Standards-to-Standards Linkage Under Title III: Exploring Common Language Demands in ELD and Science Standards

CSE Technical Report 667

Alison L. Bailey
CRESST/University of California, Los Angeles

Frances A. Butler
Language Testing Consultant, Manhattan Beach, CA

Edynn Sato
WestEd

December 2005

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 Language Proficiency
Alison L. Bailey, CRESST/University of California, Los Angeles; Frances A. Butler, Language Testing Consultant, Manhattan Beach, CA; Edynn Sato, WestEd

The work reported herein was supported by the California Department of Education.

STANDARDS-TO-STANDARDS LINKAGE UNDER TITLE III: EXPLORING COMMON LANGUAGE DEMANDS IN ELD AND SCIENCE STANDARDS

Alison L. Bailey
CRESST/University of California, Los Angeles

Frances A. Butler
Language Testing Consultant, Manhattan Beach, CA

Edynn Sato
WestEd

Abstract

Under Title III of the No Child Left Behind Act ([NCLB], 2001b) every state needs to show linkage between state content standards and state English language development (ELD) standards as input to the development of state English proficiency tests. This paper argues that Title III presents a unique opportunity to explore how different content standards can be linked on a common dimension. The paper focuses on evaluating the degree to which content standards, such as English language arts (ELA) and science standards, overlap with ELD standards in terms of implicit and explicit language demands placed on students. This helps ensure that language learners are exposed to types of language that will assist them in being successful in academic contexts.

The No Child Left Behind Act ([NCLB], 2001a) is proving to be system-transforming legislation that has caused educators to place more emphasis on the academic performance of English language learner (ELL) students including the assessment of their developing proficiency of English in academic contexts (NCLB, 2001b). In this process, the relationship between English language development (ELD) standards and English proficiency tests must be direct and the ELD standards clearly linked to the standards for English language arts (ELA), mathematics, and beginning in the 2005-2006 school year, science content areas.

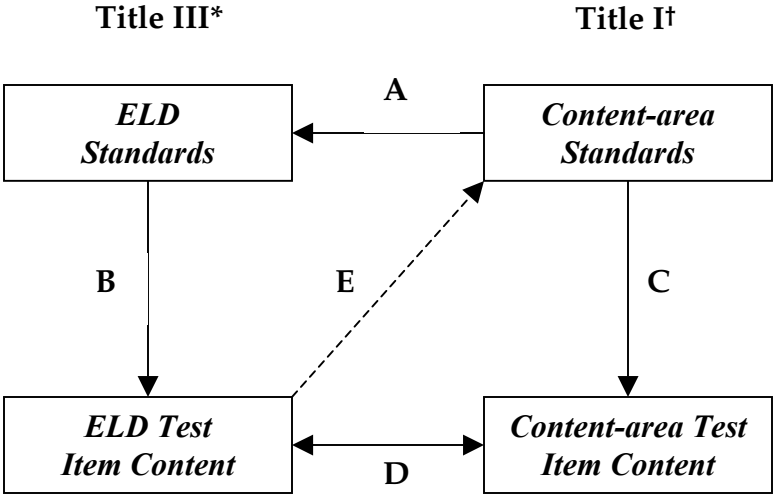
Under Title III of the NCLB (2001b) every state needs to show linkage between state content standards and state ELD standards as input to the development of state English proficiency tests. The non-regulatory guidance to states for the implementation of Title III requirements does not specify the procedure or

methodology by which this linkage should be established (U.S. Department of Education, 2003). However, this linkage undertaking is very different from the demonstration of alignment between standards and assessments. In previous studies of state standards and assessments, researchers, publishers, and states themselves have evaluated the concordance between standards statements and test items. Such studies have created methodologies to determine to what degree test items measure the skills described in content-area standards documents (e.g., Herman, Webb, & Zuniga, 2002, in press; Plake, Buckendahl, & Impara, 2001; Porter, 2002; Porter, Smithson, Blank, & Zeidner, in press; Rothman, 2004; Webb, 1997, 1999, in press; Wixson, Fisk, Dutro, & McDaniel, 2002), and even to what degree alternative sets of standards for the same content area are linked (e.g., Eckhout, Plake, Smith, & Larson, in press; WestEd, 2004).

In contrast, there is no set of procedures for establishing evidence of linkage across the standards of different content areas. Consequently, we argue that Title III presents a unique opportunity to explore how different content standards can be linked on a common dimension. In this instance, we are interested in evaluating the degree to which content standards, such as ELA and science standards, overlap with ELD standards in terms of implicit and explicit language demands placed on students. This evaluation is important because strong linkage between state content standards and ELD standards will help ensure that language learners are exposed to types of language that will help them be successful in academic contexts. In turn, strong alignment between ELD standards and ELD assessments will help ensure that states build accurate and meaningful measures of student achievement and instructional effectiveness (Fast & Hebbler, 2004; Herman, 2004).

The rationale for both evaluating standards-to-standards linkage and striving for it where it may presently be lacking is demonstrated by Figure 1. Figure 1 provides an overarching view of the relationships between standards and tests in the NCLB (2001a, 2001b) context and thereby illustrates the integration of Titles I and III. The relationship labeled A, which is the focus of this paper, requires evidence that state content standards are linked in some fashion to state ELD standards. The purpose of this linkage is for content-area standards to inform the ELD construct represented in the ELD standards. While the nature of this linkage is largely unspecified in the NCLB Act (2001b), we interpret this linkage to mean that the language implicitly or explicitly required to meet the content standards be echoed in the ELD standards. The relationships labeled B and C are representative

of the alignment between a test and the standards on which the test is based (see Bailey, Stevens, Butler, & Huang, 2005, for recent ELD test development efforts of this kind). Specifically, arrow B represents test alignment under Title III and arrow C represents test alignment under Title I (for the areas of mathematics, ELA, and science). Arrow D represents the necessary reciprocal connections between students' performance on a state ELD test and their potential performance on content-area tests. One connection addresses student readiness: ELD test performance is an indicator that a student is competent to understand the content tests administered in English; another addresses validity: students need to pass the ELD test at a certain level of proficiency for the interpretation of the content tests to be valid.



*Title III: Language instruction for limited English proficient and immigrant students.

†Title I: Improving the academic achievement of the disadvantaged.

Figure 1. Linkage and alignment schema for integration of Title I and Title III.

The broken arrow labeled E between the ELD test items and the content-area standards represents the opportunity states have to show alignment between their ELD assessment and content-area standards (e.g., by including subject matter language on the assessment or by incorporating requisite language demands in the content standards), as they respond to the mandates of the NCLB Act (2001b) and

pay increased attention to the linkage between their content standards and ELD standards and assessments.

