interpret, etc., and thus is interwoven with cognition such that it is nearly impossible to exclude one from consideration of the other. For us the challenge is to create prototype ALP tasks that respect the interwoven nature of language and cognition but that have as their goal the assessment of language ability. (pp. 24-25)

In this research, we have followed a bottom-up procedure that allows the data to guide the analyses rather than first identifying features based on expert opinion and then performing the analyses. Figure 1 provides a visual representation of the research framework for the evidentiary bases we examined.

Type of Evidence:	Standards			Textbooks		Video Sample
Content Area:	Science	Math	ESL	Science	Math	Science
Type of Analysis:						
Organizational Features	n/a	n/a	n/a	X	X	X
Language Functions	X	X	X	X	X	X
Structural Features	n/a	n/a	n/a	X	X	X
Lexical Features	n/a	n/a	n/a	X	X	X

Figure 1. Research framework for evidentiary bases.

The three types of evidence in this work—standards, textbooks, and an excerpt from a classroom video—are shown in boxes in the top row of the figure. The second row indicates the relevant content areas covered for each type of evidence. The types of analyses conducted are listed in the first column on the far left of the figure. An X indicates that an analysis was conducted for a given type of evidence and content area.

In approaching the standards, we began by describing and categorizing the language needed to complete tasks implicit in each document. In our analyses of textbooks, we began by surveying and describing the organizational structure of the textbook chapters; language functions emerged from the reading and content tasks in these texts. Subsequent analyses then allowed for description of language structures associated with these functions and tasks. To characterize in some detail the structural and lexical features of the texts, more fine-grained analyses were performed. A similar approach was followed in analyzing the video sample to determine if the same methodologies could be applied to this type of extant data source. The following section describes the first type of evidence presented in this document, subject-matter and ESL standards.

Evidence From Standards Documents

As mentioned above, in the current national climate of increased accountability at all levels of public education, standards are viewed as providing targets for best classroom practices in content areas. Standards are aspirational in the sense that they represent attainment that state boards of education—and the teachers and content-area experts they convene—feel all students should strive to reach. Their impact is pragmatic as well since sanctions and rewards are now based on student progress tied to the standards. Because of the prominent role of standards in current educational dialogue, we felt they should be considered as sources of evidence in our current research framework. We investigated (1) national and state-level subject-matter standards because they help provide the foundation for subject-matter instruction and (2) the national-level ESL standards because they should articulate language demands that cut across subject-matter standards. This approach could help establish clear linkage between content and ESL standards. We focused our efforts this year on fifth-grade science and math to show how information from standards along with evidence from other sources can be synthesized to inform test development at a specific grade level.

Standards Data Sources

For each subject, sets of standards were developed by a number of different national organizations, and while no specific sets have been officially sanctioned by the U.S. Department of Education, the federal government provided support for the development of many of the national standards. We selected the *National Science Education Standards* (National Research Council [NRC], 1996) for our work in science because they were used in the development and alignment of the science textbooks analyzed here. For math, we selected the National Council of Teachers of Mathematics (NCTM) standards *Principles and Standards for School Mathematics* (2000) because NCTM was the first professional U.S. organization to develop standards. In essence, NCTM laid out the process for developing standards that has been followed by other organizations. Finally, for ELD, we examined the national *ESL Standards for Pre-K-12 Students* (Teachers of English to Speakers of Other Languages [TESOL], 1997), the only existing national standards for English as a second language.

In addition to national standards for science and math, standards from California, Florida, New York, and Texas for the same subjects were selected for analysis because these states have the largest ELL populations in the nation

(Padolsky, 2002). The state standards documents analyzed were published by the California Department of Education (1997, 1998), the Florida Department of Education (1996a, 1996b), the New York State Education Department (1996), and the Texas Education Agency (1998a, 1998b).

Generally, *academic content standards* describe what students should know and be able to do in the various content areas, and *performance standards* give "concrete examples and explicit definitions of what students have to know and be able to demonstrate" (Section 3, Goals 2000). The content standards we examined include "key ideas" for various topics along with specified behaviors or indicators[5] that describe how students will demonstrate learning. While there are general similarities such as these among the standards, both national and state, there are also differences with regard to methods of organization, format, and style. These differences, in part, reflect the different purposes for which the standards were written. National standards are intended to guide the development of state standards. They are largely conceptual and focus on processes of learning. States can build on the national standards by adapting them to meet their own goals, which tend to focus more on the specifics of subject-matter content. The ESL standards differ from both the national and state standards in their emphasis on language. They are structured around descriptors that focus on the features of language needed to complete different types of tasks, e.g., language functions, such as summarizing a piece of text or describing an experiment.

In addition to the differences mentioned above, the standards are organized differently across grades with some standards indicating expectations at a single grade level (California and Texas) while other national and state standards indicate expectations for grade clusters (NRC, NCTM, TESOL, Florida, and New York). Since our focus for the work reported here was at the fifth grade, we examined the fifth-grade standards for California and Texas, and the grade clusters 3-5 for NCTM and Florida, 5-8 for NRC and New York, and 4-8 for TESOL. Table 1 provides the standards data sources, grade or cluster, and specific content topics reviewed for science and mathematics in this report.

---

[5] Some states explicitly use the label *indicators* in their content standards (e.g., New York), while other states (e.g., California) do not. For the purpose of our analyses, we defined the word *indicators* as statements that describe learning objectives.

Table 1

Data Sources, Grades, and Topics Analyzed for National and State Subject-Matter Standards

Standards data source	Grade / Cluster	Topics analyzed: Science	Topics analyzed: Math
National	5-8 / 3-5	Science as inquiry	Geometry
California	5	Physical sciences	Measurement and geometry Statistics, data analysis and probability
Florida	3-5	The nature of matter Energy Force and motion	Measurement Geometry and spatial sense Algebraic thinking Data analysis and probability
New York	5-8	Physical setting	Measurement Uncertainty (probability)
Texas	5	Scientific processes	Geometry and spatial reasoning Measurement Probability and statistics

Since one of the purposes of the current research was to develop a methodology for systematic analyses of different data sources, only a small number of pages from the standards were selected for this exploratory research, with the aim of increasing the sample size in future analyses. To begin, we randomly selected a page from each standards source. So while the number of pages selected from each source is the same, the topics vary. For science, the language demands implicit in the tasks associated with the following topics were analyzed: energy, force and motion, matter, physical sciences, scientific inquiry, and the scientific processes. The math topics include algebraic thinking, data analysis, geometry, measurement, probability, spatial reasoning, and statistics.

Selection of ESL standards for analysis differed from selection of subject-matter standards because the purpose of the ESL standards is to delineate the language students need both in and out of the classroom. The TESOL ESL standards specify three goals for pre-K-12, each of which has its own set of three standards. Those goals are: (a) language competency in social settings, (b) language competency in academic settings, and (c) cross-cultural awareness. For our purposes, we selected one academic language standard for the grade cluster 4-8 (Goal Two, Standard Two), which indicates academic language needed across subjects. We move now to the method of analysis used to identify the potential language demands implicit in the language of the standards.

Method of Analysis for Standards Documents

Developing a methodology for analyzing the standards was complicated by differences referenced above across the standards used in the study. Since the emphasis of the present research is on the language students need to fulfill subject-matter requirements, we were interested in looking at performance indicators and potential student outcomes to evaluate the language demands. However, since standards differ in both the way they are organized and their approaches to describing student outcomes, a comparison of language demands was not straightforward. In some standards, performance indicators are explicitly stated; in others, the reader must use the standard to hypothesize what a learning outcome might look like. Table 2 highlights some of these differences in the language used and structure of the indicators from the standards analyzed for this work.

Table 2

Examples of Indicators From National and State Subject-Matter Standards

Standards data source	Grade / Cluster	Example indicators	
		Science	Math
National	5-8 / 3-5	"Students should develop the ability to refine and refocus broad and ill-defined questions" (NRC, 1996, p. 145).	Students "describe location and movement using common language and geometric vocabulary" (NCTM, 2000, p. 164).
California	5	"Students know all matter is made of atoms, which may combine to form molecules" (California Department of Education, 1998, p. 14).	Students "know the concepts of mean, median, and mode; compute and compare simple examples to show that they may differ" (California Department of Education, 1997, p. 22).
Florida	3-5	The student "understands the arrangement of planets in our Solar System" (Florida Department of Education, 1996b, p. 2).	The student "generalizes a pattern, relation, or function to explain how a change in one quantity results in a change in another" (Florida Department of Education, 1996a, p. 2).
New York	5-8	Students "explain daily, monthly, and seasonal changes on earth" (New York State Education Department, 1996. p. 32).	Students "develop critical judgment for the reasonableness of measurement" (New York State Education Department, 1996, p. 20).
Texas	5	"The student is expected to: describe and compare life cycles of plants and animals" (Texas Education Agency, 1998a, p. A-19).	"The student is expected to: use experimental results to make predictions" (Texas Education Agency, 1998b, p. A-21).

The examples in Table 2 illustrate that both national and state standards are broad in scope, leaving considerable latitude in application. To address the standards, teachers may employ a range of tasks or instructional strategies to facilitate student learning and assessment. The national standards—NRC (1996) science and NCTM (2000) math—emphasize broad areas of content knowledge with descriptions of ways students can demonstrate learning. The sample from the national science standard focuses on scientific inquiry, whereas the sample from the national math standard focuses on using the language of mathematics to describe spatial relationships. In contrast, examples from the state subject-matter standards focus on specific content topics. For instance, state science topics include atoms, the solar system, seasonal changes, and life cycles. Similarly, state math topics include data analysis, algebraic thinking, measurement, and probability.

Despite the fact that the standards are presented in these varying formats, they all include content learning statements, which are followed by either an indicator that describes student expectations or a statement of intended learning outcomes. We analyzed potential expected student outcomes for each indicator. Tasks the student might perform to show mastery of the standards were hypothesized. Using the hypothesized task as the basis, we identified language functions that would be embedded in the language students might produce to complete the task, either orally or in writing. Table 3 shows one example each from science and math; each includes a sample indicator, a hypothesized task, and language functions that correspond to the tasks.

In the New York science example, a potential task to demonstrate learning would be to plan and conduct an investigation of the night sky. The predominant language functions needed to complete the sample task include *description*[6], *explanation,* and *hypothesis.* A student would need to plan an investigation while hypothesizing findings. Then the student would describe the arrangement, interaction, and movements of the earth, moon, and sun. Also, a student would need to explain celestial phenomena and principles of relative motion and perspective. Similarly, the math example from Texas illustrates how a student would use investigation to demonstrate knowledge of probability. A student would plan and conduct a probability experiment and justify the results to make predictions and

---

[6] The functions identified in this report are presented in noun form for consistency with previous CRESST reports that discuss ALP. They are, however, defined in verb form in the glossary because we have drawn from recognized sources that usually define these terms as verbs.

Table 3

Standard Indicators With Hypothesized Tasks and Language Functions

Standard	Indicator from standard	Hypothesized task	Language functions inferred
New York science (Cluster 5-8)	"Students: explain daily, monthly, and seasonal changes on Earth" (New York State Education Department, 1996, p. 32).	"Plan and conduct an investigation of the night sky to describe the arrangement, interaction, and movement of celestial bodies" (New York State Education Department, 1996, p. 32).	Description Explanation Hypothesis
Texas math (Grade 5)	"The student is expected to: use experimental results to make predictions" (Texas Education Agency, 1998b, p. A-21).	Plan and conduct a probability experiment and use the results to make predictions.	Explanation Justification Prediction

explain probability. The dominant language functions include *explanation*, *justification*, and *prediction*.

As mentioned above, since there are many differences across the national and state standards, it was more difficult to hypothesize performance tasks that would demonstrate knowledge of the content for some standards than others. Also, since ways to measure mastery are numerous, ranging from limited response formats such as a true/false item to a performance task in which a student might be asked to *describe* chemical reactions orally or to *explain* in writing how atoms rearrange to form a product with different properties, the limitations in hypothesizing tasks are clear. All of the tasks and language functions recorded in the analysis are hypothetical in that they may never be employed by some teachers or classrooms but may be used frequently by others. Our examination of the other types of evidence may help confirm patterns reported here in the occurrence of language functions hypothesized in our work with the standards.

With these limitations in mind, a sample page from the subject-matter standards for both science and math was selected and analyzed by two raters. Then, the two raters together reviewed the functions that were assigned to each standard. If there were differences between raters in the functions assigned, the standard was discussed in detail and consensus was reached. [7] After the rating process was complete, a list of all the functions was compiled and the frequencies calculated.

The descriptors and progress indicators for the TESOL ESL standards that focus on AL use in the classroom were reviewed by three raters who each extracted a list of occurring language functions. The three lists were compared and consensus reached. Frequencies were not tabulated for the ESL standards because the focus of those standards is already on the language needed to complete a variety of different tasks across content areas. For example, in the TESOL indicator, "define, compare, and classify objects (e.g., according to number, shape, color, size, function, physical characteristics)" (TESOL, 1997, p. 87), a task explicitly stating functions is provided. It would be redundant to hypothesize other tasks and functions. We turn now to a discussion of the results of the standards documents analyzed.

Results and Discussion—Standards Documents

The analyses of the samples from standards documents focused on language functions that emerged from hypothesized tasks associated with the standards' indicators. Table 4 provides a summary of the language functions found in the sample pages analyzed.

A comprehensive list of all functions that emerged in the analysis is provided down the left side of the table. Then, in the next column, a bullet indicates each function that appeared in the TESOL ESL standard analyzed. For the national and state subject-matter standards, the number of instances each language function was identified in each source is included in the corresponding cell. Subtotals are provided in separate columns for both science and math. The total number of instances each function emerged in the subject-matter standards is recorded as a grand total in the last column on the right.

Overall, in our analysis of the sample pages from the standards, some language functions clearly emerged more often than others. A total of 26 different language functions were identified, of which six functions were dominant across the national and state subject-matter standards for science and math. They are, in order of

[7] Interrater reliability was not calculated for the analyses discussed in this paper, since a critical part of the research was to form and refine a methodology for articulating features of academic language. Rater training with established norms is needed to accurately calculate rater reliability. Reports of future research will indicate the level of agreement between raters.

Table 4

Summary of Language Functions From Standards Documents Analyzed

Language functions	TESOL ESL	Science						Math						Grand total
		NRC (5-8)*	CA (5)	FL (3-5)	NY (5-8)	TX (5)	Sub-Total	NCTM (3-5)	CA (5)	FL (3-5)	NY (5-8)	TX (5)	Sub-Total	
Analysis	•	6				2	**8**	3	4	10	8	2	**27**	**35**
Argument	•													
Classification	•					1	**1**	1					**1**	**2**
Comparison/ Contrast	•	2	2	7	3		**14**	5	3	2	1	1	**12**	**26**
Critique		1					**1**							**1**
Definition	•							1					**1**	**1**
Description		6	7	17	13	5	**48**	17	6	9	8	4	**44**	**92**
Enumeration			4	1	2		**7**							**7**
Evaluation	•					5	**5**							**5**
Explanation	•	7	4	9	8	7	**35**	5	3	8	9	2	**27**	**62**
Generalization				1			**1**			1			**1**	**2**
Hypothesis	•	3			1	1	**6**	2		1			**3**	**9**
Identification	•	4	3	5	2		**14**	9	1	2	4		**16**	**30**
Inference	•					1	**1**							**1**
Inquiry	•	4					**4**							**4**
Interpretation	•					1	**1**							**1**
Justification	•	4		1	1	2	**8**	1	1	5	5	1	**13**	**21**
Labeling										1			**1**	**1**
Negotiation	•													
Organization	•	4				2	**6**		1	1			**2**	**8**
Persuasion	•													
Prediction		2						1		1		1		**5**
Retelling	•													
Sequence	•			1			**1**							**1**
Summary	•	4				1	**5**							**5**
Synthesis	•					2	**2**			1			**1**	**3**
	Totals:	47	20	42	30	30	**169**	45	19	42	35	11	**152**	**322**

* Grade/Clusters

frequency with number of occurrences in parentheses: *description* (92 instances), *explanation* (62), *analysis* (35), *identification* (30), *comparison/contrast* (26), and *justification* (21). The other 21 functions were identified 9 or fewer times each.

