

**California English Language Development
Standards & Assessment:
Evaluating Linkage & Alignment**

Study conducted by CTB/McGraw-Hill
for the State of California Department of Education

Report compiled and written by
Anne Murphy, Alison Bailey, & Frances Butler

July 21, 2006

Acknowledgements

We would like to acknowledge the many people who contributed their time and expertise to this study from its initial conception through its design, implementation, and interpretation of results. Edynn Sato of WestEd and the other members of the CELDT Technical Advisory Group provided the methodological framework and gave invaluable advice at each stage of the study. Mark Fetler, Fred Dobb, and Monte Blair of the California Department of Education and Ross Green, Lorena Houston, and Deborah Busch of CTB/McGraw-Hill also provided leadership and expert advice. The ex raters who conducted content analyses with dedication and skill were Pam Herron, Julie Smiley, Peggy Tharpe, Gabriel Martinez, Lisa Maria DiFranco, Teri Werner, Kenneth Boyte, and 15 expert California educators. Michelle Boyer directed the Alignment Workshop and provided technical support along with Marie Huchton and Jason Farris of CTB/McGraw-Hill. Kobi Goessel, Sandra Snell, Beth Naranjo, and Hilary Silver gave programmatic support, and Wendy Roscher led copy editing of the final report.

Table of Contents

Acknowledgements.....	3
Table of Contents.....	4
List of Tables	6
List of Figures.....	8
List of Abbreviations	9
Definition of Terms.....	10
Executive Summary	12
Introduction.....	15
Literature Review.....	19
Research on Alignment of Content Standards and Assessments.....	19
Research on Linkage Between Sets of Standards.....	22
Defining Language Demands for Correspondence Across Standards.....	22
Pilot Study Summary	24
Methods & Analysis	29
Materials & Protocol.....	30
Reviews & Ratings	33
Data Summarization & Evaluation	34
Results.....	37
Overall.....	37
Linkage	37
Alignment	44
Protocol Evaluation.....	51
Discussion.....	53
Findings & Implications	54
Conclusion	59
References.....	61
Appendix A.....	71
Linkage & Alignment Protocol.....	71

Appendix B 75
 Decision Rules Developed 75
 Appendix C 77
 Crosswalk Analysis Procedure 77
 Appendix D 82
 Linkage Results 82
 Appendix E 91
 Alignment Results 91
 Appendix F 120
 Qualitative Notes 120
 Appendix G 125
 Rater Participants 125
 Appendix H 128
 Alignment Study Workshop Agenda 128
 Appendix I 130
 Consensus Data Collection Requirements 130

List of Tables

Table 1 Linkage Analysis Language-Demand Code Frequencies	38
Table 2 Linkage Analysis Percent Ratability of Standards	39
Table 3 Linkage Analysis at Percent Each Modality Rating	40
Table 4 Linkage Analysis Percent at Each Complexity Rating	41
Table 5 Alignment Analysis Item Frequencies by Grade Span	44
Table 6 Alignment Analysis of CELDT Listening	45
Table 7 Alignment Analysis of CELDT Speaking	46
Table 8 Alignment Analysis of CELDT Reading	46
Table 9 Alignment Analysis of CELDT Writing	47
Table 10 Percent Items by Proficiency Level Coded by Modality & Complexity: K – 2	48
Table 11 Percent Items by Proficiency Level Coded by Modality & Complexity: 3 – 5	48
Table 12 Percent Items by Proficiency Level Coded by Modality & Complexity: 6 – 8	49
Table 13 Percent Items by Proficiency Level Coded by Modality & Complexity: 9 – 12	49
Table 14 Strengths and Weaknesses of CELDT-to-ELD Standards Alignment	50
Table 15 Strengths of the Linkage/Alignment Protocol	51
Table 16 Recommendations for Improvement of the Linkage/Alignment Protocol	52
Table D.1 Grade 2 Linkage Rating Results: Ratability, Modality, & Complexity	83
Table D.2 Grade 2 Linkage Rating Results: Language Demands	84
Table D.3 Grade 5 Linkage Rating Results: Ratability, Modality, & Complexity	85
Table D.4 Grade 5 Linkage Rating Results: Language Demands	86
Table D.5 Grade 7 Linkage Rating Results: Ratability, Modality, & Complexity	87
Table D.6 Grade 7 Linkage Rating Results: Language Demands	88
Table D.7 Grade 9 Linkage Rating Results: Ratability, Modality, & Complexity	89
Table D.8 Grade 9 Linkage Rating Results: Language Demands	90
Table E.1 Grades K – 2 Alignment Rating Results Summary	92
Table E.2 Grades K – 2 Alignment Rating Results by ELD Proficiency Level	93
Table E.3 Grades K – 2 Listening Alignment Results	94
Table E.4 Grades K – 2 Speaking Alignment Results	95
Table E.5 Grade 2 Reading Alignment Results	96