Common Language Demands to Link ELD and Science Standards

The language demands found to be common to both ELD and content standards can provide the mechanism for linkage that states must demonstrate as part of their ELD assessment development efforts. Language demands can be discrete linguistic skills such as specific syntactic structures used to convey certain types of information (e.g., compare/contrast constructions such as the comparative *greater/less than*, Spanos, Rhodes, Dale, & Crandall, 1988). Language demands may also be at the functional or organizing level of written and spoken text, in which case academic language functions such as *explanation* and *description* found in classrooms and textbooks (e.g., Bailey, Butler, LaFramenta, & Ong, 2001/2004; Bailey, Butler, Stevens, & Lord, in press; Butler, Bailey, Stevens, Huang, & Lord, 2004) would be identified in both the content and ELD standards.

Since standards-to-standards linkage studies generally have not compared a particular aspect or dimension of learning (e.g., language demands) *across* content areas, this small-scale study explores methods for evaluating the degree of correspondence across content areas on the language dimension. The task of cross-content area comparisons on the dimension of language is made complex by the fact that content-focused standards (e.g., science standards) often lack reference to the language structures and functions needed by students to access and demonstrate proficiency in the content. Therefore, in an attempt to more fully ascertain the implicit language demands required of a student gaining proficiency in the content-area standards, a second approach was developed. It was anticipated from prior work on standards (Bailey & Butler, 2002/2003) and language demands (e.g., Bailey et al., 2001/2004; Bailey et al., in press; Butler et al., 2004) that the content standards would not yield sufficient information by which to reliably determine language demand type and complexity. These findings motivated us to include a second exploratory protocol that examines the linkage between the language demands expressed in the ELD standards and *standards-based* instructional materials in science (e.g., standards-based textbooks, and other published curriculum materials) in the anticipation that more explicit language demands would be articulated in such material. In this study we focused on a standards-based lesson plan as an illustration of this second approach.

Thus, two exploratory protocols were developed and tested as part of the work reported here: first to evaluate the state ELA and science content standards-to-ELD standards linkage using standards documents and frameworks as the source materials (i.e., standards-to-standards linkage), and second to evaluate the state science content standards-to-ELD standards linkage using standards-based science instructional materials (e.g., lesson plans), when there was insufficient information to determine the language requirements associated with a science standard (i.e., standards-to-standards-based instructional materials linkage).

In this report, we focus on the development of methodologies and relate preliminary results of the link between the California ELD standards (Grades 3-5 span) and the state science content standards (Grade 5). Information from such efforts can be used by the state to determine how well the California English Language Development Test ([CELDT], CTB/McGraw-Hill, 2004), the ELD test used in California, captures the language demands required of students to meet content-area standards. The study focuses on science because this content area is found to be language-rich according to empirical studies of language use in science classrooms and textbooks (Bailey et al., 2001/2004; Bailey et al., in press; Butler et al., 2004). We focused on the fifth grade for two key reasons. First, the work cited above has empirically documented language demands at the upper elementary grades when students are expected to have already learned to read and are using reading to learn other subject matter. This work provides the basis for operationalizing language demands used in the analyses reported here. Second, focus on a key transitional grade such as fifth grade (commonly the last grade in elementary school) is helpful because it sets the target criteria by which teachers and students can judge the end point of students' developmental progress throughout the elementary grades. The target is the use of language in fifth-grade mainstream content-area classes. Progress can be operationalized as how well students are acquiring and, from an instructional perspective, being equipped with, the necessary language demands expected of students as they graduate the elementary level (see Bailey, in press, for further discussion).

The following questions were addressed:

1. To what degree do state ELD standards reflect the language demands of state academic content standards?
2. To what degree is each protocol effective in yielding evidence that will help states meet the requirements of the NCLB? What considerations and refinements are needed, if any?

Method

This section describes the processes of instrument development, rater training, application, and evaluation used to determine linkage between the ELD and content-area standards using an initial standards-to-standards linkage protocol, and also describes the development of a second linkage protocol for ELD standards-to-standards-based instructional materials.

Procedures and Materials

ELD, and science standards statements (California Department of Education [CDE], 1999, 2000) were rated using the standards documents and the state science frameworks document (CDE, 2003). The frameworks document was included because it provided more clarity as to the intended focus and instructional application of the standards. For the standards-to-standards-based instructional materials linkage, we examined the degree to which the language contained in lesson plans could be linked to language demands reflected in the ELD standards.

Instrumentation

The protocols are comprised of both rating and evaluation instruments that were developed and refined for this pilot study. The following paragraphs describe three applications of the instruments in detail.

1. Rating of language demands in standards documents. To determine whether the standard contained sufficient information to identify the language requirements for proficiency with the skills/knowledge stated in the standard, we identified the language demands of each standard statement. Operational definitions and examples for language demands fall into two general categories: (a) Linguistic Skills and (b) Academic Language Functions. They were based on prior work by Bailey, Butler, and colleagues (e.g., Bailey et al., 2001/2004; Bailey et al., in press; Butler et al., 2004) as well as others (e.g., Schleppegrell, 2004; Short, 1993; see Tables 1 and 2 for a list of language demands and definitions).

Table 1

Language Demands: Linguistic Skills[a]

Linguistic Skill[b]	Definition/Application[c]
Phonemes	identify, manipulate, and produce the individual sounds that make up spoken words. (L, S)[d]
Syllables	identify the division of words into the smallest units of sequential speech sounds, composed of a vowel sound or a vowel-consonant combination. (L, R)
Morphemes	identify and distinguish the smallest unit of meaningful sound in language (i.e., words, roots, or affixes). (L, R)
Vocabulary words	identify and determine meaning of spoken or written words or short phrases in context; produce spoken or written words relevant to a particular context. (L, S, R, W)
Phrases & sentences	determine meaning of spoken and written phrases and sentences; generate original phrases and sentences using grammatical forms. (L, S, R, W)
Sound-symbol correspondences	identify the relationship between letters of written language (graphemes) and the individual sounds (phonemes). (R, W)
Written English conventions	recognize and apply written English conventions (i.e., punctuation, capitalization, spelling, paragraph structure, format [including text features]). (R, W)

[a] A language demand is categorized as either a *Linguistic Skill* or an *Academic Language Function* based on whether the demand is fundamental to the development and use of language or is a contextual application of language. Some language demands are grouped together because they were judged to be similar or represent multiple levels of essentially the same demand.

[b] Of what the standard requires students to show knowledge.

[c] "The ability to…"

[d] Related Language Modalities: Listening (L), Speaking (S), Reading (R), Writing (W).