In comparing the results of the analysis of the TESOL ESL standard with the national and state standards results, we found only four language functions (*argumentation, negotiation, persuasion,* and *retelling)* exclusive to the TESOL ESL standard analyzed. In contrast, the most frequently noted function in the national and state subject-matter standards, *description*, did not appear in the TESOL ESL standard analyzed. However, at this grade cluster, *description* does appear in the TESOL ESL standard that addresses social language.

In comparing the national and state subject-matter standards analysis results, we found a greater number of instances of language functions associated with the national standards overall than with individual state standards in both science and math. In the national standards samples analyzed, a total of 47 instances of language functions were identified for science and 45 for math. In the California standards, 20 instances were identified for science and 19 for math; in Florida, there were 42 instances for both science and math; in New York, 30 for science and 35 for math; and in Texas, there were 30 science and 11 math.

Last, in comparing the functions associated with science and math, a slightly larger number of total instances of language functions was observed for the science standards (169 instances) analyzed than for the math standards (152). This may be because participation in science classes can require a broader range of tasks, such as developing research questions, designing an investigation, synthesizing and summarizing data, and organizing information into tables and charts. Some math standards indicate tasks that are completely computational, thus no language functions were associated with those standards. In fact, many of the language functions that were recorded for the math standards were based on the assumption that a student must read and solve word problems, which might include performing a set of procedures, such as to *identify* the math operations needed to solve a problem, *justify* the choice of the math operations selected, and *describe* the procedures followed.

Although we have noted trends across the samples from the standards sources, especially with regard to the most frequently occurring functions, our work to date indicates that for the purpose of operationalizing academic English, evidence of language use from standards is primarily indirect. The standards allow us to hypothesize about the kind and level of language students need in order to show mastery of a particular standard, but in subject-matter standards there is often no explicit reference made to the language needed to demonstrate content mastery. So while standards can help inform our work in a general way by offering potential academic contexts for a broad range of language use, the language students must actually understand and produce can only be observed in classroom written materials and oral language. Thus, by looking at sources of language in use—textbooks, tests, and teaching—the impact of standards and their articulation of academic English can best be realized. With this in mind, we now turn to the analysis of the textbooks.

Evidence From Textbooks

We chose textbooks as one source of evidence for operationalizing AL because of their prominent role in the academic curriculum. While we acknowledge variation in the type and quality of textbooks available, the textbooks used in classrooms across subject areas should provide an important source of information for students on specific topics being covered. Sometimes textbooks are used to introduce material, while at other times they serve to reinforce material teachers have introduced through other means. No matter how they are used in individual classrooms, students must be able to process the information they present and carry out the activities or tasks within them that are assigned by teachers. As with the standards discussed above, we focused our textbook analysis efforts this year on fifth-grade science and math.

Textbook Data Sources

Since California schools were used previously for evidence of teacher talk in classrooms (Bailey et al., 2004), the decision was made to use textbooks approved by the California Department of Education for analysis in this phase of the research. Fifth-grade science textbooks from three different publishers and fifth-grade math textbooks from two of the same publishers were used: *McGraw-Hill Science* (Daniel et al., 2001), *Houghton Mifflin Science: Discovery Works* (Badders et al., 2000), *Harcourt Science* (Frank et al., 2000), *Harcourt Math* (Maletsky et al., 2002), and *Houghton Mifflin Mathematics* (Greenes et al., 2002). The textbooks were used to describe text features. It was not our purpose to compare the textbooks in any way. The specific topics chosen from each book and the type of analyses conducted are provided in Table 5 below.

Table 5

Topic Selections and Analysis From Fifth-Grade Science and Math Textbooks

Textbook	Topic	No. of pages	Type of analyses
McGraw-Hill Science (Daniel et al., 2001)	Weather	29	Functional, grammatical
	Matter	15	Functional, grammatical
	Plants	2	Lexical, structural
Houghton Mifflin Science: Discovery Works (Badders et al., 2000)	Weather	24	Functional, grammatical
	Matter	11	Functional, grammatical
	Cells	2	Lexical, structural
Harcourt Science (Frank et al., 2000)	Weather	25	Functional, grammatical
	Matter	10	Functional, grammatical
	Cells	2	Lexical, structural
Harcourt Math (Maletsky et al., 2002)	Ratio	38	Functional, grammatical
	Multiplication	1	Lexical, structural
Houghton Mifflin Mathematics (Greenes et al., 2002)	Ratio	54	Functional, grammatical
	Multiplication	1	Lexical, structural

The topics selected were of comparable length and subject matter across the textbooks. The number of pages reviewed for the topics differs because of the types of analyses conducted. The *matter, weather,* and *ratio* sections of the texts were used for broad analyses of functional language. The short selections (*plants, cells,* and *multiplication*) were used for exploratory structural and lexical analyses, laying the methodological groundwork for future analyses. We now turn to a description of the methodology used to analyze the texts.

Method of Analysis for Textbooks

Our purposes in analyzing selections from the textbooks were to characterize (a) the nature of the language students are expected to read and understand and (b) the language demands implicit in the tasks they are called upon to perform. The analyses that document the language features of the textbook selections are compared and summarized in Table 6.

To provide as complete a picture as possible of the language students must understand and use to process information in the textbooks reviewed, we analyzed

Table 6

Science and Math Fifth-Grade Textbook Analyses Performed

Analyses performed	Purpose
Review of textbooks	Describe textbook organization and features
Language functions in the textbooks	
(1) presentation of information	Identify functions exemplified in texts students must comprehend.
(2) directed activities	Identify functions students must use to perform tasks.
(3) grammatical features of functions	Describe the grammatical features that characterize the most frequently identified language functions.
Structural analyses of textbook selections	
(1) Textual and linguistic features	Describe features of typical extended texts.
(2) Descriptive features (a) average sentence length (b) average length of noun phrases (c) frequency of embedded clauses	Analyze selected features that help characterize a text.
Lexical analyses of textbook selections	
(1) Analysis of academic vocabulary	Identify two types of academic vocabulary in text samples: general and specialized.
(2) Lexical measures (a) frequency of words with three or more syllables (b) frequency of morphologically derived words[a] (c) percent of words that are low frequency according to a general corpus collection (d) percent of words that are low frequency according to a fifth-grade corpus collection	Apply lexical measures (a-d) to characterize the text and potentially to predict academic vocabulary identifications made by researchers in (1).

[a]A morphologically derived word is a word whose category changes when an affix is added, typically a suffix such as *–ation*, *-ly*, and *–ance*. Examples include *verb/noun* pairs like *observe/observation* and adjective / adverb pairs like *easy/easily*.

the textbook formats in order to describe the features of textbook organization in science and math and also to help provide an organizational method for finer-grained analyses. Specifically, we described the organizational structure of the textbook chapters. The language functions exemplified in the texts and implicitly required in the tasks emerged, and grammatical features of three frequently occurring functions were described. A description of the textual and linguistic features of the texts was prepared and an analysis of selected descriptive features, which was aimed at characterizing academic text, was conducted. Finally, lexical analyses were conducted for the purposes of developing a methodology for systematically identifying academic vocabulary and characterizing the lexical features of an academic text.

Textbook organization and features. As indicated above, the textbooks were examined to determine the general organization and chapter formats. Four researchers, working independently, reviewed the science and mathematics textbooks and identified three main pedagogical purposes: *forecasting*, *presentation of information*, and *directed activities*. Forecasting sections are intended to stimulate student interest and activate prior background knowledge. Presentation of information sections serve to communicate the content of the lessons, which may include information about objects, phenomena, events, processes, explanations, or descriptions of the skill or skills addressed in the lesson. Directed activities sections consist of activities and tasks based on new material recently presented, as well as review of previous material, and may include projects or experiments in science and word problems in math.

Functional language in the textbooks. In presentation of information and directed activities section types, we identified, respectively, the language functions occurring in the texts that students must process in order to understand the content and the language functions that students are called upon to produce in answering questions and completing tasks. Four researchers compared analyses on the set of functions that emerged. Definitions and criteria for identification of the functions were discussed and sharpened**.** (Reminder: see Appendix A for a glossary of the functions as they have been used in the current research.) The resulting list of functions was compared with those identified by previous researchers in their analyses of texts. The functions identified in the present research were similar to those noted in prior research (Bailey et al., 2004; Chamot & O'Malley, 1994; Coelho, 1982; Herber, 1978; Mohan, 1986; Short, 1993; and Vacca & Vacca, 1996), which gave us confidence in the initial identifications made. The frequency of each function in the presentation of information and directed activities sections was then calculated. See Appendix B for the procedures followed.

In the science and math textbooks, sentences in the presentation of information sections exemplifying the frequently occurring language functions of *comparison/contrast*, *description*, and *explanation* were identified and recurring sentence types and grammatical structures were described. In the directed activities sections, we identified sentences directing students to carry out an activity for which the expected response included the performance of one of the identified language functions. For each language function, we identified typical sentence types and grammatical structures.

Structural analyses of textbook selections. For both science and math, a representative sample text identified from the presentation of information section was first selected. Researchers then described the nature of the language in the sample to provide a summary overview of the type of language found in those sections.

Selected descriptive features of academic language in both science and mathematics textbooks were also analyzed. The purpose was to identify variables that might be useful not only for characterizing academic language but also for distinguishing it from everyday language, as well as to identify variables with potential for articulating differences among content areas and across grade levels**.** Our goal in this study was to pilot these measures using a small text sample, with the objective of expanding the sample size later.

To identify features that potentially contribute to the characterization of academic language, we reviewed previous research addressing sources of difficulty in text (e.g., Abedi & Lord, 2001; Abedi, Lord, Hofstetter & Baker, 2000; Abedi et al., 1997; Chall & Dale, 1995; Klare, 1974; and Zakaluk & Samuels, 1988). Previous conceptual and methodological approaches were reviewed, including Bailey, et al. (2004); Coelho (1982); Dale and Cuevas (1992); Halliday (1975); Pimm (1987); Richards (1978); Short (1993); Snow, Met, and Genesee (1989); Spanos, Rhodes, Dale, and Crandall (1988); Stevens, et al. (2000); and Wellington (1994). The features selected for analysis here were: (a) average sentence length, (b) average length of noun phrases, and (c) frequency of embedded clauses within a sentence. These features were selected because of their potential usefulness in describing the language used and in helping to determine the complexity and accessibility of a given text for readers (see discussion in Abedi et al., 1997). Appendix C describes the procedures followed for the three features.

Lexical analyses of textbook selections. Systematic identification of academic vocabulary presents a special challenge to researchers. Reliable methodologies for distinguishing academic vocabulary from non-academic vocabulary have not yet been developed, and there is a lack of empirical research on this distinction. In this segment of our research, the emphasis was on developing an approach for distinguishing between academic and non-academic vocabulary, and between general and specialized academic vocabulary, as well as characterizing the lexical features of an academic text.

As an initial step, the researchers adopted Chamot and O'Malley's (1994) broad definition of academic language as a point of departure: "the language that is used by teachers and students for the purpose of acquiring new knowledge and skills… imparting new information, describing abstract ideas, and developing students' conceptual understanding" (p. 40)—in short, the kind of English language skills needed for school success. Further, general academic vocabulary was defined as vocabulary words "used across multiple content areas" (Stevens et al., 2000, p. 11), for example, *function* and *system*; and specialized academic vocabulary was defined as "vocabulary unique to content areas" (Stevens et al., p. 11), for example, *hypotenuse* and *trapezoid*.

From text samples, four researchers, working independently, identified words that might be classified as general or specialized academic vocabulary items according to our working definitions. The results were then compared and consensus word lists generated. In addition, the following measures were selected in the effort to develop a methodology for characterizing the lexical features of academic texts: (a) frequency of words with three or more syllables, (b) frequency of morphologically derived words, (c) percent of words that are low frequency according to a general corpus collection (Zeno, Ivens, Millard, & Duvvuri, 1995), and (d) percent of words that are low frequency according to a fifth-grade corpus collection (Zeno et al., 1995). These measures were selected because they were judged to be among those most likely to show consistent differences between academic and non-academic texts. See Appendix D for the procedures followed in the lexical analyses.

Textbook Organization and Features

This section describes the results of the review of the textbook organization and features for both science and math. Each subject area is discussed separately below.

Science. The science textbooks we reviewed are organized into a series of sections and subsections called chapters, units, and lessons. As indicated above, in general, the science textbook chapter organization reflects three basic pedagogical purposes: (a) forecasting, (b) presentation of information, and (c) directed activities. Typically, parts of each chapter are devoted to each of these purposes.

Forecasting sections generally appear at the beginning of each chapter to introduce the topic, but also sometimes within the chapter to focus, orient, and guide the student through the material. In the chapters we analyzed, sections devoted to forecasting made up about 10% of the text. From the content of the forecasting text sections, we inferred that the sections were designed to help students form an expectation of the content and activities in the chapter; become curious and possibly speculate or formulate questions about the subject matter; relate the content to their prior knowledge; and then identify main points and begin to build a new schema or revise or extend an old one. Table 7 provides examples of forecasting from the science texts.

Table 7

Forecasting Examples From the Fifth-Grade Science Textbooks

Purposes	Sentence forms	Examples
Establish main idea, define the inquiry, or engage student curiosity	Declarative statement with explicit subject "you" and a future tense verb; or	"In Chapter 4 you'll learn about how weather is predicted." (Daniel et al., 2001, p. 147)
	Direct question, informal structure	"Do you know why water, a liquid, can also be a solid (ice) and a gas (water vapor)?" (Frank et al., 2000, p. C2)
Relate the topic to familiar contexts and / or help student associate new chapter content with previous knowledge	Declarative statement; or	"When you take ice cubes from the freezer of your refrigerator, you are removing solid chunks of water." (Badders et al., 2000, p. C4)
	Imperative inviting student to speculate or imagine	"Imagine that you are in a place that has no newspapers, radio, computer, or TV." (Badders et al., 2000, p. E55)
Suggest comprehension strategies or provide study guidance	Imperative with specific directions; or	"Look for these signal words to help you compare and contrast." (Badders et al., 2000, p. E94)
	Imperative guiding student to other parts of the textbook or other information sources	"See Science and Math Toolbox page H9 if you need to review *Using a Balance*." (Badders et al., 2000, p. C7)

The examples illustrate how purpose, context, and placement of the text help to shape the forecasting section. For instance, the second example, a direct question, could be used for assessment, but because of its placement at the beginning of the chapter it serves to establish the main idea of the chapter on matter and its properties. Many of the sentences used in these sections involve a simple imperative or declarative statement, or a question that is typically rhetorical. Because the forecasting sections were relatively small and their implementation uncertain, no analyses of the language functions were conducted at this stage of the research.