Table E.6 Grade 2 Writing Alignment Results.....	97
Table E.7 Grades 3 – 5 Alignment Rating Results Summary.....	98
Table E.8 Grades 3 – 5 Alignment Rating Results by ELD Proficiency Level.....	100
Table E.9 Grades 3 – 5 Listening Alignment Results.....	101
Table E.10 Grade 3 – 5 Speaking Alignment Results	102
Table E.11 Grades 3 – 5 Reading Alignment Results	103
Table E.12 Grades 3 – 5 Writing Alignment Results	104
Table E.13 Grades 6 – 8 Alignment Rating Results Summary.....	105
Table E.14 Grades 6 – 8 Alignment Rating Results by ELD Proficiency Level.....	107
Table E.15 Grades 6 – 8 Listening Alignment Results.....	108
Table E.16 Grades 6 – 8 Speaking Alignment Results.....	109
Table E.17 Grades 6 – 8 Reading Alignment Results	110
Table E.18 Grades 6 – 8 Writing Alignment Results	111
Table E.19 Grades 9 – 12 Alignment Rating Results Summary.....	112
Table E.20 Grades 9 – 12 Alignment Rating Results by ELD Proficiency Level.....	114
Table E.21 Grades 9 – 12 Listening Alignment Results.....	115
Table E.22 Grades 9 – 12 Speaking Alignment Results.....	116
Table E.23 Grades 9 – 12 Reading Alignment Results	117
Table E.24 Grades 9 – 12 Writing Alignment Results	118
Table E.25 Overall Percent Alignment Frequency Distribution for All Modalities.....	119
Table F.1 Example Notes on Science Standards	120
Table F.2 Example Notes on Mathematics Standards	121
Table F.3 Example Notes on ELA Standards	122
Table F.4 Example Notes on ELD Standards	123
Table F.5 Example Notes on CELDT Items.....	124
Table G.1 Linkage & Alignment Rater Groups.....	125
Table G.2 Demographic Information for Raters.....	126

List of Figures

Figure 1 Linkage and Alignment Schema for Integration of Title III and Title I.....	17
Figure 2a Schema Representing the Pilot Study Process for Evaluating Linkage and Alignment	27
Figure 2b Schema Representing the Process for Evaluating Linkage and Alignment in the Current Study	30
Figure A.1 Language Demands and Definitions.....	72
Figure A.2 Levels of Complexity for Rating Language Demands	73
Figure A.3 Linkage Rating Sheet.....	74
Figure A.4 Alignment Rating Sheet.....	74
Figure C.1 Crosswalk Table Design	78
Figure C.2 Example Image of Crosswalk Analysis	81

List of Abbreviations

<i>CCSSO</i>	Council of Chief State School Officers
<i>CDE</i>	California Department of Education
<i>CELDT</i>	California English Language Development Test
<i>CRESST</i>	National Center for Research on Evaluation, Standards, and Student Testing
<i>DOK</i>	Depth of Knowledge
<i>EL</i>	English Learner
<i>ELA</i>	English Language Arts
<i>ELD</i>	English Language Development
<i>NCLB</i>	No Child Left Behind Act of 2001

Definition of Terms

The definitions below pertain to the terms as they are used in this report.