Table 2

Language Demands: Academic Language Functions[a]

Academic language functions[b]	Definition/Application[c]	Notes
Identification	identify a problem, need, fact, etc., explicit in a text; recognize it, and show that it exists.	Includes information and important details, fact, and opinion.
Labeling	produce the term corresponding to a given definition.	
Enumeration	name things separately, one by one.	In labeling a picture, the picture may be interpreted as a definition.
Classifying	divide things or their attributes or properties into groups according to type.	
Sequencing	arrange, or order things.	
Organization	give structure to something (e.g., information or data).	
Definition	say what the meaning of something, especially a word, is.	
Interpretation	determine and/or demonstrate understanding of the intended meaning of something, as distinct from the literal meaning.	
Comparison/ contrast	examine or look for differences and/or similarities between two or more things.	
Explanation	offer reasons for, or a cause.	Includes supporting details. Used to code standards requiring expository writing.
Description	say or write what someone or something is like.	Used to code standards requiring narrative writing.
Inquiry	seek information by forming questions.	
Prediction	say that an event or action will happen in the future, especially as a result of knowledge or experience.	
Generalization	infer a trend, an opinion, principle, or make a conclusion based on facts, statistics, or other information.	
Inference	reason from circumstance; surmise.	
Hypothesis	form an idea or explanation for something that is based on known facts but has not yet been proved.	
Retelling	relate or tell again, possibly in a different form.	
Summary	express the most important facts or ideas about something or someone in a short and clear form.	Includes restating in own words.

Table 2 (cont.)

Academic language functions[b]	Definition/Application[c]	Notes
Analysis	identify the parts of a whole and their relationship to one another.	
Synthesis	identify the relationships among two or more ideas or other textual elements.	
Argument	discuss a point of view with the purpose of creating agreement around a position or conviction.	
Negotiation	engage in a discussion with the point of creating mutual agreement from two or more different views.	
Persuasion	convince others of something.	
Critique	review or analyze critically.	Includes understanding and knowledge of main idea, context, purpose, audience, point-of-view.
Evaluation	use critical reading and thinking to judge and assign meaning or importance to a particular experience or event.	

[a] A language demand is categorized as either a *Linguistic Skill* or an *Academic Language Function* based on whether the demand is fundamental to the development and use of language or is a contextual application of language. Some language demands are grouped together because they were judged to be similar or represent multiple levels of essentially the same demand.

[b] What the standard requires students to use.

[c] "The ability to…"

If there was sufficient information, that is, if the standard was ratable, we specified the language demand or demands present in the standard based on the wording in the standard and relevant section(s) of the framework. We did not infer meaning beyond that which was reflected in the wording of the standard and framework. For example, in Writing–Strategies and Applications, standard WSA4A, "Write a persuasive composition by using standard grammatical forms," has sufficient evidence to determine its language demands. The standard explicitly refers to the specific academic language function *persuasion* and specific linguistic skills (grammatical forms). However, Standard WSA5A, the ELD Writing Strategies and Applications standard at the Advanced level, "Independently use all steps of the writing process," does not provide sufficient information to determine the requisite language demands. The standard refers to knowledge of a process, but

gives insufficient information about the language involved in the process or the application of the process.

Next, we recorded the modality or modalities to which the standard applied (i.e., Listening, Speaking, Reading, Writing) and rated the level of complexity of the language demand reflected in the standard. We determined which of three levels of complexity was reflected by the knowledge/skills of a standard (i.e., low, medium, high). Complexity ratings were determined at the highest level of each skill required in a standard. A *low* level of complexity required recognition, identification, comprehension, or classification of information, use of simple or familiar language, or a low level of detail and depth of content to present explicit information. A *medium* level of complexity required interpretation, application, or simple analysis of information, use of increasingly complex and varied language, or a higher level of detail and greater depth of content to present primarily explicit information. A *high* level of complexity required more complex analysis, synthesis, or evaluation of information, use of predominantly complex and varied language, or a predominantly high level of detail and depth of content to present both explicit and implicit information. Finally, we noted whether there appeared to be an emphasis on one of the modalities or demands to which a standard applied and wrote comments related to the standard's grade appropriateness and fairness, as well as possible bias issues.

2. Evaluating linkage across standards documents. A crosswalk instrument was used to facilitate the systematic analysis of the relationship between the ELD standards and the content-area standards, based on the language demands and complexity reflected in the standards. The crosswalk instrument captured both the breadth and depth of linkage across the standards documents. Each standards statement was entered into the matrix cell corresponding to the language demand and level of complexity of the standard. Breadth was evaluated by comparing the range of language demands reflected in the ELD standards with the range of language demands reflected in science. Depth, or complexity, was evaluated by comparing the range of complexity levels within each language demand in the ELD and science content standard.

3. Evaluating linkage through standards-based instructional materials. The second exploratory approach to evaluating linkage did not involve alternative analyses; rather, it required application of the developed standards-to-standards linkage protocol to a different data/document source (i.e., instructional materials)

that could potentially provide more complete information about the language demands of the standards. For the purpose of this pilot study, we compared the language demands of the ELD standards with a fifth-grade science lesson plan. The lesson plan (Shannon, n.d.), selected for illustrative purposes, was taken from a link to the Los Angeles Unified School District instructional materials Web page. The analysis makes the assumption that teachers tend to use lesson plans as the model for how they would present the science content (e.g., require a class to listen to their explanations and require the class to independently read instructions to create an experiment). As with the standards-to-standards linkage, the lesson plan was independently rated, the individual ratings were discussed, and consensus established.

Rating Process

For the standards-to-standards linkage protocol, raters were selected for their expertise in one or more of the following areas: (a) academic content knowledge (i.e., science), (b) test development experience, (c) alignment expertise, and (d) knowledge of ELD. Independent rating of standards by three raters was followed by discussion comparing analysts' judgments of the language demands and complexity levels for each standard—as well as their coding rationale—in order to establish consensus in applying the protocols. Consensus regarding complexity level, for example, was reached when analysts agreed not only on the complexity level of a standard, but also on the specific language in the standard that merited the particular complexity rating. Instructional materials were independently rated by two raters using the ELD standards and fifth-grade science lesson plans and then discussed to establish consensus. When necessary, decision rules were created to further ensure consistent interpretation and application of the criteria to standards. Training ended when we achieved consensus on a set of ratings that were accurate and consistent with the requirements of this study's protocol and criteria.

Results

The results are reported in four sections: (a) the degree to which ELD standards are ratable and the language demands that are represented in them, (b) the degree to which science standards are ratable and the language demands that are represented in them, (c) the degree of linkage between the two sets of standards in terms of breadth and depth of language demands, and (d) parallel information using a standards-based science lesson plan for illustration.