Roughly two thirds of each chapter reviewed is devoted to presentation of information in the form of expository text, in which information about objects, entities, phenomena, events, and processes is presented. A pedagogical feature typically found in the presentation of information sections reviewed is the presence of questions checking reading comprehension. They are typically short, direct assessment-type questions that check the student's comprehension of content in the preceding paragraphs, sometimes under a heading such as "check your understanding." Many ask about a simple fact given in the preceding text. Others ask for a description, and a few require an explanation (e.g., to provide a cause) or enumeration (e.g., to list). Textbooks vary in the relative emphasis on different tasks. One text in this study included nearly three times as many comprehension checks as another.

Approximately a quarter of the science textbook chapters reviewed consisted of suggestions for directed tasks and activities, intended for use as classroom learning activities or assessments, or for use as homework assignments. The texts call for students to represent information displayed in different modes (e.g., to describe information presented in a map or graph). Some of the activities require students to provide nonverbal physical responses as well (e.g., to carry out an experiment).

Mathematics. Selections from two fifth-grade mathematics texts were analyzed using procedures similar to those used for the science texts. The math textbook chapters are each divided into multiple lessons with the major focus on problem-solving practice. In general, chapters begin with a brief forecasting section, followed by a series of lessons, each of which includes a presentation of information section, consisting of a sample problem and solution, and a directed activity section, consisting of practice problems.

Chapters begin with a brief introductory segment we identified as the forecasting section. These sections serve to relate topics of familiar contexts to the student, introduce topics, and orient the student to new material. The forecasting sections have various formats. Some lessons begin with a one-sentence learning objective, while other lessons begin with a sample word problem, which includes a brief review of relevant knowledge or skills. Because of the relative brevity of the forecasting sections (approximately 8% of the total pages analyzed), we did not identify language functions occurring in them.

The presentation of information sections (approximately 30% of the total pages analyzed) are composed of lessons that begin with a sample story problem and solution, sometimes incorporating definitions of new terms. As with science, math textbooks also include questions that check student comprehension.

More than half of the pages analyzed (approximately 62%) are devoted to directed activities. Each lesson contained practice problems, cumulative tests and problem sets for review. Some problems require calculation and application of routine algorithms, typically numbers or symbols. For other activities the student might be required to recall factual information, concepts, or terminology. Word problems require the student to apply conceptual understanding to new data. At the end of each chapter there is typically an assessment section including a cumulative review of material from all the lessons. Directed activity sections may also include a review of skills and concepts from previous chapters, or may specifically address standardized test preparation.

Functional Language in the Textbooks

This section describes the results of the analyses of the language expectations implied in the textbook selections analyzed, including the language that students must process and the language that students are asked to produce. Then, grammatical features of three frequently occurring functions in the textbooks and standards analyses are discussed.

Functions identified in the science and math textbooks. Table 8 provides the findings of the analyses across the science and math presentation of information and directed activities sections.

Table 8

Instances of Language Functions: Fifth-Grade Science and Math Textbooks for Presentation of Information and Directed Activities Sections

	Science textbooks		Math textbooks	
Language functions	Presentation of information	Directed activities	Presentation of information	Directed activities
Classification	3	3		1
Comparison / Contrast	5	15		5
Definition	18	4	4	
Description	72	42	13	11
Evaluation		5		
Explanation	29	18	6	9
Hypothesis		5		
Inference		10		1
Interpretation		1		1
Justification		6		3
Labeling	12	3	1	4
Prediction		12		1
Sequence				3
Total no. of instances	139	124	24	39

The column down the left side of the table provides a comprehensive list of all functions identified in the pages analyzed. Then, in the next two columns, results of the science textbook analyses are reported, with separate columns for presentation of information and directed activities. The last two columns consist of the functions found in the presentation of information and directed activities sections of the math textbooks. For both science and math, the number of times a function occurs is noted. Finally, the total number of instances of language functions for each textbook section is provided in the row at the bottom of the table.

As mentioned above, in the science textbooks most of each chapter is devoted to presentation of information in the form of expository text. A total of 139 instances of language functions were identified in the presentation of information sections reviewed. In the directed activities sections, science texts consist mainly of questions and tasks, which were analyzed and categorized according to the types of linguistic responses that students would be expected to produce (i.e., the language functions).

In the pages analyzed for directed activities, 124 instances of language functions were identified. The most frequent language function, for both presentation of information and directed activity sections, is *description*, followed by *explanation*.

In the math textbooks, texts in the presentation of information sections are mainly composed of questions, declarative statements, and sample word problems with solutions. Twenty-four instances of language functions were identified in those sections. In the directed activities sections, language functions implicit in the tasks were analyzed. A total of 39 instances of language functions were identified in the pages reviewed. As with science, the most frequently occurring language functions identified in the math pages analyzed are *description* and *explanation*.

Overall, two functions emerged as predominant across the textbooks, *description* and *explanation*, regardless of whether a student is expected to read and comprehend text or complete a task. However, the frequency of other functions depends on whether the presentation of information or directed activities section is being considered. For the presentation of information sections in the science and math textbooks, *definition* was the third most frequently occurring function, whereas *comparison/contrast* was infrequent or non-occurring; in the directed activities sections, *comparison/contrast* was third most frequent, whereas *definition* was infrequent or non-occurring in the pages analyzed. Although these results are preliminary, they indicate that there may be differences in the functional language students are expected to comprehend as opposed to the language they may be required to produce.

Grammatical features of three frequently occurring language functions. Drawing on the results of the identification of functions in the textbooks and also in the standards documents, three frequently occurring functions across the two sources were selected for further review. *Description* and *explanation* were the most frequently noted functions across the evidentiary sources. *Comparison/contrast* was the third most frequently noted function in the directed activities language of the textbooks and was also identified in the standards as a function that occurs frequently. The same selections of texts used above in the analysis of the functions (i.e., the long selections shown in Table 5) were used to describe the grammatical features that appeared in text samples that contained the three functions. These preliminary descriptions of grammatical features of language functions are presented in alphabetical order below.

Comparison/contrast. Information on the grammatical features of the language function *comparison/contrast* in the presentation of information sections came only from science. No data are included in this part of the discussion for math because the function did not appear in the presentation of information samples reviewed. In our science samples, *comparison/contrast* usually occurred within the context of a larger description. English includes a variety of comparative structures. Many were identified in our fifth-grade science samples. Adverbial comparisons of process included the *more* and *–er* forms (e.g., *more quickly than* and *faster than*) as in "During the day, the land heats up more quickly than the water does…" (Frank et al., 2000, p. B41). Comparative adjective forms (e.g., *higher than* and *gets colder than*) and quantity comparisons (e.g., *more water than*) were also typical. Other patterns (e.g., the use of an introductory word such as *instead* to contrast with a previous statement) occurred occasionally.

Comparison/contrast within the directed activities sections of science and math showed marked similarities in sentence structure patterns across the two. In setting up directed activities, both questions and imperative statements were used. In instances where a similarity or difference was assumed, a *how* question with an adjective form such as *related to* might be used to prompt the student to provide specific information about a process or phenomenon (e.g., "How is a tornado related to a thunderstorm?" [Daniel et al., 2001, p. 173]) In other instances, the student might be directed to *compare* a process or phenomenon (e.g., Compare the two circle graphs). Another typical pattern mirrored the comparative *more* and *–er* patterns seen in the directed activity section for science—a *which* question with an adjective form (e.g., "Which material—soil or water—heats up faster? "[Frank et al., 2000, p. B38).

Description. For the language function *description*, the presentation of information sections in the science texts in this research used a variety of sentence formats to describe processes, events, and phenomena as well as objects. Most sentences contained multiple clauses, though some single-clause sentences occurred as well. The sentences were evenly divided between active transitive and intransitive structures. There were a few predicate adjective and predicate nominal structures (e.g., "Cirrus clouds are high clouds made mostly of ice crystals" [Frank et al., 2000, p. B16]), and a few passive voice clauses (e.g., "…when warm, humid air is pushed into the atmosphere" [Frank et al., 2000, p. B52]). Almost all of the sentences were in the simple present. Descriptions of physical objects provided details on physical attributes such as size, appearance, and composition and were typically of the form: X *is/are* Y where Y might be either a noun or an adjective (e.g., “Made of tiny pockets of air surrounded by thin walls of silica, aerogels are nearly transparent” [Daniel et al., 2001, p. 207]). In addition, analogies were occasionally used in describing a process or object (e.g., “Tornadoes are…like deadly whirling brooms. They can sweep away anything in their paths” [Daniel et al., p. 160]).

The presentation of information sections of the math textbooks shared the use of the X *is/are* Y form with science (e.g., “The ratio of dog books to the total number of books he checked out is part to whole” [Maletsky et al., 2002, p. 542]). However, these sections in the math textbooks tended to focus on providing sample problems, which served as descriptions by taking the student through the steps required to solve the problem. Imperative sentences typically served to describe the steps (e.g., “Count the small squares in the figure” [Greenes et al., 2002, p. 530]) and have, thus, been identified as serving a descriptive meta-function in the presentation of information sections in math.

The directed activities sections in science mirrored the imperative forms of the presentation of information sections in math. Whereas the imperatives in math described a computational process for the student, the directed activities imperatives in science asked the student to do the describing, generally of a physical process, phenomenon, or object (e.g., “Write a description of how a thunderstorm forms” [Daniel et al., 2001, p. 163]). Typically students were asked to write a description of something or record what happened to something. In the directed activities sections in math, students were also asked to describe, in this case, typically a computational or problem-solving process, but the structure used to set up the mathematical descriptions generally involved the question word *how* (e.g., “How would you show 100% on a 10-by-10 grid? [Greenes et al., 2002, p. 571]).

Explanation. In the presentation of information sections in science, *explanations* were most frequently given through the use of the main verb *cause* and an infinitival clause where X and Y were noun phrases: “X causes Y infinitive phrase” (e.g., “Winds blow into the low-pressure areas, and the rotation of the Earth causes them to spiral around the low” [Frank et al., 2000, p. B54]). In the same sections in the math textbooks examined, the explicit use of *cause* was virtually non-existent. Rather, the text often described a procedure that the student had to perform in order to accomplish an objective. The description gave the student detailed information that could be used to solve the problem. In this form, the procedure led to accomplishing the objective; the sentence describing the relationship between the procedure and the objective was regarded as an *explanation* (e.g., "You can use scale drawings and map scales to determine actual sizes and distances" [Maletsky et al., 2002, p. 548]). In a different form, the objective was stated first, followed by the procedure (e.g., "You can find a percent of a number by changing the percent to a decimal or a fraction and multiplying" [Maletsky et al., p. 565]).

In the directed activities sections of the science texts, students were typically asked to give reasons or causes for generalizations or physical phenomena by way of explanation. The use of *why* questions was dominant with many containing an implicit factual claim (e.g., "Why do you think predicting a severe storm is so difficult?" [Daniel et al., 2001, p. 173]). Other questions asked the student to provide the cause for some phenomenon (e.g., "Why do hurricanes form over oceans?" [Frank et al., 2000, p. B57]). Some questions were embedded after an imperative verb such as *explain* [e.g., "Explain why forecasting the weather can be a difficult job" [Greenes et al., 2002, p. E63]). In the math texts, students were often asked to support or give reasons for choices in problem solutions or mathematical representations. As in science, direct *why* questions were asked (e.g., "Why is it easier to compare decimals than it is to compare fractions?" [Greenes et al., p. 538]). In addition, the imperative *explain* was used with *why* (e.g., "Explain why the order in which you write a ratio is important" [Maletsky et al., 2002, p. 542]). The imperatives *explain* and *explain why* were also used after an information question or a statement (e.g., "Would you use a fraction or a decimal to find 25% of 120? Explain" [Maletsky et al., p. 575].).

Discussion. It is important to note that while the grammatical features for *comparison/contrast*, *description*, and *explanation* have been discussed separately for each function in both science and math, in general, science texts especially consisted of integrated presentations in which one or more language functions were embedded in another function. For example, the sentence, "During the day, the land heats up more quickly than the water does, so the breezes blow from the sea to the land" (Frank et al., 2000, p. B4), describes and compares as well as explains. It describes the direction of the breezes; it compares the rate at which land and water heat up; and it utilizes the *description* and the *comparison* to provide a reason, an *explanation*, for the phenomenon described. In our analyses, we have examined each function on its own in an attempt to characterize the most salient features for test development purposes, but we have not lost sight of the interrelated nature of the functions in both oral and written texts. We will further explicate the features of the three functions discussed here, and others that have been identified as well, in matrices and tables in our test development documents. In addition, we will draw on the interrelationship of the functions in the development of our academic language tasks. We turn now to a discussion of the structural analyses of the textbook selections.

Structural Analyses of Textbook Selections

Selections of representative text from the presentation of information sections were reviewed for both science and math; textual and linguistic features of typical texts from each are first described below. Three measures that help to characterize features of a text were also applied, as outlined above. We give the results here to illustrate the methodology but with the caution that they were based on a very small sample size.

Textual and Linguistic Features

Science. A typical presentation of information text sample in the fifth-grade science textbooks used in this study is an expository, multi-paragraph passage consisting of short paragraphs between two and seven sentences in length, such as in the following sample text.

> **Severe Storms**
>
> *Thunderstorms*
>
> About 2000 thunderstorms are taking place on Earth at any given moment. A **thunderstorm** can be a very strong storm with a lot of rain, thunder, and lightning.
>
> A thunderstorm begins to form when warm, humid air is pushed high into the atmosphere. As the warm air is pushed upward, it begins to cool, and a cloud forms. Soon, the weight of the condensed water vapor becomes too much for the air to support. The water falls to the ground, pulling cool air with it.
>
> Electric charges build up in the cloud. The charges increase until they are so strong that electricity travels through the air as lightning. It may travel between parts of the cloud or between the cloud and the Earth's surface.
>
> The air along the path of a lightning bolt is heated to temperatures that can be greater than 28,000°C (about 50,000°F). This intense heat makes the air expand so fast that the shock waves make the sound of thunder.
>
> Most thunderstorms are over within an hour. The precipitation and cool air moving downward through a thundercloud stop more warm air from moving up into the cloud. Sometimes, however, the cool air rushing down to the Earth's surface pushes more warm air upward to form another thundercloud. (Frank et al., 2000, pp. B52-B53).[8]

[8] Pages B52-B53 are from HARCOURT SCIENCE, Grade 5, Pupil Edition. Copyright © by Harcourt, Inc. Included by permission of the publisher.

Sentences in the presentation of information example above are primarily statements of fact that describe the phenomenon of a *thunderstorm*. The topic is concrete; the language largely descriptive. The vocabulary is a combination of familiar, commonly used words such as *rain*, *thunder,* and *lightning* and more technical, scientific vocabulary such as *condensed water vapor, electric charges,* and *shock waves*. The effect is often an extended definition. There is frequent use of multiclausal sentences, as well as prepositional phrases and participial modifiers. Present tenses are prevalent, typically simple present with the occasional progressive form. Some modal verbs (e.g., *can*, *may*) appear, though infrequently. Active voice predominates, but some passive constructions are used as well.

Math. A typical presentation of information section in the fifth-grade mathematics textbooks used in this study consists of a word problem with an interrogative and a solution. The purpose of the word problem is typically to introduce the concept being taught.

Equivalent Ratios

Learn About It

After you choose a paint color when buying paint, the clerk mixes the paint by adding small units of color to a base color.

Henry has chosen a color that requires 2 units of blue for every 6 units of yellow. If the clerk uses 6 units of blue for Henry's order, how many units of yellow will he need?

You can find an **equivalent ratio** that has 6 as its first term.

Different Ways to Find Equivalent Ratios

You can find an equivalent fraction.

Step 1
Write the first ratio as a fraction in simplest form.