<i>Alignment</i>	The correspondence between the content of the ELD test and the ELD standards
<i>Breadth</i>	Range in types of language demand
<i>Complexity</i>	Relative number of parts or elements represented in a standard or assessment item and operationalized as the number of elements relevant to an item's or standard's language demands (e.g., number of modalities, pictures, distinctions required, references between different elements; for more, see Appendix B)
<i>Content Standards</i>	Statements that define the knowledge, concepts, and skills students should acquire at each grade level in a given content area
<i>Depth</i>	Range of complexity of language demands
<i>Discourse</i>	Language organized into oral or written text (beyond the level of a sentence)
<i>ELD Standards</i>	Statements that define the knowledge, concepts, and skills students should acquire to be proficient in English in the four language domains of Listening, Speaking, Reading, and Writing
<i>Frameworks</i>	Blueprints for implementing the content standards, including suggested curriculum and pedagogy to be used in standards-based education
<i>Language Demands</i>	Linguistic features (phonological, lexical, syntactic) and language functions (discourse) frequently encountered and commonly learned in the K – 12 academic context
<i>Language Function</i>	Part of discourse which focuses on communicative purpose; pragmatic use of language
<i>Lexical</i>	Pertaining to the words of a language as distinct from grammar; vocabulary
<i>Linkage</i>	The correspondence between the content of the ELD standards and content standards

<i>Modality</i>	Four response modes used in language proficiency assessments under Title III: <i>listening, speaking, reading, and writing</i>
<i>Objective</i>	Statements used to specify achievable learning goals within a given ELD or content standard
<i>Phonological</i>	Pertaining to the sounds of language; often related to Reading skills
<i>Ratability</i>	Identification of language demands within a standard statement, objective, or assessment item
<i>Standard</i>	Statements that define the knowledge, concepts, and skills students should acquire to be proficient at a specific grade level in a given content area or discipline (e.g., ELD, ELA, Mathematics, Science)
<i>Syntactic</i>	Pertaining to grammatical forms as distinct from vocabulary; grammar
<i>Title III</i>	Section 3001 of the No Child Left Behind Act of 2001: Language Instruction for Limited-English-Proficient and Immigrant Students

Executive Summary

Under Title III of No Child Left Behind Act of 2001 (NCLB), English language development (ELD) standards must be linked to content standards and the English language proficiency test must be appropriately aligned to the ELD standards. States are held accountable for the academic progress of English learners (EL) and for their ability to sustain academic performance after they exit EL programs. This report details the evaluation study conducted for the State of California to provide evidence of linkage and alignment. *Linkage* is defined here as the correspondence between the ELD standards and content standards, and *alignment* is defined here as the correspondence between the ELD standards and the ELD assessment.

Sato, Lagunoff, Worth, Bailey, and Butler (2005) developed an approach appropriate to assessments of ELD to evaluate linkage and alignment based on research in the areas of education and applied linguistics. Their approach looked for correspondences across standards and tests in terms of *language demands*. The language demands include both *linguistic features* (phonological, lexical, syntactic) as well as *academic language functions* (discourse) commonly learned in the K – 12 context. This study used the language-demands approach in the evaluation of linkage of the California ELD standards to the state content standards. The content areas of English Language Arts, Mathematics, and Science were analyzed for four grade levels (Grades 2, 5, 7, and 9); in Grade 9, Biology and Algebra I were analyzed for Science and Mathematics, respectively. Also, the study evaluated the alignment of the Form E California English Language Development Test (CELDT, CTB/McGraw-Hill, 2005) to the ELD standards. All ELD standards at four grade spans (K – 2, 3 – 5, 6 – 8, and 9 – 12) and 396 CELDT Form E items were used to evaluate alignment.

Analyses examined (a) *ratability* (whether sufficient evidence exists in the items themselves to rate language demands); (b) *modality* (whether an item’s language demands occur in *listening, speaking, reading, and writing*); (c) *complexity* (related to depth-of-knowledge and the nature of the content or skills reflected in an item or

standard); and, (d) specific *language demands* (types of linguistic skill and/or academic language function codes).