Evaluating the Language Demands of ELD Standards (Grades 3–5 Span)

A linkage protocol was used to take account of the language demands expressed in the California ELD standards for the interaction between language and content reflected in content-area standards as well as the variability with which language constructs are addressed in these standards. The ELD standards are distributed across five proficiency levels: (a) Beginning, (b) Early Intermediate, (c) Intermediate, (d) Early Advanced, and (e) Advanced. They are also organized to the four modalities, with Listening and Speaking forming one unitary strand, and Reading and Writing forming two additional standards.

Ratability of the ELD standards. The first step of standards analysis was to determine whether analysts could identify language demands in a given standard. Within the Grades 3–5 ELD standards, 100% of the Listening and Speaking standards (28) and Reading standards (91) had sufficient information to be "ratable." All but one Writing standard, WSA5A cited above, (44 out of 45 or 98%) were judged to be ratable. For each ratable standard, relevant language demands (linguistic skills and academic language functions), language modality (listening, speaking, reading, and writing) and complexity level (low, medium, and high) were identified.

Language demands of the ELD standards. Table 3 shows the distribution of the language demands for the Grades 3–5 ELD Standards across modalities. The language demands include both linguistic skills and academic language functions.

For Listening and Speaking, the 28 standards contain 42 occurrences of language demands; of those, 20 (47.6%) are linguistic skills and 22 (52.4%) are academic language functions. Of the 20 linguistic skills, 19 are *phrases and sentences* with the remaining skill in the *phonemes* category. The 22 academic language functions cover six language demand categories. The most commonly occurring language demand is *phrases and sentences* (67.9%). Fourteen standards (50%) require one language demand. The other 14 (50%) require multiple language demands. We were able to code 2 of the 7 linguistic skills and 6 of the 14 academic language functions to the ratable Grades 3-5 ELD Listening and Speaking standards.

Table 3

Distribution of Language Demands Across Modalities in the Grades 3–5 ELD Standards and Ratable Grade 5 Science Standards by Number (and Percentage)

Language demand	Listening & Speaking	Reading	Writing	Science[a]
Linguistic skill				
Phonemes	1 (3.6%)	3 (3.3%)		
Morphemes		9 (9.9%)		
Syllables		1 (1.1%)		
Sound-symbol correspondences		1 (1.1%)	1 (2.3%)	
Vocabulary words		27 (29.7%)	4 (9.1%)	1 (7.7%)
Phrases and sentences	19 (67.9%)	19 (20.9%)	18 (40.9%)	3 (23.1%)
Written English conventions		5 (5.5%)	17 (38.6%)	2 (15.4%)
Academic language function				
Analysis and synthesis				
Argumentation, negotiation, and persuasion			2 (4.5%)	
Classifying, sequencing, and organization		4 (4.4%)	3 (6.8%)	6 (46.2%)
Comparison and contrast		5 (5.5%)		
Critique and evaluation	3 (10.7%)	11 (12.1%)	1 (2.3%)	1 (7.7%)
Definition		1 (1.1%)		
Description		2 (2.2%)	9 (20.5%)	1 (7.7%)
Explanation	2 (7.1%)	2 (2.2%)	3 (6.8%)	1 (7.7%)
Identification	4 (14.3%)	22 (24.2%)		1 (7.7%)
Inquiry	5 (17.9%)	1 (1.1%)		1 (7.7%)
Interpretation	1 (3.6%)	7 (7.7%)		
Labeling and enumeration			1 (2.3%)	3 (23.1%)
Prediction, generalization, inference, and hypothesis		4 (4.4%)		3 (23.1%)
Retelling and summarizing	7 (25.0%)	2 (2.2%)	1 (2.3%)	

Note. These demands are not mutually exclusive, hence the total percentage of standards in columns 2 through 4 can sum to over 100.

[a]Thirteen of 38 Grade 5 science standards were judged "ratable" according to this study's criteria.

For Reading, the 91 standards contain 126 occurrences of language demands; of those, 65 (51.6%) are linguistic skills and 61 (48.4%) are academic language functions. Of the 65 linguistic skills, more than two thirds belong to the *vocabulary words* and *phrases and sentences* categories. The 61 academic language functions cover 11 language demand categories, with over half belonging to the *identification* and *critique and evaluation* categories. The most commonly occurring language demand is *vocabulary words*, which was identified in 27 (29.7%) of the 91 standards. Fifty-seven of the 91 standards (62.6%) require one language demand. The other 34 standards (37.4%) require multiple language demands. We were able to code all 7 of the linguistic skills and 10 of the 14 academic language functions to the ratable Grades 3–5 ELD Reading standards.

For Writing, the 44 ratable standards contain 60 occurrences of language demands; of those, 40 (66.7%) were linguistic skills and 20 (33.3%) were academic language functions. Of the 40 linguistic skills, more than 35 belong to either the *phrases and sentences* or *written English conventions* categories. The 20 academic language functions cover 7 language demand categories, with nearly one half belonging to the *description* category. The most commonly occurring language demand categories were *phrases and sentences,* reflected in 18 (40.9%) of the 44 standards, and *written English conventions,* reflected in 17 (38.6%) standards. Thirty standards (68.2% of ratable standards) required one language demand. The other 14 (31.8% of ratable standards) required multiple language demands. We were able to code 4 of the 7 linguistic skills and 7 of the 14 academic language functions to the ratable Grades 3–5 ELD Writing standards.

Worth noting in this analysis are the patterns that differentiate the modalities in terms of both linguistic skills and academic language functions. In only a few instances do the demands cut across all modalities. It will be important in future work to determine if these patterns hold up across grade levels.

Language modalities of the ELD standards. We examined the distribution of Grades 3–5 ELD standards across the four language modalities: (a) listening, (b) speaking, (c) reading, and (d) writing. All four modalities are represented in these standards. While the standards were written to a specific modality, the listening and speaking and the reading standards included cross-modality skills. For example, of the ratable standards for listening and speaking, 13 (46.4%) were coded to one modality and 15 (53.6%) to multiple modalities. The multiple modalities distribution for listening and speaking included 15 (53.6%) for listening, 27 (96.4%)

for speaking, with reading and writing each represented in one (3.6%) standard. For the writing standards, only the writing modality is represented. All 44 ratable writing standards were coded to that modality.

Complexity levels of the ELD standards. The Grades 3–5 ELD standards contained language demands at each of the three complexity level ratings—low, medium, and high—with the majority across modalities at the low and medium levels. Table 4 shows the distribution of standards across the complexity levels.