Step 2
Find a number to multiply that gives 6 as the new first term.

Step 3
Multiply the numerator and denominator by 6. (Greenes et al., 2002, p. 520).[9]

---

[9] From HOUGHTON MIFFLIN MATHEMATICS, Level 5. Copyright © 2002 by Houghton Mifflin Company. Reprinted by permission of Houghton Mifflin Company. All rights reserved.

The word problem above provides a scenario that leads to a question, with a conditional ("If…") clause, to be answered by performance of a mathematical operation. Most verbs are in the simple present tense with one use of perfect and a use of the modal *can*. Voice is active. The vocabulary consists primarily of familiar everyday words, though a new concept, *ratio,* is introduced. The new term is bolded. Following the problem statement, steps leading to a solution are illustrated.

Descriptive Features

As mentioned above, the structural analyses for this initial stage of research were performed on small samples of text (1-2 pages per selection) from the science and math textbooks and included three measures: average sentence length, average length of noun phrases, and embedded clauses as a percent of total clauses. The fifth-grade science textbook samples had a total of 2,359 words and 202 sentences. The average sentence length was approximately 12 words. The average length of noun phrases was two words with a total number of 667 noun phrases. In the science textbook samples, embedded clauses made up 36% of the total number of clauses (317).

The math samples had a total of 981 words and 130 sentences. The average sentence length was approximately eight words per sentence. The average length of noun phrases was approximately two words with a total number of 251 noun phrases. Embedded clauses made up 17% of the total number of math clauses (161).

Lexical Analyses of Textbook Selections

Researchers identified general and specialized academic vocabulary items in the brief selections used in the structural analyses above. As expected, some of the same lexical items were found across the texts since the topics were the same. For example, all three science textbooks contained the general academic words *basic, function, system,* and *waste.* In terms of specialized academic vocabulary, the samples included *cell, organ,* and *tissue.* In the science samples, approximately 7% of the total number of words (2359) were identified as specialized and 10% were identified as general or nonspecialized academic vocabulary. In the math samples, approximately 3% of the total number of words (981) were identified as specialized math vocabulary and approximately 9% were identified as general academic vocabulary.

As mentioned in section 4.2, four independent measures were applied to the texts to help characterize the lexical features. For fifth-grade science, on average, approximately 8% of the words contained three or more syllables. Approximately

2% were morphologically derived. According to the general word frequency list (Zeno, et al., 1995) used in this study, which includes all grades first through college, low-frequency words varied across samples from about 6% to 14% of the total number of words on a frequency list for fifth grade, and from 8% to 15% were identified as low-frequency words from a general word list.

On average, approximately 4% of the words in the math textbooks had three or more syllables, and 1% of the words were morphologically derived. The percentage of low-frequency words varied across the math samples from about 5% to 17% when looking at fifth grade only, to 11% to 28% when looking at the complete list.

We investigated the extent to which these lexical measures might identify the same words as those identified by the researchers as academic vocabulary—that is, whether independent measures might produce results similar to those of the researchers' selections. One complicating factor in identifying academic vocabulary is the issue of multiple word meanings. For example, in addition to its more common meaning, *body* is used in academic contexts differently, as in *a body of information*. Although a word such as *body* is concrete when used in its everyday context, its academic context may be conceptually abstract, possibly making it difficult for ELLs to draw meaning from the word in a new context unless it has been explicitly taught (Scarcella & Zimmerman, 1998; Stevens et al., 2000). Since the frequency list we used is not sensitive to word senses, this subtlety is not reflected in the data. Ongoing investigations are evaluating the methodology for reliably identifying general and specialized academic vocabulary by using these lexical measures.

Discussion—Textbooks

In our analyses of selections from fifth-grade science and math textbooks, we looked at textbook organization, language functions, the structural features of texts, and academic vocabulary. These analyses were all performed to help provide descriptive detail about AL that would support test development. The initial analyses were broad and provided an overview of AL use in textbooks. Data reduction allowed us to continue to focus our attention more selectively on the salient features of language that would be important to assess in an ALP test.

First, analysis of textbook organization provided characterization of the pedagogical purposes for which language is used and thus provided a context for further analysis. Three pedagogical purposes were identified—forecasting, presentation of information, and directed activity—but only the latter two figured in subsequent analyses. Second, a range of language functions was identified in both of those two sections for science and math with marked overlap of six functions: *classification*, *comparison/contrast*, *definition*, *description*, *explanation*, and *labeling*. (See Table 8.) Of the six, *description* and *explanation* occurred in both sections in both science and math. The results of these analyses suggest candidates for inclusion on any test design effort for ALP tests. We looked at grammatical features associated with specific functions. Three of the six frequently occurring functions were selected for analysis—*comparison/contrast*, *description*, and *explanation*. Similarities in features across the same functions in science and math were noted. Patterns of occurrence will be taken into consideration in subsequent test development steps. In addition, the frequent embedding of functions within functions was noted for test development purposes.

Next, structural features of texts were identified to help establish criteria for text selection in the task development process. Textual and linguistic features were described for one representative presentation of information text each from science and math, and the three measures used to help characterize the features of texts (average sentence length, average length of noun phrases, and embedded clauses as a percent of total clauses) were applied. The data these initial analyses provide will serve as a starting point for future work.

Finally, preliminary vocabulary analyses were conducted on the same selections used for the structural analyses. General and specialized academic vocabulary items were identified by the researchers. Four independent measures were then applied to the texts to help characterize the lexical features and to help determine whether independent measures might be used in the future to specify the two categories of vocabulary. As with the other analyses conducted on the textbook selections in this work, the vocabulary analyses must be replicated on larger samples before any findings can be generalized.

Using Extant Video Recordings As a Data Source for Classroom Language

Previous work by Bailey et al. (2004) investigated oral language use in fourth- and fifth-grade science through a series of classroom observations. While this research was productive in providing an initial overview of academic language use at those grade levels and in one content area, a problem inherent in doing classroom observations is that inevitably the researcher cannot record everything seen or heard or may perhaps miss something important in the course of making notes about something else. To more thoroughly document the language functions and forms that emerge from classroom talk and to have more confidence in the accuracy of research findings, it is necessary to use other approaches, such as analyzing videotaped recordings of lessons, that allow researchers to revisit parts of a lesson in question and look more carefully at the language used.

Thus, for this study, we decided to explore the use of extant video recordings of classroom interaction as a data source for oral language and also for the purposes of developing a reliable and replicable methodology for empirically identifying academic language used by teachers and students. The data reported here are exploratory and limited in scope to a single activity in a lesson from a fifth-grade science class with native English speakers as well as English language learners. The purpose of reporting the work is to illustrate a process for examining academic language in use in the classroom.

The video selected is a model for training teachers; we chose a lesson that would provide samples of academic language used in the classroom at the grade level and in one of the subject areas analyzed in the standards and textbooks.

Video Sample Data Source

The language sample we reviewed is a ten-minute extract from a longer science lesson taught by a science specialist at a school in a large school district in California. It was professionally edited (e.g., a segment with classroom management activities was edited out) and includes narration and voice-over explanations by the teacher.

The selection shows a lab activity in which students are guided to observe and analyze fingerprints in order to solve a crime. The lesson demonstrates the scientific method and includes a discovery process whereby students make observations, record them, and then describe their findings in their own words. The teacher first introduces the activity and then guides the students in describing the fingerprints, grouping fingerprints according to similar characteristics, comparing the fingerprint evidence to potential suspects, and in making inferences about the identity of the robber.

Method of Analysis for Video Sample

The first step in the analysis was transcription of the video. The file was then converted into the format used in the Child Language Data Exchange System

(CHILDES), the Computerized Language Analysis (CLAN) program format (MacWhinney, 2000). The utterances, identified as intonation breaks within an individual speaker's turn, were numbered and counted. Analyses were conducted on instances of teacher/student talk only, so teacher voice-over narrations were edited out of the transcript. Of the remaining 124 utterances, 88 were teacher talk and 36 were student talk. Since this was a teacher-guided activity, there was too little student talk to analyze.

The transcript was analyzed in the same way as in the textbook analyses, according to (a) the language functions exemplified in the teacher's talk, and (b) the nature of the linguistic responses expected from the students during their participation in the classroom activity. In addition, the teacher's talk was further analyzed using the same set of lexical and structural measures used for the textbooks to determine if the same methodologies could be used reliably for future analyses of oral classroom language (see Section 4.2). The utility of using these procedures is evaluated at the end of this section.

Video Sample Findings

The pedagogical purposes identified in the video sample were forecasting, directed activity, and classroom management. A brief initial segment (4 of the 124 utterances in the transcript) consisted of forecasting (i.e., telling the students what was going to happen): *"First we're going to…. Then we're going to…."* The major part of the lesson (115 utterances) was devoted to a teacher-guided hands-on lab activity. At the end of the lesson was a brief coda (5 utterances) in which the teacher complimented the class on their good work: "*That was a great activity today*…. *Let's give ourselves a hand."*

Language use and functions. Most of the teacher talk (77% of the transcript) consisted of guiding the lab activity. The guidance consisted primarily of instructions for carrying out the activity that employed linguistic features such as imperatives and questions. Directions were sometimes given using imperatives ("Take a close look…"; "See how it…"), but more typically indirectly ("I want you to…"; I'd like you to tell me…"; See if you can…"; "What you and your partner are going to do is…"; "You need to…"). Questions were sometimes direct requests for factual information ("Would that fit…?"; "Does this one have…?"; "Who would like to…?"). However, often they were less direct, serving to plant ideas or encourage reflection (“Do you agree that…?”; “Do you think…?”; “Are there any reasons why…?”; “How many people think…?”; “So can we really conclude…?”).

During the directed lab activity, the teacher’s talk demonstrated two language functions: *description* and *labeling*, with five utterances of the transcript devoted to *description* and six utterances devoted to *labeling*. In this lesson, the teacher led the students in describing, comparing, and classifying fingerprint patterns, and in drawing inferences about who committed the crime. Accordingly, the language demands placed on the students as a part of completing these activities and tasks were, in the order from most to least frequent (as measured by number of teacher’s utterances eliciting these responses): *inference (10), classification (9), comparison/contrast (4), description (4), explanation (2),* and *evaluation (1).*

An additional feature of the teacher talk was the use of repair strategies, specifically *repetition* and *paraphrasing*. The teacher would sometimes respond to a student’s comment by repeating or paraphrasing it. In addition, the teacher also requested further *clarification* from a student.

Structural analyses. For the textbook analyses, our measures included average sentence length, average length of noun phrases, and percent of embedded clauses. With regard to average sentence length and the distinction between main clauses and embedded clauses, we found that, as studies of oral language have shown (e.g., Chafe 1988, 1994; Chafe & Danielewicz, 1987), analysis in terms of sentence units was not appropriate for spoken interaction. As these researchers suggest, oral communication is organized around intonational units rather than sentence structures. Additionally, as shown by Ono and Thompson (1995), it is often not possible to analyze oral data reliably into clear patterns of hierarchical embeddings within sentences. Therefore, in the video transcript, we only computed the average length of noun phrases. In this exploratory analysis of an excerpt from a single lesson, we identified 232 noun phrases with an average length of 1.4 words in the 88 teacher utterances. Future analyses in a similar vein will help us determine if this number and length of noun phrase is typical for a lesson of this type.

Lexical analyses. Regarding the use of academic vocabulary, we followed the same procedures used to analyze vocabulary in the textbooks. First, we used a set of criteria to identify general and specialized academic vocabulary words used in the teacher talk. Then we applied the four lexical measures used for the textbook analyses. Approximately 8% of the total number of words (717) had three or more syllables, and approximately 1% of the words were morphologically derived. According to the general word frequency list, as well as a frequency list for fifth grade (Zeno et al., 1995), approximately 12% of the words were low frequency.

Almost all words identified initially by the researchers as academic vocabulary were also identified using one or more of the four lexical measures, although the measures sometimes flagged words that were not identified by the researchers. A total of 32 words in the transcript were identified by the researchers as academic: 27 general and five specialized. An additional seven words were flagged by the lexical measures but not identified by the researchers: *grouping, lesson, looped, probably, robbery, robbing,* and *wavy*.

Discussion—Video Sample

The results presented in this section are drawn from only one video sample, and thus are preliminary. However, they do illustrate that a video sample offers researchers the opportunity to analyze functions used by teachers and expected in student responses, and academic vocabulary use. Furthermore, the analyses show that some of the same procedures used in analyzing language in textbooks can be applied to classroom discourse, yielding the same type of data sets.

Synthesis

In this section, we first revisit the questions that guided the research undertaken; then we synthesize the findings of the current research with findings from prior research that are relevant to the current test development effort.

Revisiting the Research Goals

In Section 2, three research questions were raised. The first emphasized the development of a methodological approach for the collection and synthesis of empirical data; the other two focused more specifically on the goal of operationalizing academic language for language test development applications. Below we discuss our findings for question one and then present preliminary results for questions two and three.

> Question 1: Will the proposed methodological approach provide a means for systematically collecting and synthesizing AL evidence for test development purposes?

An overarching goal in this stage of the research was to establish a methodology for collecting and synthesizing data from a variety of evidentiary sources. The initial steps toward the development of the methodology documented in this report were at times challenging and often frustrating. Because no guidelines exist for systematically extracting or describing language use across data sources for test development purposes, we drew from previous work in related areas such as discourse analysis and reading research. The research reviewed took varying methodological approaches; for example, some compared the language of textbooks with classroom discourse, while others described linguistic features that potentially impact performance on large-scale assessments. Since our purpose differed from the purposes behind the approaches we reviewed, we found little guidance in selecting appropriate measures, operationalizing terminology, applying measures equally across content areas, and so on. However the current research has provided us with a foundation for further research through the development of procedures and definitions that can be applied across a variety of data sources.

In all cases, our first task was to review and describe the organization of information in each evidentiary source. Then, using information about the formats, we followed a bottom-up procedure to analyze the data, documenting procedures and developing definitions as we went.

Specifically, regarding the methodological approach to the standards, the standards were analyzed according to expected outcomes only, because language expectations for each subject area were not explicitly stated in any of the standards selections we reviewed except the ESL standard. Since the primary purpose of the ESL standards is to guide language curriculum across content areas for ELLs, we expected to find that language requirements would be more explicitly stated. Language expectations implicit in the standards can only be verified through review of textbooks and assessments aligned to standards and through observations of classroom instruction. Therefore, although a methodology for extracting potential language use was illustrated in this report and could be replicated, the usefulness of the data gained through these efforts is restricted unless other research is carried out concurrently that makes comparisons with actual language in use to determine the extent to which textbook and test publishers, school districts, and teachers apply the standards to their materials and in their teaching.

The analyses performed on the textbooks revealed similar textbook formats across subject areas—forecasting, presentation of information, and directed activities. Forecasting and directed activities formats were also apparent in the classroom video sample analyzed. The formats provide an organization for future analyses of not only the functional language in texts and oral discourse, but also structural and lexical features.

In terms of the functional analyses performed, the current research began with a bottom-up approach. In the first round of analyses, researchers made their initial identifications without guidelines or operational definitions of the functions. Definitions for each function were then clarified and articulated through the use of examples. This research has resulted in preliminary rating guidelines and working definitions of the functions, in addition to a means of systematically extracting the data in a way that can inform the development of relevant language tasks. On this basis, we are continuing our efforts toward operationalizing the language functions, and developing a system for training and norming raters so that reliability can be established in future efforts.