A total of 1,381 content and ELD standards were analyzed with an overall ratable of 74.0%. Few standards in Mathematics and Science provided enough information to allow for a rating on a language dimension (21.1% and 21.9%, respectively) while most ELD and English Language Arts (ELA) standards were ratable (96.2% and 95.8%, respectively). A total of 396 CELDT Form E items were analyzed with 100% ratable. These 396 items, 99 per grade span, represented 20 ELD standards in Grade K – 2, 25 in Grades 3 – 5, 21 in Grades 6 – 8, and 19 in Grades 9 – 12.

Overall, the findings of this study confirm many of the results reported by Sato, et al. (2005). Linkage was inconsistent across groups of standards. Most ELD and ELA standards were ratable (greater than 90%); however, only 13% to 38% of the Math and Science standards were ratable using both the standards themselves and accompanying frameworks documents. There were variable degrees of alignment between ELD standards and the CELDT depending on language demands and proficiency levels. Strongest alignment appeared in items in the Reading and Writing sections on the language-demand dimensions. Reading and Speaking items generally showed moderate to strong alignment across sections of the test on all dimensions (i.e., ratable, modality, complexity, and language demands), especially in the lower two grade spans. Items in the Writing sections were weakly aligned on the modality and complexity dimensions, and Listening items showed the weakest alignment on all dimensions. The 6 – 8 grade span showed the weakest alignment of the four spans, especially on complexity and language-demand dimensions.

Implications of the study are described below.

1. If the correspondence between California’s ELD standards and content standards is strengthened on all language dimensions, the CELDT can be better aligned to the content standards. The development of CELDT items that reflect more academic language functions and higher levels of complexity would improve overall alignment of the system. (See Figure 1 for a visual representation of the

various relationships involved in assessment systems including Title III and Title I requirements.)

2. The CELDT item development should target the following priorities for the purpose of strengthening alignment: (a) items that incorporate academic language functions in the form of high-frequency, non-specific, academic vocabulary; (b) items that address specific language demands at the word and sentence levels (e.g., *identification, labeling and enumeration, description*); and (c) items that reflect greater depth (i.e., complexity). Given that the ELD standards and the CELDT tend to have lower complexity ratings, higher levels of complexity in item development will improve alignment, as well as the test's ability to discriminate at the higher proficiency levels.

3. Although classroom and formative assessments are well suited to assess certain academic language functions at the discourse level, increased emphasis on academic language functions is needed in order to bring the ELD standards and CELDT into better compliance with NCLB Title I and Title III requirements (Figure 1).

4. Each content area has standards with unique linguistic demands, lexicons, expectations, and assumptions. Further work to understand these unique aspects of content standards and how they are related on the language demands dimensions would help to inform the effort to bring the ELD standards and the CELDT into better compliance with NCLB Title I and Title III requirements (Figure 1).

5. The linkage and alignment methodology used for this study promises to be a useful tool for item development and has the potential for aiding in the development of curricular materials. The approach taken here will help test developers and state assessment decision makers by to improve the processes that guide item development and fill gaps in alignment.

Introduction

Under Title III of No Child Left Behind Act of 2001 (NLCB), English language development (ELD) standards must be linked to content standards and the English language proficiency test must be appropriately aligned to the ELD standards. States are held accountable for the academic progress of English learners (EL) and for their ability to sustain academic performance after they exit EL programs. Bibian's (2006) case study of a California urban middle school found that over half of the re-designated EL students were returned to EL programs after three years after failing to sustain their academic performance. Bibian suggested that her findings illustrate the need to better align the Title III system with the Title I system.