For listening and speaking, 6 standards contained language demands at multiple levels. All of the language demands rated as high complexity were in the *phrases and sentences* language demand category. For reading, 11 standards contained language demands at multiple levels. Six standards did not have enough information for analysts to determine a complexity level for a language demand; these standards were rated indeterminate. Five of the 6 indeterminate standards were associated with the *vocabulary words* language demand. In these instances, the standard did not specify the vocabulary, and analysts believed that the difficultly level of the vocabulary word would influence the complexity level. For writing, 7 standards contained language demands at multiple levels.

Evaluating the Language Demands of Science Content Standards (Grade 5)

A linkage protocol used to evaluate the science standards accounts for the interaction between language and content reflected in content-area standards, as well as the variability with which language constructs are addressed in these standards. The California science standards are organized according to the following six scientific areas: (a) Physical Sciences, (b) Life Sciences, (c) Earth

Table 4

Distribution of Grades 3–5 ELD Standards and Ratable Grade 5 Science Standards across Complexity Levels by Number and (Percentage)

Complexity level	Low	Medium	High	Multiple levels
Listening and speaking	13 (46.4%)	17 (60.7%)	4 (14.3%)	6 (21.4%)
Reading	41 (45.1%)	51 (56.0%)	3 (3.3%)	11 (12.1%)
Writing	17 (39.5%)	29 (67.4%)	5 (11.6%)	7 (16.3%)
Science	2 (15.4%)	8 (61.6%)	4 (30.8%)	1 (7.7%)

Note. The combined percentage is greater than 100% because of multiple complexity levels on one standard.

Sciences (Earth's Water), (d) Earth Sciences (Weather), (e) Earth Sciences (The Solar System), and (f) Investigation and Experimentation.

Ratability of the science standards. Of the 38 science standards for fifth grade, 13 (34%) had sufficient information about language demands to be judged ratable. The science standards judged to have indeterminate language demands contained explicit wording regarding content knowledge, but did not contain the language needed by students to learn or demonstrate their knowledge. Much of the language in these standards referred to "knowing" a concept, process, or relationship, but not to the specific scientific language used in the standards and framework descriptions. Moreover, "knowing" does not clearly convey the level of complexity required; that is, it is unclear whether "knowing" requires a lower-complexity skill such as recognition or identification, or a higher-complexity skill such as analysis or synthesis. Thus, for such standards, we did not assume the standard required knowledge or use of scientific terminology, nor did we assume a particular level of complexity (i.e., low, medium, high), unless it was explicitly stated in the standard or framework. As a result, for 25 of the 38 science standards, language demands could not be identified.

Language demands of the science standards. Table 3 also shows the distribution of language demands in the 13 ratable science standards. Only 11 (52.4%) of 21 possible types of language demand could be coded in the standards. In any given standard multiple language demands could occur; 4 standards (30.8% of ratable standards) required one language demand, and 9 standards (69.2% of ratable standards) required multiple language demands. Thus, there were a total of 23 occurrences of language demands coded, 6 (26.1%) of which were linguistic skills and 17 (73.9%) of which were academic language functions.

Of the six instances of linguistic skills coded, half (three) were *phrases and sentences*, two were *written English conventions*, and one was *vocabulary words*. The ratable science standards were not judged to reflect the other four linguistic skills; however, given the content-focus of these standards, we would not expect them to emphasize linguistic skills such as *phonemes, morphemes, syllables*, and *sound-symbol correspondences*.

Most of the language demands coded to the ratable science standards reflected academic language functions. The most commonly occurring language demand belonged to the *classifying, sequencing, and organization* category (i.e., 6 of 17 instances of academic language functions coded), which was reflected in 6 (46.2%) of the 13

ratable standards. Three instances of academic language functions were coded to *labeling and enumeration* as well as to *prediction, generalization, inference,* and *hypothesis*. One instance was coded to each of the following academic language functions: (a) *critique and evaluation*, (b) *description*, (c) *explanation*, (d) *identification*, and (e) *inquiry*. The ratable science standards were not judged to reflect the remaining 6 academic language functions.

Language modalities of the science standards. We examined the distribution of ratable Grade 5 science standards across the four language modalities: (a) Listening, (b) Speaking, (c) Reading, and (d) Writing. Twelve (92.3%) of the 13 ratable standards were coded to all four language modalities and reflected academic language functions. In each of the 12 cases, modality was not explicit in the standard.[1] The standard coded to only one language modality was "Students know water vapor in the air moves from one place to another and can form fog or clouds, which are tiny droplets of water or ice, and can fall to Earth as rain, hail, sleet, or snow." With consideration of its accompanying framework description, "By learning basic meteorology from texts…" this standard was coded to the reading modality because of the explicit mention of learning from texts.

Complexity levels of the science standards. Table 4 also shows that the ratable Grade 5 science standards contained language demands at each of the three complexity levels. One standard contained language demands at multiple levels of complexity. The majority of language demands reflected in these ratable standards were coded to the medium level of complexity; that is, in 8 (61.6%) of 15 instances, the language demands required students to use, for example, language in interpretation, application, or simple analysis, or use language with complex construction. In 4 (30.8%) instances, the complexity level was coded as high. In 2 (15.4%) instances, the complexity level was coded as low.

Evaluating Linkage Between the ELD and Science Standards

We based judgments of linkage between the ELD standards for Grades 3-5 and Grade 5 science standards on analyses of the language demands and complexity reflected in the standards (i.e., based on consensus ratings of the standards). The breadth of the relationship between standards was evaluated by comparing the range of language demands reflected in the ELD standards with the range of language demands reflected in the set of standards for the science content area.

[1] Since the modalities are often closely interrelated and difficult to distinguish, standards were coded to all four modalities when one was not explicitly referenced.

Depth, or complexity, was evaluated by comparing the range of complexity levels within each language demand in the ELD and the science content area standards.

Breadth of linkage. Although only 34% of the science standards were judged ratable, all of the language demands (100%) reflected in these ratable standards are included in the ELD standards for Grades 3-5. Moreover, this coverage includes language demands associated with both linguistic skills and academic language functions. Thus, for linkage measured in terms of breadth of coverage, all of the language demands in the ratable science standards were also present in the ELD standards.

Depth of linkage. One hundred percent of the language demands reflected in the ELD standards are at or below the complexity level of the demands in the ratable science standards. The ELD standards for Grades 3-5 cover 9 (82%) of the 11 language demands at the same complexity level as that reflected in the ratable science standards. One language demand was coded as medium complexity for science and low complexity for ELD (i.e., *labeling and enumeration*). Another language demand was coded as high complexity for science, and for ELD was coded as low, medium, and of indeterminate complexity (i.e., *inquiry*). Thus, with the two exceptions indicated, the depth of coverage or complexity for the subset of ratable science standards is matched to the ELD standards.