The methodological approach for identifying structural characteristics and features of texts was established and refined through the current research. The structural measures were applied to small selections of the textbooks in each content area and should be replicated on larger text samples. The language structures associated with selected functions were also described. These analyses must be expanded to cover additional, frequently occurring functions. Replicating the analyses on texts in other content areas and grade levels will result in profiles of academic language, which can be used to judge the representativeness of a text selected for test development. These analyses will also enable us to differentiate the complexity of language across grade levels.

The methodology used to identify general and specialized academic vocabulary was refined by developing rating guidelines based on the issues that arose while rating and clarifying the definitions that guide classification of general and specialized academic words. In the next phase of research, larger selections of text will be analyzed, which will lead to more comprehensive guidelines for classifying different types of vocabulary words and greater confidence in the findings. What shows the greatest promise in terms of vocabulary though is the set of measures used to characterize lexical features of a text. In the analyses conducted on the video sample, the same academic vocabulary words identified by researchers were also flagged by the three lexical measures. Our findings suggest that using the same or a similar set of measures may result in a more systematic, objective method for identifying academic vocabulary in the future.

Last, regarding the video sample analyses, the research presented here was exploratory since only a segment of one teacher-directed lesson was reviewed. While we are unable to generalize from the findings, the methods of analysis seem promising for using extant videos as a data source because we were able to systematically extract similar data sets using the same categories and analyses, when appropriate, as those used with the standards and textbooks. Thus, we will be able to compare and synthesize data from future analyses across data sources, providing a fuller description of academic language use.

> Questions 2 and 3: What are the salient features of AL in fifth-grade science and math textbooks? Which of these features are common to the two content areas of science and math?

In the current research, three areas of the analyses allowed us to identify salient features in the science and math textbooks and thus provided data for drafting sample specifications and tasks. Those areas are (1) organization of material, (2) text type, and (3) language functions. First, the textbook analyses indicated that science and math textbooks follow similar organizational formats: forecasting, presentation of information, and directed activities. These formats provide a means of categorizing information about the language that students are expected to understand in textbooks, as well as the language they are expected to produce. This information will play a role in the development of AL tests that reflect broad aspects of language use.

Second, in terms of texts, our research shows that the science and math texts differ considerably from each other. Science texts tend to be composed of expository passages with illustrations and supplementary graphics. Math texts are dominated by practice problems and contain many short imperative and interrogative sentences. Usually the only extended math texts are word problems and special sections intended to relate math concepts to the students' lives or to other subject areas. Our preliminary data show that sentences tend to be longer in science than in math.

Third, the analyses of functional language identified in the textbooks also yielded data that are useful for test development purposes. Although texts in science and math vary considerably, the language functions students must use in the two content areas and the language structures associated with those functions are similar. Language functions were analyzed on two levels in the current research, according to the language students would have to read and process in the presentation of information sections, and according to the language students would need to produce to complete a content-area task in the directed activities sections. The three most frequently identified functions listed in alphabetical order were the same in both science and math textbooks. These were *definition*, *description*, and *explanation* in the presentation of information sections and *comparison/contrast*, *description*, and *explanation* in directed activities sections.

The textbook analyses alone in this study provided data on the language functions students must process when reading; only the analyses of the textbooks can be synthesized since the standards selections reviewed provide little or no specific information about language use. For both science and math, the predominant language function in the presentation of information sections is *description*, followed by *explanation* and *definition*. Other functions appear, as indicated in Table 8 (see page 27), but less frequently.

In considering the language functions students need to complete a task, information was synthesized from both the standards and textbooks. In the standards selections and science and math textbooks, the predominant functions are the same: *description* and *explanation*; these are followed by *analysis* in the standards and *comparison/contrast* in the science and math textbooks. Table 9 shows these findings, listing functions from most to least frequent.

The results of this study indicate that *description* and *explanation* are the predominant language functions across the two data sources, regardless of whether a student is reading a text or doing a task. However, beginning with the third most frequently identified function noted above (see also Table 8 in Section 4), some language functions appeared more frequently in the presentation of information sections of the textbooks than in directed activities tasks and vice versa. For example, in the presentation of information sections, *definition* was the third most frequent function for both science and math, whereas it did not appear at all among the language functions identified in directed activities tasks. Instead, *comparison/contrast* appeared as the third most frequent function, with *analysis* in that position for the standards. Future work will help us determine if this finding is indeed a pattern and if so what the implications for test development are.

Table 9

Predominant Language Functions in Standards and Textbooks

	Standards	Science textbooks	Math textbooks
Language input	n/a	Description Explanation Definition	Description Explanation Definition
Expected language output	Description Explanation Analysis	Description Explanation Comparison/Contrast	Description Explanation Comparison/Contrast

In the standards analyses for science and math, *description* and *explanation* also emerged as the most frequently identified functions. However, the number of different functions identified in each content area varied. In science, 20 language functions were identified; in math, 13. The same pattern holds true when comparing the functions identified in the textbooks and the number of instances each function is identified, although the pattern is weaker. This finding suggests students may need to control a wider range of language functions in order to meet the standards requirements in science than in math at the fifth grade.

Although this research is preliminary and the results are not based on large samples, a pattern of specific language functions is emerging from the data across evidentiary sources—in particular, the two dominant functions, *description* and *explanation*, in the texts, with the addition of *comparison/contrast* when taking standards into consideration. Future research at other grade levels, for the same content areas and for other content areas across grade levels, will help determine what distinguishes language demands from grade to grade and across subjects. For example, if *description* and *explanation* are shown to be the most frequently occurring functions for other subjects and at different grade levels, how are these functions structurally manifested in those other contexts? We currently have preliminary data on the structures associated with functions in fifth-grade science and math textbooks, which show some similarities. However, we do not yet have enough information to make distinctions across grade levels and content areas, though being able to make such distinctions will be critical to articulating levels of proficiency and language sophistication in future test development efforts. Furthermore, the structural and lexical analyses discussed in this report need to be replicated on larger text selections so that data can be synthesized and applied for test development purposes.

Synthesis With Other Research

In this section, we synthesize the findings of the current study with findings from prior research that draw from the same types of evidentiary data sources. Few studies, however, articulate information about data sources (e.g., the specification of grade level) or empirical approaches (e.g., methodology for examination of textbooks) with enough specificity to make valid comparisons possible. Nevertheless, where results are compatible across studies, the information is helpful in improving our understanding of AL and moving us toward the operationalization of AL as a construct. Table 10 is a result of the synthesis of each data source and serves as a critical resource for specification development. A synthesis for each type of data source follows below.

Standards

In the present research, samples from subject-matter standards were analyzed according to hypothesized student output, which resulted in the identification of a range of language functions. These results and those from the analysis of a sample page from one of the ESL standards are synthesized in Section 6.1 with findings from the textbook analyses. Since no other research has been identified to date that analyzes standards according to implied language demands, no further comparisons are possible here.

Textbooks

Data in the current study on organizational formats, functions, grammatical features of functions, and structural features of the language in textbooks are discussed below in conjunction with similar data from other studies.

Organizational formats. In the current research, we identified organizational formats used across science and math textbooks—forecasting, presentation of information, and directed activities. We found no data elsewhere to parallel the forecasting data in this study. Bailey et al. (2004), however, identify an assessment function in the analysis of classroom discourse that parallels the "comprehension check" feature in this study in presentation of information. The comprehension checks serve to assess students' understanding of the text (e.g., reading skills such as whether or not students can identify the main idea or read for detail). Many questions in the directed activities tasks in the textbooks also serve a similar function. Other research does not include specific information about organizational formats. However, Coelho (1982) elaborates on the language skills students need for different classroom tasks, according to modality. She identifies readings skills needed in order to use and understand textbooks and reference books, such as reading for main idea, skimming and scanning for detail, and inferring word meanings; these skills correspond to the comprehension check feature and assessment function identified above.

Language functions. We have shown in this study that *description* and *explanation* are the two most frequently occurring functions in the presentation of information and directed activities selections reviewed for science and math. A review of three other CRESST research studies that analyze the use of language functions indicates that all but six of the 26 functions identified in the current research were identified in the other CRESST studies (Bailey et al., 2004; Butler et al., 1999; Stevens et al., 2000), across grade level clusters and content areas. Of the 26 functions, some were identified more frequently across the studies than others. The function *comparison/contrast* was the most frequently identified, appearing in all four CRESST studies; this was followed by *description* and *explanation*, which were identified in three studies; and *classification* and *inference* in two.

We identified 13 functions in a review of Coelho's (1982) informal analyses of the language needed to complete tasks in secondary school classrooms, across geography, history, and scientific English subject matter. They are: *argument*, *classification*, *comparison/contrast*, *definition*, *description*, *explanation*, *generalization/example*, *hypothesis*, *inquiry*, *justification*, *organization*, *persuasion*, and *sequence*.[10]

Short's (1993) analyses of language used in middle school history classrooms by students and teachers identify the following eight language functions: *comparison*, *definition*, *description*, *evaluation*, *explanation*, "*give example,*" *justification*, and *sequence*. She also identifies six textbook structures that reflect five of the same functions listed above, in addition to *enumeration*. All of these functions were identified in CRESST research as well, with the exception of *give example*, which was identified in Bailey et al. (2004) as a "form of support" (p. 39) and Butler et al. (1999) as the rhetorical form generalization / example. Table 10 summarizes the language functions that have been identified across the research studies discussed above.

---

[10] Per footnote 6 (p. 13), language functions are listed in noun form for the purpose of consistency across this document; however, some original sources refer to the functions in verb form (e.g., *to describe)*.

Table 10

Language Functions Identified Across CRESST and Other Research

Grade cluster	3-5	3-5	3-5	3-5	6-8	6-8	6-8	9-12
Content area	Math, Sc	Math, Sc	Sc	SS	SS	SS	Hist	Geo, Hist, Sc
Data source	Standards	Textbooks & printed materials	Oral language	Textbooks	Textbooks	Content tests	Textbooks & oral language	Textbooks
Research study	Current research	Current research; Bailey et al. (2004)	Current research; Bailey et al. (2004)	Butler et al. (1999)	Butler et al. (1999)	Stevens et al. (2000)	Short (1993)	Coelho (1982)
Language functions								
Analysis	•					•		
Argument	•							•
Classification	•	•	•	•	•	•		•
Comparison/ Contrast	•	•	•	•	•	•	•	•
Critique	•							
Definition	•	•	•		•		•	•
Description	•	•	•	•	•		•	•
Enumeration	•						•	
Evaluation	•	•	•			•	•	
Explanation	•	•	•	•	•		•	•
Generalization	•						•	•
Hypothesis	•	•						•
Identification	•	•		•	•	•		
Inference	•	•	•	•	•	•		
Inquiry	•							•
Interpretation	•	•						
Justification	•	•		•	•		•	•
Labeling	•	•	•					
Negotiation	•							
Organization	•			•	•			•
Persuasion	•							•
Prediction	•	•						
Retelling	•							
Sequence	•	•		•	•		•	•
Summary	•			•	•			
Synthesis	•							

Note: The language functions that are bolded above were the most commonly identified functions. Sc = Science; SS = Social Science; Hist = History; Geo = Geography.

Each specific grade cluster, content area, data source, and research study is shown in the rows across the top of the table. The far left column provides a complete list of all the functions identified in the research studies cited in the table. Of the 26 functions identified, eight appear frequently across these studies and data sources. These eight functions are bolded in the table. The function *comparison/contrast* appears in all the studies; *classification, description, explanation,* and *sequence* appear across all but one; and *definition, inference,* and *justification* appear across all but two.

Grammatical features of language functions. In the present research, we reviewed the short textbook selections in science and math for the purposes of describing the grammatical features of three frequently occurring language functions—*comparison/contrast*, *description*, and *explanation*. These initial analyses revealed similar structures for each function across the two content areas, including comparative adjective forms and adverbial comparisons, use of logical connectors[11] such as *instead* to show contrast, and imperative forms directing students to *compare*. For description, the simple present was used frequently with adjective forms to make the descriptions more vivid. Explanations were often given using the main verb *cause* and an infinitival clause.

The other CRESST studies used in this synthesis that analyze textbooks or printed materials (Bailey et al., 2004; Butler et al., 1999) do not include an analysis of the grammatical features of functional language as in this study. In Bailey et al. (2004), sentence structure and language functions are analyzed. Example sentences are provided from which the associated grammatical features can be extracted, such as adverbial comparatives for *comparison/contrast*, logical connectors in *description*, and the use of modals with *explanation*. In Butler et al. (1999), lengthy descriptions of texts at each grade level are provided. From these, we could extract examples of structures associated with *comparison/contrast*, such as a range of logical connectors that show conflict or contrast (e.g., however); *description* (e.g., the use of subordinate clauses); and *explanation* (e.g., logical connectors that show cause / reason, such as *so that*).

[11] We adopted the term *logical connectors* from Celce-Murcia and Larsen-Freeman (1983), who categorize connectors into several groups. In this paper, we are using the following subcategories: cause / reason (as, because), condition (if…then); conflict / causal (however, conversely, while); effect / result (so that, as a result); exemplification (such as, like); replacement (rather, instead); sequential (first, initially); and similarity (similarly, equally).

Structures that are associated with the three functions above are discussed in Coelho (1982), although not overtly in conjunction with grammatical analyses of language functions. Coelho names comparative and superlative forms of adjectives and adverbs as typical means of expressing *comparison/contrast*, in addition to logical connectors, such as *like*, *similarly*, *on the other hand*, and *the same as*. For *description*, she names sequence words that help describe a process, such as *first*, *then*, and *finally*. A variety of other language functions are mentioned in the context of helping to fulfill the function of *description*, such as the use of *generalization-example* “to describe prevailing conditions” (Coelho, p. 64). The embedding of functions within one another is also noted in the current research in Section 4 above. Last, she lists some of the rhetorical signals used to express cause and effect (*explanation*), for example, *as a result of*, *consequently,* and *causes*.

In their discussion of syntax, Dale and Cuevas (1987) give examples of frequently occurring *comparison/contrast* structures used in math, for example, *greater than/less than*, *n times as much*, and *–er than*. They do not explicitly analyze the grammatical features of *description* and *explanation*, but they do cite examples of logical connectors that are used to signal various situations that include these functions, such as *if…then* constructions, *such that*, and *because*.

Short (1993) also does not explicitly analyze grammatical features of the three functions we analyzed in this research, although she does note the frequent use of causative words in history texts, such as *as a result* and *so,* which signal *explanation*, and temporal phrases, which are used frequently in *description*.

Table 11 synthesizes information about grammatical features of functions from the CRESST research and Coelho (1982), Dale and Cuevas (1987), and Short (1993).

Information about the studies is provided in the rows across the top of the table (e.g., grade cluster, content area, and data source). The grammatical features are organized according to the three most frequently occurring functions in this study (*comparison/contrast*, *description*, and *explanation*) down the left column. Although the synthesis of this information is preliminary, we can see that certain categories, such as adverbial comparatives under *comparison/contrast* and logical connectors under *description*, were identified in as many as five of the seven data sources. A broader literature review and further analyses should help not only to confirm these initial findings, but also to expand the categories within each language function.