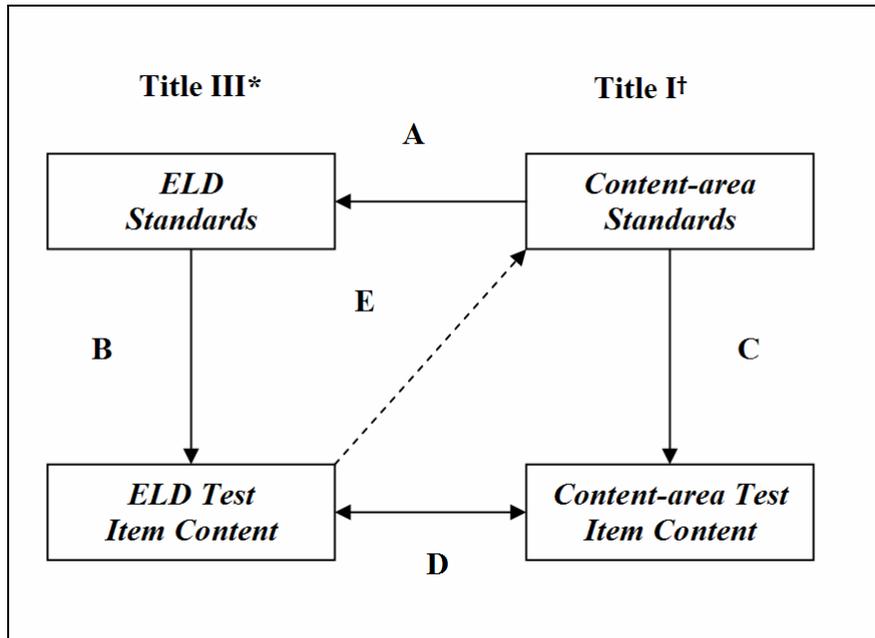
This report details the approach taken by the State of California to provide evidence of this linkage and alignment. As part of this approach, the California Department of Education (CDE) sponsored a pilot study at the fifth-grade level to determine the effectiveness of linkage and alignment methods and to refine the evaluation instruments (Sato, Lagunoff, Worth, Bailey, & Butler, 2005). Procedures and results of the pilot study were then used as a basis for the broader evaluation presented here.

The pilot study developed processes for the analyses of (a) *linkage*, or the correspondence between the ELD standards and the content standards, and (b) *alignment*, or the correspondence between the ELD test and the ELD standards.¹ Because a purpose of developing these processes was to evaluate language proficiency assessment systems, a methodology was developed in terms of *language demands*, those linguistic skills and academic language functions stated both explicitly and implicitly in the standards, frameworks, and test items. Language demands appeared to be the most appropriate dimension on which to determine correspondences between different sets of standards and between the state's ELD standards and the California English Language Development Test (CELDT).

¹ The correspondence or relationship between two types of standards is referred to here as *linkage* and, following much of the literature, reserves the term *alignment* for the correspondence between a test and standards.

For the current study, language demands were operationalized as the linguistic features (phonological, lexical, syntactic) and academic language functions (discourse) frequently encountered and commonly learned in the K – 12 academic context. The current report extended the work of the pilot study to include additional grade levels and added Mathematics and all English Language Arts (ELA) standards to the existing fifth-grade level findings. Mathematics, ELA, and Science are core content areas assessed under NCLB and assumed to be essential bodies of knowledge and skill for all students. For analysis of the content standards, the sample of grades was selected (Grades 2, 5, 7, and 9). Given limitations of time and resources, these grades were selected to represent (a) the range of primary, middle, and secondary standards; (b) a range of developmental levels; and (c) each of the ELD grade spans (K – 2, 3 – 5, 6 – 8, and 9 – 12). In Grade 9, Biology and Algebra I were used for Science and Mathematics, respectively, since Science and Mathematics standards were written for the high-school, 9 – 12 grade span. All ELD standards and CELDT Form E items were used in the current study to cover the K – 12 grades.

The NCLB of 2001 stated the need to establish clear relationships between standards and tests (e.g., Bailey, in press). Figure 1 provides a visual representation of the relationships between standards and testing in the NCLB context.



*Title III: Language instruction for limited English proficient and immigrant students (NCLB, 2001a)

†Title I: Improving the academic achievement of the disadvantaged. (NCLB, 2001b)

*Figure 1 Linkage and Alignment Schema for Integration of Title III and Title I (Copyright permission granted by National Center for Research on Evaluation, Standards, and Student Testing and to reproduce Figure 1 by Bailey, Butler, & Sato, 2005. Also reprinted in *Applied Measurement of Education* by Bailey, Butler, & Sato, in press.)*

This study focused on the Title III portion of Figure 1, in which arrows represent the relationship between content standards and ELD standards (Arrow A) and the relationship between the ELD standards and the ELD test (Arrow B). The relationship represented by Arrow A, which captures the linkage between the two types of standards, is intended to stress the importance of the ELD standards that reflect the type of language inherent in the content areas. Relationship B, which illustrates the progression from ELD standards to the state ELD test, shows the continued emphasis on the language identified in the content areas. The broken arrow labeled E between the ELD test and the content standards illustrates the potential for states to demonstrate alignment between their ELD assessment and their content standards. The totality of the Title III relationships

illustrated in Figure 1 can potentially provide strong evidence of a coherent system in response to NCLB mandates.