Evaluating Linkage Through Standards-Based Instructional Materials

Preliminary findings from the standards-to-standards linkage analyses suggested that, for the majority (66%) of Grade 5 science standards, there is insufficient information to determine the language demands needed by students to access or demonstrate proficiency in the content area. As discussed previously, although there appears to be a high degree of linkage between the science and ELD standards, this linkage is based on a very narrow portion of the science standards. For this reason, the second approach is explored to help determine linkage on the language variable between the two sets of standards.

Case study of application to a science standard (Grade 5). For this study, the selected lesson is a proxy for Grade 5 Life Science Standard 2b (CDE, 2000). This standard, for which the lesson plan was selected, was identified as "unratable" in the initial analyses. The lesson plan may be relevant to other science standards as well, but the protocol requires identifying each standard for language demands analysis on an individual basis.

Life Science Standard 2b states:

> 2. Plants and animals have structures for respiration, digestion, waste disposal, and transport of materials. As a basis for understanding this concept:
>
> > b. Students know how blood circulates through the heart chambers, lungs, and body and how carbon dioxide (CO2) and oxygen (O2) are exchanged in the lungs and tissues.

The standard serves as the objective for the lesson plan. The chosen lesson plan shows a link to the standards and provides richer language information than the standards alone. It first provides a two-paragraph *Lesson Background* that explains the purpose of the ribs in the body, followed by a description of the diaphragm and its role in relation to the ribs and the lungs. Next, the lesson title to be presented to the students, *The Breathing Machine*, is indicated along with a list of materials needed for students to make a model of their lungs to help them demonstrate how the lungs and diaphragm work. Finally, the model-building steps are provided with some explanation throughout of what the steps illustrate. The final goal is a class discussion during which the students are to conclude "that breathing is a mechanical process by which there is an interaction between the organism and the surrounding air; the lungs and other parts of the respiratory system perform this mechanical process." The information provided to the teacher in the lesson plan is rich in academic language.

One aspect of this instructional materials approach that uses the lesson plan, as has been done here, allows analysts to more directly capture whether the language demands are required of the students (i.e., productive language) or made on students by the teacher (i.e., receptive language).

Ratability, language demands and complexity levels of the science lesson plan. There was sufficient detail for analysts to determine potential language demands at primarily the academic language function and vocabulary levels. We coded 50 instances of seven identifiable types of language demands in the life sciences lesson plan. There were 41 instances of the language demands in the lesson plan at the *vocabulary words* level, including both general academic language words such as *demonstrate, attach, discuss, perform*, and *composed of*, which can all cut across content areas, and specialized content-specific words such as *diaphragm, organism, respiratory system, carbon dioxide*, and *exhale*.

There are just two complex phrases we identified as potentially demanding due to the nature of the concepts or language functions these structures support. For example, the compare/contrast function required use of the contrastive temporal phrase *the same as when*, and the second phrase turns a verbal phrase into part of a complex nominalization used as the head noun phrase (*Changing air pressure outside the lungs is*…).

Of the five types academic language functions identified, *explanation* and *description* were required of both the teacher and the students (i.e., the students would be required to both listen to and speak (the two language modalities that we identified in this lesson plan); that is, comprehend and produce these two types of language functions). *Comparison and contrast*; *classifying, sequencing, and organization*; and *definition*, on the other hand, occurred just once each. They required the student to comprehend their use by the teacher but did not require the student to produce them.

The lesson plan contained language demands at low to high complexity levels. The complexity levels are presented here by the type of demand coded. Note that the complexity level implies varying degrees of complexity *within* a grade level (obviously language demands rated "low" complexity for the fifth grade may rate "high" for lower grades).

Breadth and depth of linkage with science lesson plans. We report linkage between the ELD standards and the language of the science lesson plan separately for listening and speaking. For listening, there was no match in terms of *vocabulary words*. In terms of *phrases and sentences* there was a match (both the ELD standards and the science lesson plan contained medium- and high-complexity language demands). The *explanation*, and the *classifying, sequencing, and organization* language functions found in the science lesson plan were each matched at a comparable level of complexity on the ELD standards. *Comparison and contrast* was found in both the science lesson plan and the ELD standards, but the ELD standards rated as requiring only a low level of complexity and the lesson plan rated as requiring a high level of complexity. *Description* and *definition* found in the lesson plan had no corresponding language function in the ELD standards.

For speaking, *vocabulary words* identified in the science lesson plan matched with the identified language demands in the ELD standards at the medium level of complexity only. There were no standards identified as requiring high-complexity vocabulary, which the lesson plan appeared to require. In terms of *phrases and*

sentences there was a match (both appear to contain medium- and high-complexity language demands in speaking). The *explanation* and *description* academic language functions, which students appear to be required to produce at a medium level of complexity in the science lesson plan, correspond with the same language functions at the same level of complexity in the ELD standards.

Discussion

First summarizing our findings in terms of the research questions, we asked to what degree are state ELD standards linked to state academic content standards? The answer to this question is important in order to ensure that English learners attain the level or range of English language proficiency needed to facilitate progress in the content areas. Across the subset of Grade 5 science standards we determined to be ratable (34%), there appears to a match in both the breadth and depth of language demands reflected in these standards and the Grades 3–5 ELD standards. That is, all of the language demands reflected in the ratable science standards are also reflected in the ELD standards, and the ELD standards are at the same or a lower level of complexity as the ratable science standards. However, a major caveat to these findings is that while 34% of the Grade 5 science standards were ratable, 66% were judged as having insufficient information to determine their associated language demands. Moreover, the modality or modalities in which knowledge of these standards can be displayed is almost never made explicit.

At present, language demands related to science standards are more likely reflected in the planned instruction of these standards rather than in the standards themselves, in the process of teachers helping students to access and demonstrate proficiency in the knowledge and skills described in the science standards. Thus, an approach for uncovering these instructionally embedded language demands was considered in an attempt to provide additional information about linkage between the ELD standards and science standards. Such an undertaking can be considered a study of the validity of the ELD standards in terms of their utility to teachers preparing ELL students for the language of the content areas. Or, more simply put, the second linkage approach we developed helps to answer the question, do ELD standards get represented in planned instruction for science (Personal communication, Barbara Merino, June 2005)?