Table 11
Synthesis of Grammatical Features of Language Functions in Textbooks and Printed Materials

Grade cluster	3-5	3-5	3-5	6-8	Multiple	6-8	9-12
Content area	Math, Sc, SS	Sc	SS	SS	Math	Hist	Geo, Hist, Sc
Research study	Current research	Bailey et al. (2004)	Butler et al. (1999)	Butler et al. (1999)	Dale & Cuevas (1982)	Short (1993)	Coelho (1982)
Functions							
Comparison/Contrast							
Adverbial comparatives	•	•	•		•		•
Comparative adj forms	•	•			•		•
Equative comparative forms					•		
Imperative verb forms	•						
Logical connectors: a) Conflict or contrast b) Exemplification c) Replacement d) Similarity	• (c)	• (b)		• (a,b,c,d)			• (a,b,d)
Description							
Imperative verb forms	•	•					
Logical connectors: a) Effect/result b) Exemplification c) Sequential d) Similarity	• (a)	• (b)		• (a,b)	• (b,c)		• (b,d)
Modals			•	•			
Nominal structures	•				•		•
Passive voice	•				•		•
Phrasal verbs		•		•			
Predicate adj structures	•						
Prepositions		•			•		•
Simple past				•			
Simple present	•	•	•				
Subordinate clauses (e.g., relative clauses)		•		•	•		
Temporal phrases			•			•	
Explanation							
Imperative verb forms	•						
Logical connectors: a) Cause/reason b) Condition c) Effect/result			• (a,b,c)	• (a,b)	• (a,b,c)	• (a)	• (a,b,c)
Modals	•	•					
Verb *cause* w/ infinitive	•						•

Note. Sc = Science; SS = Social Science; Hist = History; Geo = Geography; Adj = Adjective.

Structural features of academic texts. In order to characterize features of typical texts, textual and linguistic features of texts (e.g., text types, syntactic features, etc.) were described and three measures were applied—average sentence length, average length of noun phrases, and percent of embedded clauses—in the current research. Here, we first discuss the types of texts found in textbooks in science and math; then we discuss the textual and linguistic features of the texts. Finally, we summarize the findings on sentence length. Findings for average length of noun phrases and percent of text composed of embedded clauses are provided in Section 4 but are not discussed further in this section because comparable analyses were not identified in the studies synthesized.

Types of texts found in textbooks. Marked differences were observed between the types of texts that appear most frequently in science and social science textbooks compared to the math textbooks. In the present study, the science texts were composed of mostly expository passages. In the social sciences, Butler et al. (1999) describe texts for Grades 5 through 8 as primarily expository, with a range of rhetorical functions occurring within. Short's (1993) discussion of middle school history texts indicates that they are expository as well. The math texts in this study, on the other hand, were primarily composed of single sentences and questions. Word problems form the most frequent paragraph structure observed in math texts. Dale and Cuevas (1987) describe math discourse as "chunks of language" that are made up of "sentences or groups of sentences or paragraphs" and suggest that word problems are an example of a subcategory of math texts (p. 337).

Textual and linguistic features of the texts. The current research identified linguistic features of the science text selections reviewed, which include concrete topics, use of prepositional phrases and participial modifiers, occasional use of passive voice, and a format that frequently defines concepts and terms. In math, the use of conditional clauses and modal verbs and the predominant use of present tense were noted.

In science, Coelho (1982) also notes the use of prepositional phrases in geography, as well as definitional structures modified by various types of clauses such as relative clauses. Bailey et al. (2004) identified complex embeddings; comparative structures; temporal-, adverbial-, and relative clauses; as well as the use of prepositional phrases and modal auxiliaries in their review of printed materials used in fourth- and fifth-grade science classrooms.

In math, Dale and Cuevas (1987) note the use of comparative structures, prepositions, and passive voice. They also note that math discourse lacks redundancy and paraphrasing, which may make it denser than other content areas and may also contribute to shorter yet not necessarily less complex sentences. Abedi et al. (1997) identified comparative structures, relative clauses, and passive voice, among other factors, as difficult linguistic features in a set of eighth-grade NAEP items. They also identified abstract presentations of information typical of expository text as a source of problems in math.

In descriptive analyses of fifth- through eighth-grade social sciences texts and printed materials, Butler et al. (1999) note a variety of linguistic features. At fifth grade, these features include concrete topics, frequent use of enumeration, the dominance of simple tenses, and use of prepositional phrases and temporal markers, which in many ways is similar to the findings for the science texts mentioned above. At sixth grade, texts are described as more abstract with the frequent use of compound noun and predicate phrases, embedded definitions, and of past tense in both active and passive voice. In addition to the features already cited in the paragraphs above, Short (1993) points out that textbooks have a typically "dense presentation" (p. 9).

The discussion of the linguistic features of texts in the preceding paragraphs provide an indication of the ranges of foci researchers have taken in describing the linguistic nature of science, math, and social science texts. We have not synthesized these data in tabular form here because the categories of analysis are not parallel. Nevertheless, we will draw from the information in these studies to refine the data collection process for our next phase of research, such that the data across content areas will be more amenable to comparison in a systematic way.

Findings on sentence length. Average sentence length provides a sense of the amount of information students must process within a specified structural unit and is part of the descriptive information that helps characterize text of a certain genre. Table 12 summarizes information drawn from this study and others about sentence length.

In the current research, the average sentence length was typically 12 words in science and 8 in math textbooks. Average sentence length for the fourth- and fifth-grade printed science materials analyzed in Bailey et al. (2004) was 13 words. In Butler et al. (1999), social sciences texts had an average sentence length of 15 words at the fifth grade, 14 words at the sixth, 19 words at the seventh, and 17 words at the eighth. We see variability within the fifth grade across content areas with the average sentence length increasing from math (8 words) to science (13 words) to

Table 12

Average Sentence Length in Student Textbooks and Materials

Research study	Grade level / Cluster	Content area	Average sentence length
Current research	5th grade	Math	8
		Science	12
Bailey et al., 2004	4th-5th grade	Science	13
Butler et al., 1999	5th grade	Social Science	15
	6th grade		14
	7th grade		19
	8th grade		17

social science (15 words). Future analyses on a larger scale will help to confirm these findings and expand our understanding of how length may help to characterize features of different content areas and grade clusters.

Classroom Discourse

In the present research, we tested the use of the methodology from the standards and textbook analyses with one language sample extracted from a classroom video. We identified the organizational formats of the lesson and the language functions used by the teacher and expected of the students. Although these results must be viewed with caution, we did note some similarities to research findings elsewhere, which suggests that future analyses could yield promising data in terms of comparability with other CRESST research and possibly other prior research. For example, similar organizational formats were identified in the textbooks and video sample in this study and in the teacher talk samples in Bailey et al. (2004), though different terminology was used for each. In this study, concept instruction is referred to as *presentation of information,* while it is called *science instruction* in Bailey et al. (2004); the organizational format identified in the textbook and video samples in this study as directed activities is referred to as *process/application instruction* in Bailey et al. (2004).

In the current study, the teacher used three repair strategies: c*larification*, *paraphrasing*, and *repetition*. The same three were also noted in Bailey et al. (2004); however, instances of repetition in that study were classified as *clarification*.

In the video sample, we identified the use of two functions by the teacher, *description* and *labeling*, and six functions expected in student responses—*classification, comparison/contrast*, *description*, *evaluation*, *explanation,* and *inference*. *Inference* was the most frequently noted in the video sample. Bailey et al. (2004) also identified the use of *comparison/contrast, description*, and *explanation*.

Summary

Drawing from the synthesis of research in section 6.2, we developed a draft matrix of academic language features found in textbooks and printed materials to be used in the development of test specifications. Table 13 is our first attempt to aggregate the data for the reading modality in a visual display. As our work continues, the matrix will be refined and expanded.

The matrix is organized across the top according to the organizational formats identified in Section 4—forecasting, presentation of information, and directed activities. The far left column displays five major categories: text types, task types, reading skills, language functions, and vocabulary. Text types and task types help to form the basis of information from textbooks from which reading text specifications and tasks will be developed. Reading skills, language functions, and vocabulary form the construct being operationalized—academic language.

Within the five main categories, subcategories were drawn from the synthesis in Section 6.2. For text and task types, we list the types identified in the textbook selections in this study. The content under the reading skills and language functions categories was identified in the synthesis. Within the subcategory of language functions, we have identified grammatical features associated with three frequently identified functions: *comparison/contrast*, *description*, and *explanation*. The features shown in the matrix appeared in at least three different research studies in the synthesis. Future research will identify the features of the other five frequently occurring functions identified in 6.2 (*classification*, *definition*, *inference, justification*, and *sequence*).

Vocabulary was included in the draft matrix because it is central to understanding a text. However, more work is needed to allow us to more clearly articulate the most critical features of vocabulary and determine how they should be

Table 13
Draft Matrix of Features of Academic Language From Textbooks and Printed Materials

	Organizational format		
	Forecasting[a]	Presentation of information	Directed activities
Text types			
Sentence		•	•
Paragraph		•	•
Multi-paragraph		•	
Visual or graphic (with or w/o text)		•	•
Task types			
Completion of Graphic Organizer			•
Generation of lists			•
Matching			•
Multiple-choice			•
Open-ended response (short, long, or numeric answer)		•	•
Reading skills			
Identify main idea		•	•
Locate supporting details		•	•
Language functions			
Comparison/contrast			
Adverbial comparatives		•	•
Comparative adj forms		•	•
Logical connectors		•	•
Description			
Logical connectors		•	•
Nominal structures		•	•
Passive voice		•	•
Prepositions		•	•
Simple present tense		•	•
Subordinate clauses		•	•
Explanation			
Logical connectors		•	•
Vocabulary			
Identify meaning in context		•	
Draw meaning from embedded definition(s)		•	

[a]No detailed analyses of academic language features in the forecasting sections were carried out in this stage of the research.

assessed. Students must be able to recognize and understand high frequency vocabulary for their grade level in order to draw meaning from texts, whether the words are classified as non-academic or academic. In addition, students must be able to draw meaning from definitions of vocabulary that are embedded in texts.

Grade level and content area have not yet been included in the matrix though both are relevant in different ways. Grade level gives an indication of the level of language complexity students need and will play a role in determining the range of proficiency levels in an assessment system. The grade-level data presented in the matrix come from studies conducted at a single grade level or multiple levels from third through eighth grades. A priority in our ongoing research agenda is specification of the trajectory of language complexity across grade levels. Content area information provides the materials associated with the language students must understand. The content of the matrix represents information drawn from three subject areas: math, science, and social sciences. Additional information, including average sentence length, the number of embedded clauses, average length of noun phrases, and other linguistic features of texts, is currently being compiled to help form text profiles that can be used for text selection and task development.

In this section, we have shown that empirical research that describes the nature of academic language can be systematically documented, analyzed, and then compared with other research. While the findings are preliminary, we are able to extract enough information to help guide the development of prototype specifications and tasks; that is, features of texts that help guide text selection and features of language use that help guide task development. We turn now to a description of the test development efforts based on the data extracted from the analyses in this study.

Test Development Efforts

In this section of the report, we illustrate how data generated from our research efforts to date can be applied to the initial steps in drafting specifications and developing prototype academic language proficiency tasks[12] for use in the English language assessment of ELLs. The long-term goal is to expand the work shown here into a comprehensive framework and test specification document that can serve as a

[12] Following Davidson and Lynch (2002), we use *tasks* to mean both individual *items*, such as a multiple-choice item, and constructed-response *tasks*, such as producing a writing sample.

foundation for producing a range of operational academic language proficiency assessments for different purposes.

The Test Development Process

The development of valid and reliable test instruments requires a highly structured process from conceptualization to implementation. Each step in the process serves a specific purpose. The process as a whole helps assure the quality of the tests being developed. Bailey and Butler (2002) provide a description of the steps to be followed in a principled test development undertaking. The process they describe provides the validity foundation for tests being produced.

> When done well, test development is a complex process that begins with determining what construct or constructs and associated skills are to be assessed. The process generally begins with a needs analysis that helps set the parameters for how the test(s) is to be used. A framework document is then developed to characterize the construct being tested. The framework draws from the research literature in the relevant fields and identifies gaps that may require additional research to help solidify the content base for next steps. The construct articulated in the framework must then be operationalized for actual test development. That is, the content in the framework document must be synthesized and translated into a working format/paradigm (facilitated by the creation of matrices) that will lead to test specifications, which in turn will guide task development. (pp. 23-24)

The long-term goal of our work is to provide a comprehensive framework and test specification document with prototype tasks that operationalizes academic English sufficiently to serve as a content base for teachers and test developers to use in developing their own language assessments. The document would serve as a point of departure for the production of test specifications and tasks tailored to specific needs (Bailey & Butler, 2002). Previous CRESST research (Bailey et al., 2004; Butler et al., 1999; Stevens et al., 2000) and the work reported in this document were synthesized (see Section 6) to provide the initial content base for creating sample draft specifications and tasks. Research is ongoing so that the content base can be expanded to include a continuum of language uses that increase in complexity and sophistication with grade level and proficiency. We turn now to the development of draft specifications and exemplar prototype tasks.

The Development of Draft ALP Test Specifications and Prototype Tasks

In this section, we present the considerations that went into the development of the draft specifications and tasks, including: (a) the selection of a language modality for this initial effort, (b) the selection of text types, (c) the selection of appropriate task formats, and (d) the issue of task content.

Language modality: reading. Our first consideration in developing sample prototype specifications and tasks was which language modality should be the focus of our initial efforts. It is our view that all modalities are equally important in the educational arena, since students must be able to function well across modalities in order to learn and achieve in U.S. classrooms. Reading was selected due to the reading demands placed on students both in the classroom and on standardized content assessments, which play a critical role in the current climate of accountability-driven educational reform.[13] Later work will expand to writing, speaking, and listening, including integrating modalities to mirror language as it is often interwoven in classroom contexts.

Text selection. An important consideration in developing any task or test is deciding what materials to use. In this case, our intent is to use authentic content-area materials for assessing language proficiency; specifically, our focus this year is on fifth-grade science and math textbooks. We reviewed five textbooks in this study and prepared a list of all the text types that appear in both the science and math. A total of seven text types were identified, for example, single words or phrases, single sentences, and multi-paragraph extended texts. Based on this review, we determined that the most prevalent text type in science textbooks is the multi-paragraph expository text; single sentences are more prevalent in math. Therefore, initial task development for the language used in science was based on multi-paragraph selections from science textbooks. However, for math, single-paragraph texts were selected rather than single sentences, because they form a richer language base from which to develop tasks. Single-paragraph texts in math are predominantly word problems, which are reflected in our text selections. Criteria for the selection of reading texts were drafted and used to select sample texts for task development. Table 14 provides those criteria.

[13] It is important to note the distinction between reading comprehension tests and second language proficiency reading subtests in English. Reading comprehension tests tend to focus more on factual information and conceptual understanding, sometimes requiring students to apply background knowledge; whereas second language reading tests tap a full spectrum of language skills (e.g., general comprehension, reading strategies, a range of vocabulary, and grammatical structures) and do not emphasize content or background knowledge.

As noted above, the initial text selections for task development were made from science and math textbooks used in mainstream English classrooms in California, the same textbooks used in this study. Specific guidelines were followed in terms of text length and type, in addition to ensuring that the language functions found to be predominant in this research were also present in the selected texts. For our purposes, topics are restricted to general or introductory topics, since more specialized topics may result in difficulties not associated with language. Finally, we avoided texts that could be problematic for reasons such as those specified in Table 14 as "Text problems to avoid." If we pilot the tasks drafted in the next sections, we will take the additional step of obtaining publisher permission to use the example texts.