Underlying linkage and alignment efforts such as those illustrated in Figure 1 and described in this report is the assumption that the content standards (the starting point) include valid representations of the constructs to be learned and assessed – in this case, the language of school (as well as the content material for each discipline). However, content standards have not traditionally included explicit statements of the type of language needed to demonstrate knowledge of the content in a given area (Butler, Lord, Stevens, Borrego, & Bailey, 2003/2004). A process such as the one reported here offers the opportunity to evaluate the quality of standards and tests *vis a vis* the articulation of language demands and to implement procedures for improvement where weak linkage or alignment is identified. In the current study, the following questions were addressed:

1. In what ways and to what degree do the state ELD standards reflect the language demands of state academic content standards?
2. In what ways and to what degree does the CELDT reflect the language demands of state ELD standards?
3. In what ways and to what degree is the evaluative process effective in yielding evidence to help the state meet requirements of the NCLB Act?
4. What implications do the analyses suggest for the improvement of both standards and the CELDT?

The overriding goals of the current effort were to (a) establish a process by which the State of California can continuously monitor its efforts to assure linkage between the ELD and content standards and (b) evaluate the alignment of the CELDT with the ELD standards. The state strives to refine its ELD standards to more closely reflect the language of school across content areas. Results from this study will inform future item development for the CELDT.

Literature Review

This review summarizes four areas relevant to the study. The first section describes prior research on establishing alignment between content standards and assessments designed to test knowledge in those content areas. Next, it is established that little research has evaluated the linkage between different sets of standards across different content areas. The third section provides a brief overview of *language demands* in the academic context and their benefits as an analytical tool. Finally, a summary of the pilot study methods, main findings, and recommendations is presented as a basis for the current, expanded study (Sato, et al., 2005).

Research on Alignment of Content Standards and Assessments

Prior work in the area of standards and test alignment has evaluated the correspondence between standards and test items. These studies have created procedures to evaluate the degree to which test items measure skills described in standards documents, and they have focused primarily on three main dimensions of concordance: content, articulation of alignment across grade and age levels, and equity concerns (e.g., Herman, Webb, & Zuniga, in press, 2002; Plake, Buckendahl, & Impara, 2001; Porter, 2006, 2002; Porter, Smithson, Blank, & Zeidner, in press; Rothman, 2004; Webb, 2006, 1999, 1997; Webb, Horton, & O’Neil, 2002; WestEd, 2004a, 2004b, 2004c, 2003; Wixson, Fisk, Dutro, & McDaniel, 2002).

Within the content dimension, researchers have focused on the following to evaluate the degree of alignment: categorical concurrence of content, depth of knowledge (DOK) consistency, range of knowledge correspondence, and balance of representation of content of the standards on a given test. *Categorical concurrence of content* evaluates to what degree set standards and assessments incorporate the same content. *DOK consistency* refers to the match in complexity of content contained on the standards and related assessments. Webb (1997) placed complexity on a scale of 1 – 4 denoting the shift from simple Level-1 recall, through Level-2 skill/concept processing, to Level-3 strategic thinking to Level-4 extended thinking or complex reasoning over time. *Range of*

knowledge correspondence captures comparability in the breadth of knowledge in the standards and assessments. *Balance of representation of content* evaluates the degree to which objectives within standards are emphasized on a test. The distribution of objectives should be equivalent across standards and assessments.

Articulation of standards to test alignment across grade and age levels has dealt with issues of *cognitive soundness* and *cumulative growth order* so that assessments can best portray expected developmental sequence for knowledge and concept learning. Additionally, Webb (1997) has called for alignment studies to pay attention to equity and fairness of standards and assessments as they reflect the backgrounds of all students.