While we examined lesson plans for this pilot study, other types of documents may also be considered by states to provide data that yield richer descriptions of

language demands related to content-area standards. These could include such materials as teacher guides to interpretation of standards, standards-based textbooks, and activity guides (e.g., California Science Teachers Association, 1999). The selection of standards-based instructional materials, if such an approach is adopted, should be carefully controlled to assure a pre-existing relationship with the appropriate state standards. For example, a procedure for the selection of lesson plans could be established. Two areas for consideration are quality control (use of state-funded/sponsored and/or state-endorsed standards-based lesson plans—often created in conjunction with content-area teacher professional organizations) and state agreement (the state must first confirm that the chosen lesson plans are meaningful in terms of quality, recency, utility, and widespread adoption).

In the case study examination of a fifth-grade science lesson plan, we found that the best match between language represented in the ELD standards and the language represented in the standards-based lesson plan was in the type and complexity of syntactic structures, followed by the variety in academic language functions found in both documents. Vocabulary demands did not match well across the two; in particular, the level of complexity suggested by the ELD standards was primarily lower than the level of complexity of vocabulary found in the lesson plan.

In terms of the second key research question regarding the performance of the methodology, we asked, to what degree is each protocol effective in yielding evidence that will help states meet the requirements of the NCLB legislation (2001b)? A related question we asked was, what considerations and refinements are needed, if any? We found that the protocols allowed for uniformity of application across the ratings. Their initial refinement underscored the need to clearly define academic language functions as distinct from general cognitive functions where possible, as well as the need to group multiple discrete demands (e.g., phoneme deletion, substitution) under larger units of analysis (e.g., phonemes).

However, the protocols in this work took into account only those content-area standards for which language demands can be determined. Thus, a high percent of linkage could suggest a greater amount of correspondence than actually exists in absolute terms (e.g., only 34% of the science standards in this study had identifiable language demands, but 100% of those standards had linkage to the ELD standards). Therefore the formula for calculating degree of linkage should consider the proportion of ratable standards. Furthermore, while academic language functions were readily identifiable in the documents we analyzed, linguistic skills were not.

Use of language captured by functions was often salient in the verbs used to describe the *type* of student knowledge desired (e.g., *explain* X, *contrast* Y with Z, etc.). Therefore, analysis of the lesson plans yielded more specificity than the science standards documents and frameworks yielded. This level of detail from the lesson plans can only be suggestive of what exposure students receive in the classroom, because as already mentioned, not all teachers will necessarily use the range of vocabulary or complex syntactic structures identified in the lesson plan. In addition, lesson plans may not always reflect the different vocabulary and grammatical structures that are now found to be characteristic of the different content areas (e.g., Butler et al., 2004; Schleppegrell, 2004) because they were not originally created with such factors in mind.

It is important to reiterate that the linkage tasks undertaken in this study are fundamentally different from prior linkage and alignment studies with state academic standards. Part of the process of developing protocols for evaluating linkage between ELD and content standards has led to our identifying the issues unique to this is type of linkage and has enabled us to better clarify the Title III requirement described at the beginning of this paper. In this discussion, we divide the issues into three key areas for both conceptual and methodological consideration: (a) operationalization of the language demand construct, (b) measurement of language along an acquisitional continuum, and (c) the interaction between language and academic content.

Language Demands

The language demand construct is the commonality across the content areas of ELD and science on which we evaluated linkage. We operationalized language demand by identifying two different types—linguistic skills and academic language functions. Language demands that were categorized as linguistic skills were often problematic to rate because too little is explicitly stated to determine grammatical structures, word usage, etc. Moreover, we suggest that it is less meaningful to expect direct linkage between ELD standards that represent language features (e.g., phonemic awareness) and standards for content such as science that need never make explicit the component skills of the language modalities. States with the help of linguists and educators should decide which linguistic skills are key to content standards achievement and thus should be included in those standards, and conversely, how much language of the content areas the ELD standards should reflect.

With the findings based on analysis of academic language functions, we are in a position to ask how similar the results of this pilot are to our expectations. The results of the pilot study suggest that in the Reading modality at least, the California ELD standards that were produced in 1999 are covering the range of academic language functions we found represented in the state's ratable science content standards. However, these results are incomplete in regard to the remaining grades in science and to the other content areas under the NCLB Act (2001b), specifically mathematics and ELA.

Measurement of Language Along an Acquisitional Continuum

Language development can generally be thought to occur along a continuum, whereby less challenging skills are incrementally built upon to establish more sophisticated comprehension and production. The ELD standards in California and many other states are organized into within-grade span proficiency levels that presumably attempt to capture the developmental progression of English-as-a-second language. The five levels represented in California's ELD standards range from *Beginner* to *Advanced*. This continuum poses challenges for evaluating linkage between the ELD standards and the academic content area standards that are not organized on a developmental continuum. Unlike the ELD standards, standards for science and other content areas are made in statements of discrete topic knowledge and age-appropriate skills. The content standards do not typically categorize content knowledge into proficiency levels within a given grade. We were therefore faced with a two-step process to overcome this organizational difference across the two sets of standards documents. We first clustered similar skills across different levels of ELD proficiency into one standard statement and then created our own language demand complexity rating scheme in order to form a common metric by which to judge degree of linguistic and discourse sophistication represented in the two sets of standards.

Interaction of Language and Content

There are also difficulties with operationalization of the language demand construct due to the interwoven nature of language and content-area material. For example, at what point does knowing a word like *diaphragm* stop being an issue of ELD proficiency and become an issue of science knowledge requiring explanation and placement within the larger conceptual frame of human anatomy or the respirational system? Such interaction between content-area material and language

has made the operationalization of the academic language construct extremely challenging for linguists and educators (Bailey & Butler, 2002/2003).

One way to address validity at the lexical level is to categorize vocabulary by context of learning and use. Following Scarcella and Zimmerman (1998), in previous research Butler et al., (2004) reliably distinguished between general academic vocabulary (e.g., *evidence, demonstrate,* and *represent*) and specialized lexicons (e.g., *diameter, condenses,* and *abolitionist*), each of which students must acquire in order to become fully proficient in English in the academic setting.

Implications

While the research reported here was exploratory, the results have implications for states interpreting and attempting to comply with the educational policy laid out in the NCLB legislation (2001b). The results imply two areas in which state standards documents would benefit from greater specificity:

1. The degree of complexity of the lexical and grammatical forms expected of students at each ELD level.
2. Explicit statement of the language demands required and the modalities preferred for demonstrating mastery of the content standards.

Attention to point 2 above will benefit all students by broadening our linguistic understanding of the content areas.