Task formats. To determine which task formats would be most aligned to the formats students are exposed to in their science and math textbooks, we reviewed the long science and math selections cited in Table 5 and developed a list of all the task formats present. A total of eight different task formats were identified (e.g., completion of a graphic organizer, paragraph response, sentence completion, and short answer). We selected five formats that were also identified in social science texts (Butler et al., 1999). These task formats include: completion of graphic organizers, generation of lists, matching, multiple choice, and short answer. The mix of task formats selected reflects traditional standardized item formats (e.g., multiple

Table 14
Example Reading Text Selection Guidelines

Grade / Cluster	Elementary, Grades 3-5
Content area(s)	Science / math
Text sources	Texts should be selected from content-area texts written for English-speaking students and used in their original form.
Text length	Follow guidelines for length in the text specifications document. Multi-paragraph texts are to be selected from science textbooks and single-paragraph texts from math textbooks.
Language functions	Based on text analyses, the following language functions should be present in the selected texts: *Comparison/contrast, description,* and *explanation*. Not all functions will be present in each text.
Topics	Texts should be of a general or introductory nature and should not be highly technical, requiring teacher assistance or supplementary texts to ensure understanding.
Text problems to avoid	Efforts should be made to avoid texts with a high volume of specialized academic vocabulary; they should have sufficient content from which to generate tasks without supplemental texts or graphics; and they should not refer to other textbook activities, sections, or page numbers.

choice), as well as performance type task formats (e.g., graphic organizers). The rationale for choosing a mix of format types is that students are exposed to a range of task types in school. However, even though a range of task types occur in the materials teachers use, it is uncertain how many teachers use them and which ones they use regularly. Our previous task development work with social science texts showed that ELLs could have as much of a problem with task format as with language due to a lack of familiarity with the format (Butler et al., 1999). In our current initial test development work, we are aligning task types to tasks used frequently in textbooks and on standardized tests to help ensure the use of tasks that will not require teacher or proctor assistance to complete.

Task content. The content to be tested was drawn from the results of the current research as well as prior research (see Section 6). We selected three key areas of focus that emerged from the current research to illustrate procedures for applying findings to the development of draft specifications and task prototypes. The three areas of emphasis are: (a) reading skills, (b) language functions and associated grammatical features, and (c) vocabulary. Table 15 provides the content for the draft task specifications.

Reading skills were frequently identified in the current research on textbooks as *comprehension checks* (see Section 4) and in classroom discourse as the *assessment* function (Bailey et al., 2004). Furthermore, reading skills often form the core of many tests of reading comprehension, including identifying the main idea of a passage or the theme of a story, locating supporting details, making inferences, and using context clues. The two types selected for this initial task development effort are identifying the main idea of a passage and locating supporting details, since both are critical elements of understanding a text.

Table 15

Content for Task Specifications

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)

To read texts effectively, students must be able to interpret the textual relationships that provide meaning. Language functions play a key role in this aspect of reading, because they help students understand the purpose and content of a passage (e.g., a description of a theory vs. a comparison of two theories). In Section 6, we identified several language functions that occur frequently across content areas and language modalities. Here we selected three to illustrate the task development process: *comparison/contrast*, *description*, and *explanation*. In addition, we identified grammatical features associated with these functions, which are reflected in the draft tasks.

Last, students must be able to handle a range of vocabulary from the simplest verbs to abstract nouns; therefore, vocabulary was selected as a key area in this task development effort. In particular, students must be able to draw meaning from academic vocabulary in the context of a passage. In this effort, *specialized academic* *vocabulary* word(s) were selected for assessment if they are defined *within* the text; and *non-academic and general academic vocabulary* word(s) were selected if they are critical to understanding the main idea and supporting details of the text. Following the initial selection of vocabulary for assessment, we used the three lexical measures discussed in Section 4 to determine if the same words would be flagged as "academic" for the age range selected; all of the words selected were flagged by the measures, which gave us more confidence in our selection of words for the vocabulary task below. We move next to text selection and reading tasks.

Draft Text Selections and Reading Tasks

In the following pages we present example reading texts and tasks. The samples serve to illustrate the process exemplified in this report; that is, to show how empirical research can lead to the development of draft specifications, the selection of texts, and the development of prototype tasks. It should be noted that these texts and tasks have not yet been field tested, so they should not be considered "test ready" in their present form. Next steps would include a task tryout with native English speakers to ensure that the texts and tasks are appropriate and comprehensible, followed by revisions, as necessary, and then field testing with ELLs.

Draft specifications. The first part of the procedure in this prototype task development effort involved drafting two types of specifications: text specifications and task specifications. Texts that are selected serve as the stimulus or input material for the test tasks.

Sample text specifications were created for two reading text types: single- and multi-paragraph passages. A fully articulated specification document should include guidelines for selecting reading texts (see Table 14 above for the draft reading text selection guidelines used in this report) and specifications for each text type to be used in task development. Text specifications usually include some combination of the following elements (Butler, Weigle, Kahn, & Sato, 1996):

Typical Text Specification Components

General description	A brief definition of the text type
Length	The recommended number of words for the text type
Number of tasks per text[14]	The recommended number of tasks associated with a text in an operational form of a test
Sample text notes	Explanatory notes about the text that include a discussion of the text vis-à-vis the specified text attributes
Sample text	A text that exemplifies the text type

Task specifications usually include the following elements:

Typical Task Specification Components

General description and text type	A brief definition of the skill as operationalized in task types and the type of text that will be used
Task format	A brief description of the task type testing the skill
Stimulus attributes	Characteristics of the input to the test taker, which may include information such as a detailed description of the stem or where target-level language is found in the task
Response attributes	Characteristics of the response, which may include information such as the number of response options, characteristics of the key (correct options), characteristics of the distractors (incorrect options), or language characteristics of the response options whether multiple choice or open ended
Sample task	A task that exemplifies the task type. The key is indicated by an asterisk (*).

[14] For a given text, as many tasks as possible should be drafted and pre-tested to generate a sufficient number of acceptable tasks for an operational form of the test.

Texts selected. Using the text selection guidelines and draft text specifications, a number of texts were selected for the two text types mentioned above. In this section, we present two of the texts as examples.

The first text is a paragraph, defined as a series of related sentences (at least two, but usually no more than seven), from a fifth-grade math textbook. The sample consists of four sentences and is 42 words in length. It is a typical word problem that consists of a general topic without any specialized math words. However, it does contain a low-frequency word for this grade level, *stilts,* which will require careful attention when pilot testing items using this text. The language functions *comparison/contrast* and *description* are used. Two tasks were developed for this text and are shown below.

Sample Text #1—Single-Paragraph Drawn From Math

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? (Maletsky et al., 2002, p. 156).[15]

The second text is a multi-paragraph text, which is defined in this research as a series of at least two but usually no more than seven paragraphs with a range of 235 to 410 words. This text has five paragraphs and 240 words. Drawn from a fifth-grade science textbook, the topic is the solar system. The purpose of the piece is to compare the earth and moon; it employs the use of the language functions *comparison/contrast*, *description*, and *explanation*. Four tasks were created based on this text and are shown in the next section.

Sample Text #2—Multi-Paragraph Drawn From Science

How is the Moon Different from Earth?

Earth's moon, our nearest neighbor in space, is a far different place from earth. There is no evidence of earthquake faults as on earth's crust. There are no erupting volcanoes. In fact there is no evidence of any of the kinds of motion that earth's crust has.

Without air and water, there can be very little weathering or erosion. The moon has almost no air or water. There are no streams, no glaciers, and no wind. The only weathering and erosion is due to the impact of rocks from space hitting the moon's surface. (Daniel et al., 2001, p. 314).[16]

Draft Prototype Tasks.

A total of six prototype tasks were drafted on the basis of the two texts shown above, as well as the content specified in Table 12. The six tasks are provided and discussed below.

Sample tasks 1 and 2. The first two tasks were created for the single-paragraph math text shown above. For the first task, students are expected to demonstrate control of reading skills by locating supporting details in a single-paragraph text. To do this, students will complete a graphic organizer, in this case a table that is partially completed with information drawn from the word problem. Students must read the text, locate the details missing from the table, and then fill in the table accurately.

Person	Year	Distance	Days	From	To
Stilt walker #1	1980		158		Bowen, Kentucky
Stilt walker #2	1891	1,830 miles		Paris, France	

Sample Task #1—Reading Skills: Locate Supporting Details

Read the problem. Then complete the table.

[*Key: 3,008 miles; Los Angeles; 54 days; Moscow, Russia]

[16] From MCGRAW-HILL SCIENCE (p. 314), Grade 5. Copyright © 2001 by McGraw-Hill Science. Reprinted with permission of the publisher.

For the second task, students are required to interpret a textual relationship. In order to do this, students must be able to draw meaning from the linguistic features of a comparative structure embedded in the text. This task taps understanding of superlatives, in this case the use of the word *longest*, which are common to comparison/contrast texts and tasks. Students are expected to provide a short one- or two-word response to answer this question. The length and grammaticality of student responses are not to be assessed when scoring the answer, in order to avoid emphasis on writing skill.

Sample Task #2—Language Functions: Comparison/Contrast

Which stilt walker traveled the longest distance? _______________

[*Key: stilt walker #1; the first man; or any equivalent, correct response]

Sample tasks 3-6. The next four tasks were drafted based on the second text presented above, a multi-paragraph text. For Sample Task 3, students will demonstrate control of reading skills by identifying the main idea. A multiple-choice format was selected for this task, since it is commonly used in standardized assessments to test the same skill with native speakers. Accordingly, the prompt uses a question structure common to those assessments (e.g., "Which of the following…best…").

Sample Task #3—Reading Skills: Identify Main Idea

Which of the following is the best title for the passage? Circle the best answer.

a. How Meteorites Strike the Moon
b. How the Moon Is Different from Earth*
c. How the Moon and Earth are Similar

For Sample Task 4, students must interpret textual relationships associated with a linguistic feature of an explanatory structure (i.e., cause/effect) embedded in the text. The task format is multiple choice. Students select a sentence from the list of options that is equivalent in meaning to the prompt. This is an advanced-level task that requires students to recognize the causative structures *due to* and *caused by,* features of explanation texts, as having the same meaning.

> ***Sample Task #4—Language Functions: Explanation***
>
> 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

Task 5 calls for students to interpret a textual relationship associated with a descriptive structure embedded in the text. To complete the task, students must generate a list of features that describe the earth. The passage names features of the earth that are not features of the moon and also describes an important feature of the moon that differs from the earth, *craters*. Appearance of this word in the list would be marked as incorrect. Student answers would be scored on the completeness of their list of features drawn from the passage. Features students list that are not named in the passage will not be counted as correct, since this is not a test of background knowledge. The first answer is given as an example.

> ***Sample Task #5—Language Functions: Description***
>
> According to the passage above, what can you find on Earth that is not on the Moon? Fill in all the blanks in the list below. The first answer is given.

a.	earthquake faults	a.	earthquake faults (given)*
b.	________	b.	air
c.	________	c.	water
d.	________	d.	streams
e.	________	e.	glaciers
f.	________	f.	wind
g.	________	g.	erupting volcanoes

> [*Key: Any order acceptable]

For Task 6, students must demonstrate their understanding of key vocabulary used in the text. Students must match words from Column A to the words with the same meaning in Column B. The number of vocabulary words in this example may be adjusted according to the text and also the number of tasks needed for a particular text or test form. Words were selected on the basis of their importance in understanding key elements of the passage. The specialized academic vocabulary word *meteorite* is defined in the text as a type of rock and then used repeatedly throughout the passage. The word *impact* is used first in the same sentence as *hit*, and then hit is not used again. Students can use context to understand the meaning of the word *impact*. The other words in Column A are all used in the passage, although their equivalents are not. They are classified as non-academic and general academic vocabulary words that are used across content areas.

Sample Task #6—Vocabulary: Identify Meaning in Context, Draw Meaning From Embedded Definitions

Match the words in Column B to the vocabulary words from the passage in Column A. Put the letter of the word in Column B on the line next to the vocabulary word with the same meaning in Column A. The first one is done for you.

Column A		Column B*
meteorite	_____	a. rock
impact	_____	b. hit
shattered	_____	c. broken
produce	_____	d. make
different	_____	e. unlike
nearest	_____	f. closest

[*Key: The order of the keyed words are shown in column B directly across from each word in column A. The order of the words in Column B would be arranged randomly in an actual test.]

The two texts and the six tasks discussed above could potentially be grouped together along with other texts and tasks representative of language from other content areas for tryouts and further revisions. As we expand our test development efforts, we plan to introduce a wider variety of tasks across the content areas including performance-oriented items that integrate modalities, such as combined reading and writing activities. We hope that the use of well-aligned, integrative, and innovative assessment methods will result in a positive washback effect for curriculum and teaching.

Next Steps in the ALP Test Development Process

Test development is an iterative process that requires revisiting each step sometimes multiple times to assure the information obtained from a test is both valid and reliable. While our goal is not to produce an operational test per se, we want to be confident that the framework document, the specifications, and the prototype tasks we are developing accurately reflect the evidence of academic language use in the classroom our research describes. A notion that is important to us in this test development effort is the interrelationship of the linguistic elements in service of the language functions needed to carry out tasks. For example, it is our plan, as much as possible, to assess syntax, vocabulary, and discourse as they co-occur in the language functions rather than isolating them as discrete units. To the extent our research allows us to describe the linguistic features associated with functions, we should be able to judge mastery of specific features on the basis of use of the function in performance of the test task.

In addition to continued research that moves our analyses into other grade levels and content areas, we will begin rigorous review of the current framework, specifications, and tasks by language and subject-matter specialists. Following revisions based on input from those sources, others will be asked to generate tasks based on the specifications. In parallel, we will conduct small-scale tryouts of the tasks to be sure students are able to understand what they are being asked to do and to be sure the responses can be scored. As more research results become available, we will more fully specify the framework document and expand our test development efforts to include information from additional grade clusters and content areas.

Discussion

Our previous CRESST research and the work reported here provides a foundation for the development of English language proficiency tests for K-12 ELLs. It demonstrates how an evidence-based approach can document the type of language students must be able to understand and use to be successful in the U.S. school environment. Further, it shows at the fifth-grade level how evidence from a variety of source types—standards, textbooks, and classroom discourse—across content areas, science, mathematics, and social sciences can be synthesized to provide a coherent description of AL for English language test development purposes. This descriptive information is being used to generate draft specifications and assessment tasks that will eventually serve as models for the development of academic English language proficiency tests for ELL students.

Lessons Learned

As we went through the process of collecting and analyzing data from the multiple sources mentioned above, challenges emerged. First, the feasibility and applicability of using different types of evidence for operationalizing academic language proficiency caused us to seriously evaluate the quality and usefulness of information from each source. In working with the standards, for example, we discovered multiple problems in analyzing and generalizing across the national and state subject-matter standards and ESL standards because of conceptual and organizational differences among them. Additionally, as mentioned in Section 3 above, since the overall focus of subject-matter standards is on delivering information about content learning goals, drawing evidence of language use from these standards for the purpose of operationalizing academic English is primarily indirect. Baker, Sawaki, and Stoker (2002) point out "content standards may have been developed as broad devices for communication rather than as operational boundaries to guide the design of instruction and assessment" (p. 2). This would certainly seem to be the case in terms of AL tests.

Since language demands associated with the content learning goals are rarely made explicit in standards, we can only hypothesize about the type of language that might be employed to show mastery of content, noting that there may be a wide variety of interpretations of how to measure mastery leading to variation in how language demands are realized in the classroom. Thus, analysis of standards and the type of data yielded by the standards analyses have led us to reevaluate the role of this type of evidence in describing AL.

A second challenge faced us as we began our analysis. Differences in the nature of the evidence from current research across sources, as well as prior research, made synthesizing data for test development purposes difficult. Because the types of analyses across data sets—standards, large-scale content assessments, textbooks, and classroom discourse—vary, comparisons of findings are difficult. As we move into additional content areas and grade levels, there is a need to systematize the types of analyses performed to the extent possible so that findings are more immediately comparable. In order to produce test specifications and actual tasks, information from sources must be compatible. The methodology that has evolved from this current effort will facilitate our work as we move ahead.