We conclude with two suggestions. First, we recommend the federal government move to coordinate the efforts of states faced with devising processes to evaluate the nature and quality of linkage between ELD standards and content standards. Such efforts should include guidance on what constitutes evidence of linkage, and what will be considered adequate linkage between standards and ultimately between standards and assessments, as the schema in Figure 1 portrays. Evaluating and anticipating the issues common to all states attempting to determine standards linkage under Title III will facilitate state efforts to more effectively demonstrate standards linkage and alignment.

Second, we suggest that states consider the exploratory processes detailed in this study as a point of departure to begin the work of systematically evaluating and, as necessary, refining their standards. The characterization of academic language is still evolving; therefore, states should consider re-evaluating their standards as new knowledge about the construct emerges. Systematically keeping track of whether instruction and student performance in the content areas change as

a result of forging closer links between ELD and content-area standards should be part of the effort. Such processes are intended to help states ensure that their standards appropriately reflect intended constructs and are adequately complete and comprehensive vis-à-vis the requirements of the NCLB Act (2001b).

References

Bailey, A. L. (in press). Introduction: Teaching and assessing students learning English in school. In A. L. Bailey (Ed.), *Language demands of school: Putting academic language to the test.* New Haven, CT: Yale University Press.

Bailey, A. L., & Butler, F. A. (2002/2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document* (Final Deliverable to OERI/OBEMLA Contract No. R305B960002) (Currently available as CSE Rep. No. 611). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Reprinted with modification as: Bailey, A. L., & Butler, F. A. (in press). A conceptual framework of academic English language for broad application to education. In A. L. Bailey (Ed.), *Language Demands of School: Putting academic language to the test.* New Haven CT: Yale University Press.

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

Bailey A. L., Butler, F. A., Stevens, R., & Lord, C. (in press). Further specifying the language demands of School. In A.L. Bailey (Ed.), *Language demands of school: Putting academic language to the test.* New Haven, CT: Yale University Press.

Bailey, A. L. Stevens, R., Butler, F. A., & Huang, B. (2005). *Using standards and empirical evidence to develop academic English proficiency test items in reading.* (Final Deliverable, March 2005, to Institute of Education Sciences Contract No. R305B960002. To be available as CSE Report. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Butler, F. A., Bailey A. L., Stevens, R., Huang, B., & Lord, C. (2004). *Academic English in fifth-grade mathematics, science, and social studies textbooks.* (Final Deliverable to IES, Contract No. R305B960002; currently available as CSE Rep. No. 642). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

California Department of Education. (1999). *English-language development standards for California public schools: Kindergarten through grade twelve.* Sacramento, CA: California Department of Education. Retrieved April 21, 2004, from http://www.cde.ca.gov/re/pn/fd/documents/englangdev-stnd.pdf

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

California Department of Education. (2003). *Science framework for California public schools: Kindergarten through grade twelve.* Sacramento: California Department of Education. Retrieved April 13, 2005, from http://www.cde.ca.gov/re/pn/fd/sci-frame-dwnld.asp

California Science Teachers Association. (1999). *Making connections: A guide to implementing science standards.* Sacramento, CA: Author.

CTB/McGraw-Hill. (2004). *California English Language Development Test. Form D.* Monterey, CA: Author.

Eckhout, T. J., Plake, B. S., Smith, D. L., & Larson, A. (in press). Aligning a state's alternative standards to regular core content standards in reading and mathematics: A case study. *Applied Measurement in Education.*

Fast, E. F., & Hebbler, S. (2004). *A framework for examining validity in state accountability systems.* Washington, DC Council of Chief State School Officers (CCSSO).

Herman, J. L. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education*. New York: Teachers College Press.

Herman, J. L., Webb, N., & Zuniga, S. (2002). *Alignment and college admissions: The match of expectations, assessments, and educator perspectives.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Herman, J. L., Webb, N., & Zuniga, S. (in press). Measurement issues in the alignment of standards and assessments. *Applied Measurement in Education.*

No Child Left Behind Act of 2001a, Pub. L. No. 107-110, 1 U.S.C. (2002).

No Child Left Behind Act of 2001b, Pub. L. No. 107-110, 3 U.S.C. (2002).

Plake, B. S., Buckendahl, C. W., & Impara, J. C. (2001, June). *A comparison of publishers' and teachers' perspectives on the alignment of norm-referenced tests to Nebraska's language-arts content standards.* Paper presented at the annual Large-Scale Assessment Conference of the Council of Chief State School Officers, Snowbird, UT.

Porter, A. C. (2002). Measuring the content of instruction: uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (in press). Alignment as a teacher variable. *Applied Measurement in Education.*

Rothman, R. (2004). Benchmarking and alignment of state standards and testing. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education*. New York: Teachers College Press.

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.

Schleppegrell, M. J. (2004). *Language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shannon, P. D. (n.d.) *Lungs in the rib cage.* Retrieved April 25, 2005 from Lesson Plans Page Web site: http://www.lessonplanspage.com/SciencePELungs-TheBreathingMachineRevisedVersion25.htm

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.

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 Mathematics* (pp. 221-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

U.S. Department of Education. (2003). *Part II-Final non-regulatory guidance on the Title III state formula grants program: Standards, Assessment and Accountability* (Draft February, 2003). Washington, DC: US Department of Education, Office of English Language Acquisition.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Council of Chief State School Officers and National Institute for Science Education). Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states.* Madison, WI: University of Wisconsin-Madison, National Institute for Science Education.

Webb, N. L. (In press). Issues related to judging the alignment of curriculum standards and assessment. *Applied Measurement in Education.*

WestEd. (2004). *Alignment study: Louisiana content standards and the Louisiana Alternate Assessment (LAA).* San Francisco, CA: WestEd.

Wixson, K. K., Fisk, M. C., Dutro, E., & McDaniel, J. (2002). *The alignment of state standards and assessments in elementary reading* (CIERA Rep. #3-024). Ann Arbor: University of Michigan School of Education, Center for the Improvement of Early Reading Achievement.

Author Note

Alison L. Bailey, Department of Education, UCLA and CSE/CRESST; Frances A. Butler, Manhattan Beach, California.

Bailey and Butler developed the original Standards Linkage Protocols for a pilot study funded by the California Department of Education (CDE). WestEd staff was asked to collaborate in the refinement and implementation of the protocols. The authors thank Rachel Lagunoff and Peter Worth for research assistance, and Joan Herman, Noreen Webb, Mark Fetler, Jeanette Ganahl, Kent Hinton, and Anne Katz for helpful comments and suggestions.

Correspondence concerning this article should be addressed to Alison L. Bailey, Department of Education, Moore Hall, Box 951521, UCLA, Los Angeles, California 90095-1521.