Future Directions

In July 2002, an expert panel convened at CRESST to provide comment and recommendations for our research plan. The panel suggested that textbook analysis and analysis of classroom discourse would yield "more consistent and higher quality data than standards and expectations data" (Bailey & Butler, 2002, p. 21). Along these lines, it was decided that while the work on the standards reported here would be used to inform this test development effort, the primary long-range focus would be on textbook language, classroom discourse, and the language of standardized tests. Teacher expectation data will be sought to guide the selection and development of appropriate grade-level texts, discourse, topics, and tasks.

Our work this year has focused on standards and textbook analyses and the development of a methodology for collecting data on academic language use, including the identification of language functions in reading texts and tasks, the grammatical structures of those functions, the linguistic features of content texts at the fifth-grade cluster, and a methodology for the systematic and reliable identification of specialized and general academic vocabulary words. The initial application of these methodologies has been at the fifth-grade level in science and mathematics.

Figure 2 provides a visual representation of the types of evidence that currently inform, or will in the future inform, our characterization of AL. At a glance, the figure gives an overview of the coverage of CRESST work we have been able to achieve so far.

The columns for science and math at the elementary school level and the column for social studies at the middle school level provide the largest accumulations of evidence to date for informing test development. While both elementary science and math each have evidence from multiple sources, science is the most informative because of the evidence from teacher talk in fifth-grade science classrooms.

Grade clusters	Content areas											
	Science			Social sciences			Math			Language arts		
	ES	MS	HS	ES	MS	HS	ES	MS	HS	ES	MS	HS
	3-5	6-8	9-12	3-5	6-8	9-12	3-5	6-8	9-12	3-5	6-8	9-12
Sources of evidence for language demands*												
Differential test perf. of ELLs & EOs			X		X	X	X		X	X		X
Mainstream classroom observations	X				X							
Classroom videos	X											
Textbooks and printed materials	X	X		X	X		X	X				
Standardized tests			X		X		X		X	X		X
State standards	X						X					
National standards	X						X					
Teacher expectations												

*National ESL standards were also a source of evidence.

Figure 2. Types of evidence for operationalizing academic language.

We have produced preliminary specifications and science- and math-based AL tasks. In our research for the next year, we will continue to develop fifth-grade science and math tasks and will expand our efforts to include social sciences and language arts. These efforts will include piloting and revision of the preliminary test specifications and tasks, leading to further refinement of the preliminary test specification document to include whole-test construction guidelines. The initial focus will continue to be on reading, although data collection will cross all language modalities with the goal of developing specifications for the other modalities in subsequent years.

Long-term goals include comprehensive coverage of science, math, social science, and language arts, and the grade clusters for elementary, middle, and high school for test development, curriculum development, and professional development. The dissemination of this information to national, state, and district agencies who can use our work as a basis for meeting local, state, and national needs in these areas is viewed as an ongoing task critical to the equitable education and assessment of ELLs in U.S. schools.

References

Abedi, J., Courtney, M., & Leon, S. (2001). *Language accommodations for large-scale assessment in science: Assessing English language learners* (Final Deliverable to OERI, Contract No. R305B960002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., Courtney, M., & Leon, S. (2002). *Research supported accommodations for English language learners in NAEP* (Final Deliverable to OERI, Contract No. R305B960002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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

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

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

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

Aguirre-Munoz, Z. (2000). The impact of language proficiency on complex performance assessments: Examining linguistic accommodation strategies for English language learners. *Dissertation Abstracts International* (UMI No. 9973171).

The American Heritage dictionary of the English language (4th ed.). (2000). Boston: Houghton Mifflin. Retrieved February 24, 2003, from http://www.bartleby.com/61/

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. (2000). *Language analysis of standardized achievement tests: Considerations in the assessment of English language learners* (Final Deliverable to OERI, Contract No. R35B960002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bailey, A. L., & Butler, F. A. (2002). *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). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Bailey, A. L., & Butler, F. A. (in press) Ethical considerations in the assessment of the language and content knowledge of English language learners K-12. *Language Assessment Quarterly*.

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

Baker, E., Sawaki, Y., & Stoker, G. (2002, April). *The LEARNOME: A descriptive model to map and predict functional relationships among tests and educational system components*. Paper presented at the annual meeting of the American Educational Research Association (AERA), New Orleans, LA.

Butler, F. A., & Castellon-Wellington, M. (2000). *Students' concurrent performance on tests of English language proficiency and academic achievement* (Final Deliverable to OERI, Contract No. R30B60002). 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).

Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing, 18*(4), 409-427.

Butler, F. A., Stevens, R., & Castellon-Wellington, M. (1999). *Academic language proficiency task development process* (Final Deliverable to OERI/OBEMLA, Contract No. R305B960002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., Weigle, S. C., Kahn, A. B., & Sato, E. Y. (1996). *California Department of Education adult English-as-a-second-language assessment project: Test development plan with specifications for placement instruments anchored to the model standards* (Final Deliverable to California Department of Education, Contract No. 3151). Los Angeles: University of California, Center for the Study of Evaluation (CSE).

California Department of Education. (1997, December). *Mathematics content standards for California public schools: Kindergarten through grade twelve.* [PDF file] Retrieved April 17, 2003, from http://www.cde.ca.gov/standards/

California Department of Education. (1998, October). *Science content standards for California public schools: Kindergarten through grade twelve.* [PDF file] Retrieved April 17, 2003, from http://www.cde.ca.gov/standards/

Cambridge international dictionary of English. (2001). Cambridge, England: Cambridge University Press. Retrieved July 20, 2001, from http://dictionary.cambridge.org

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

Chafe, W. (1988). Linking intonation units in spoken English. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 1-27). Amsterdam: John Benjamins.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: University of Chicago Press.

Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). San Diego, CA: Academic Press.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

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

Coelho, E. (1982). Language across the curriculum. *TESL Talk, 13*(3), 56-70.

Dale, T. C., & Cuevas, G. J. (1992). Integrating mathematics and language learning. In P. A. Richard-Amato & M. A. Snow (Eds.), *The multicultural classroom: Reading for content area teachers* (pp. 330-349). White Plains, NY: Longman.

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

Davidson, F., & Lynch, B. (2002). *Testcraft: A teacher's guide to writing and using language test specifications.* New Haven, CT: Yale University Press.

Feuer, M., Towne, L., and Shavelson, R. (2002). Scientific culture and educational research. *Educational Researcher, 31* (8), 4-14.

Florida Department of Education. (1996a). *Mathematics: Grades 3-5.* [PDF file] Retrieved April 17, 2003, from http://www.firn.edu/doe/curric/prek12/frame2.htm

Florida Department of Education. (1996b). *Science: Grades 3-5.* [PDF file] Retrieved April 17, 2003, from http://www.firn.edu/doe/curric/prek12/frame2.htm

Frank, M. S., Jones, R. M., Krockover, G. H., Lang, M. P., McLeod, J. C., Valenta, C. J., et al. (2000). *Harcourt science* (California ed., Grade 5). Orlando, FL: Harcourt.

Goals 2000: Educate America Act, H.R. 1804, 103d Cong. (1994) §3(a)(9).

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

Halliday, M. A. K. (1975). Some aspects of sociolinguistics. In E. Jacobson (Ed.), *Interactions between linguistics and mathematics education* (pp. 64-73). Paris: United Nations Educational, Scientific, and Cultural Organization.

Herber, H. L. (1978). *Teaching reading in content areas.* Englewood Cliffs, NJ: Prentice-Hall.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly, 10*, 62-102.

MacWhinney, B. (2000). *The CHILDES-Project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Erlbaum.

Maletsky, E. M., Andrews, A. G., Bennett, J. M., Burton, G. M., Johnson, H. C., Luckie, L. A., et al. (2002). *Harcourt math* (Grade 5). Orlando, FL: Harcourt.

Mohan, B. A. (1986). *Language and content*. Reading, MA: Addison-Wesley.

National Council of Teachers of English & International Reading Association. (1996). *Standards for the English language arts.* Urbana, IL: National Council of Teachers of English & International Reading Association.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.* Reston, VA: NCTM.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (2002). *Scientific research in education*. Committee on Scientific Principles for Education Research. Shavelson, R. J., & Towne, L., Editors. Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

New York State Education Department. (1996). *Learning standards for mathematics, science, and technology*. Available from http://www.emsc.nysed.gov/ciai/mst.html

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

Olson, L. (2002, December). States scramble to rewrite language-proficiency exams. *Education Week.* Available from http://www.edweek.com

Ono, T., & Thompson, S. A. (1995). What can conversation tell us about syntax? In P. W. Davis (Ed.), *Descriptive and theoretical modes in the alternative linguistics* (pp. 213-271). Amsterdam: John Benjamins.

Padolsky, D. (2002, December). *How many school-aged English language learners (ELLS) are there in the U.S.?* National Clearinghouse for English Language Acquisition. Retrieved April 17, 2003, from www.ncela.gwu.edu/askncela/01leps.htm

Pimm, D. (1987). *Speaking mathematically*. London: Routledge & Keegan Paul.

The Random House college dictionary (Rev. ed.). (1988). New York: Random House.

Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 187-199). Harlow, Essex, England: Pearson Education.

Richards, J. (1978). *Classroom language: What sort?* London: George Allen & Unwin.

Richards J. C., Platt, J., & Platt, H. (1992). *Dictionary of Language Teaching & Applied Linguistics.* Essex, England: Longman Group.

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

Short, D. J. (1993). *Integrating language and culture in middle school American history classes* (Report to OERI). Washington, DC: Center for Applied Linguistics and the National Center for Research on Cultural Diversity and Second Language Learning.

Snow, M.A., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly, 23*(2), 201-218.

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

Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs* (Final Deliverable to OERI, Contract No. R305B960002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Teachers of English to Speakers of Other Languages, Inc. (1997). *ESL Standards for Pre-K-12 Students*. Alexandria, VA: TESOL.

Texas Education Agency. (1998a). Texas Administrative Code (TAC). Title 19, Part II, Chap. 112. *Texas Essential Knowledge and Skills for Science.* Subchapter A. Elementary § 112.7(b)(6)(C). Retrieved April 10, 2003, from http://www.tea.state.tx.us/rules/tac/chapter112/ch112a.html#112.7

Texas Education Agency. (1998b). Texas Administrative Code (TAC). Title 19, Part II, Chap. 111. *Texas Essential Knowledge and Skills for Mathematics.* Subchapter A. Elementary § 111.17(b)(12)(B). Retrieved April 10, 2003, from http://www.tea.state.tx.us/rules/tac/chapter111/ch111a.html

Vacca, R. T., & Vacca, J. L. (1996). *Content area reading.* New York: Harper Collins College Publishers.

Wellington, J. (1994). Language in science education. In J. Wellington (Ed.), *Secondary science* (pp. 168-188). England: Mackays of Chatham PLC.

Zakaluk, B., & Samuels, S. J. (Eds.). (1988). *Readability: Its past, present, and future.* Newark, DE: International Reading Association.

Zeno, S. M., Ivens, S. H., Millard, R. T., Duvvuri, R. (1995). *The educator's word frequency guide.* New York: Touchstone Applied Science Associates, Inc.

Appendix A
Glossary of Language Functions

The glossary consists of a list of working definitions for language functions identified in the standards, texts, or video sample. The functions are defined according to recognized sources where applicable; others have been adapted or developed based on the data being analyzed. Sources (and their abbreviations) used in developing the definitions below include: *The American Heritage Dictionary of the English Language* (AHD) (2000); *The CALLA Handbook* (CALLA) (Chamot & O'Malley, 1994); *Cambridge International Dictionary* (CID) (2001); *The Random House College Dictionary* (RHCD) (1988); and *Standards for the English Language Arts* (SELA) (National Council of Teachers of English & International Reading Association, 1996).

Language Functions

analyze: to identify the parts of a whole and their relationship to one another (adapted from SELA)

argue: to discuss a point of view with the purpose of creating agreement around a position or conviction

classify: to divide things into groups according to their type (CID)

compare/ contrast: to examine or look for differences and / or similarities between two or more things (adapted from CID)

critique: to review or analyze critically (RHCD)

define to say what the meaning of something, especially a word, is (CID)

describe: to say or write what someone or something is like (CID)

enumerate: to name things separately, one by one (CID)

evaluate: to use critical reading and thinking to judge and assign meaning or importance to a particular experience or event (adapted from SELA)

explain: to offer reasons for or a cause (AHD)

generalize: to infer a trend, an opinion, principle, or make a conclusion based on facts, statistics, or other information (adapted from RHCD)

hypothesize: to form an idea or explanation for something that is based on known facts but has not yet been proved (adapted from CID)

identify: to identify a problem, need, fact, etc.; to recognize it and show that it exists (adapted from CID)

infer: to reason from circumstance; surmise (AHD)

inquire: to seek information by forming questions (adapted from RHCD)

interpret: to decide what the intended meaning of something is (CID)

justify: to give a reason or explanation for something (adapted from CID)

label: to produce the term corresponding to a given definition

negotiate: to engage in a discussion with the point of creating mutual agreement from two or more different views

organize: to give structure to something (e.g., information or data)

persuade: to convince others of something (adapted from CALLA)

predict: to say that an event or action will happen in the future, especially as a result of knowledge or experience (CID)

retell: to relate or tell again, possibly in a different form (adapted from AHD)

sequence: to arrange or order things

summarize: to express the most important facts or ideas about something or someone in a short and clear form (CID)

synthesize: the process of identifying the relationships among two or more ideas or other textual elements (SELA)

Appendix B

Procedure for Language Function Analyses

Step 1: Select a topic found in national and state standards documents and in two or more textbooks within a grade level.

Step 2: Classify pieces of text (single sentences, single paragraphs, multi-paragraphs, etc.) as forecasting, presentation of information, or directed activity.

Step 3: For the presentation of information sections, identify the language functions that occur and quantify their relative frequency.

Step 4: For the directed activity sections, identify the language functions that the student is asked to perform and count the number of times each occurs.

Appendix C
Procedure for Descriptive Features Analyses

1. For each textbook, select one language-rich page, holding topic constant.

2. Calculate average sentence length per page.

3. Identify and calculate the average length of noun phrases (number of words in each noun phrase / number of noun phrases).

4. Identify and count embedded clauses; calculate as percent of the total number of clauses in passage.

Appendix D
Procedure for Lexical Analyses

1. For each textbook, select one language-rich page, holding topic constant.

2. Identify and count the number of general and specialized academic vocabulary words per piece of text.

3. Calculate the percentage of general and specialized academic vocabulary words.

   - Number of general academic vocabulary words / number of words per piece of text.

   - Number of specialized academic vocabulary words / number of words per piece of text.

4. Identify and calculate the percentage of words with three or more syllables (number of words with three or more syllables / number of words in sentences).

5. Identify and calculate the percentage of derived words (number of derived words / number of words in sentences).

6. Identify low frequency words:

   - Type the entire page into a CLAN file and run the word frequency program to obtain an alphabetical list of all the words on the page and the number of times each word was used (i.e., total number of different word types used, the total number of words (tokens), and the type / token ratio).

   - Find the frequency of each word in *The Educator's Word Frequency Guide* (Zeno et al., 1995).

   - Calculate the percentage of words with a frequency of less than 10 per million words.

   - Calculate the percentage of words with a frequency of less than 10 per million words for the selected grade level (in this case – fifth grade).