In the wake of the NCLB requirement that states show how standards and their related assessments are aligned (in terms of *comprehensiveness*, *content*, *emphasis*, and *depth*, among other criteria), Webb (in press) raised the question of how to judge what an acceptable level of alignment is. In the past, studies have presented percentages of correspondence or created an index of alignment (e.g., Webb, 1997; WestEd, 2003). However, what percentage (i.e., 50%, 75%, anything less than 100%) is sufficient for alignment to be judged as good? Webb (2006) stressed that alignment is ultimately a content analysis that needs to be conducted by content experts, and the interpretation of the results of an alignment study ultimately depends on the purpose of the assessments under scrutiny. As Webb (in press) points out in an example of this recommendation applied to DOK consistency,

If the purpose of the assessment is to differentiate between students who are proficient from students who are not, then an argument could be made that all or nearly all of the item DOK levels would be the same as the DOK levels of the corresponding objectives [standards]. However, if the purpose of the assessment is to place students on a range of proficiency levels (e.g., below basic, basic, proficient, and advanced), then it is reasonable to have items with a range of DOK levels in comparison to the objectives.

The CELDT reflects the notion of a range of proficiency levels as already built into the ELD standards (unlike standards in other content areas). Thus correspondence for criteria such as DOK across the CELDT and the ELD standards needs to be maintained. Otherwise, the standards and the test will be mismatched.

In addition to these past studies of alignment, the following key questions helped to focus the pilot study and the expanded study reported here (Fast & Hebbler, 2004, p. 33):

1. Are the assessments and the standards aligned comprehensively, meaning that the assessments reflect the full range and depth of the standards? Are the assessments as cognitively challenging as the standards?
2. Are the assessments and the standards aligned in terms of content and process, meaning that the assessments measure what the standards state students should both know and be able to do?
3. Do the assessments reflect the same degree and pattern of emphasis on the content as are reflected in the state's academic content standards?
4. Do the assessments yield scores that reflect the full range of achievement implied by academic achievement standards?
5. Do the assessments measure the knowledge and skills described in [the] academic content standards and not knowledge, skills, or other characteristics that are not specified in the academic content standards?
6. How are gaps and weaknesses identified, and what is done to improve the alignment of [the] standards and assessments? (p. 33).

These questions provided the pilot study (Sato, et al., 2005) and the present study an appropriate and best-practice framework for evaluating linkage and alignment.

Research on Linkage Between Sets of Standards

Recently, WestEd and others have evaluated the degree of correspondence between standards of the same content area. Specifically, studies have created methodologies for standards-to-standards comparisons of general education and special education of the same content area (e.g., ELA) for various states (Eckhout, Plake, Smith, & Larson, in press; WestEd, 2004a, 2004b, 2004c, 2003). However, standards-to-standards linkage studies have not generally compared a particular aspect or dimension of learning (e.g., language demands) *across* different content areas. As Bailey, et al., (in press a) pointed out:

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, overlap with ELD standards in terms of implicit and explicit language demands placed on students.

The pilot study (Sato, et al., 2005) summarized below, developed methods for evaluating the degree of correspondence across content areas on an academic English language dimension. Specifically, the language demands found to be common to both the ELD and content standards can provide the mechanism for evaluating the current nature and degree of linkage between standards and standards-based instructional materials (e.g., standards-based textbooks, lesson plans, and other curriculum materials).

Defining Language Demands for Correspondence Across Standards

To date, relatively little empirical study exists regarding the nature of academic English language and the specific language demands placed on EL students at the point that they are redesignated as fluent English proficient and enter mainstream classrooms. The paucity of empirical information in this area may be due to the finding that teachers

are rarely explicitly aware of their language expectations for students (Wong Fillmore & Snow, 2000). Research exists on the second-language-acquisition process: namely, studies of differential outcomes for learners (see Larsen-Freeman, 2000, for review) and studies of English-for-specific-purposes (e.g., as the medium for acquiring discipline-specific knowledge, Douglas, 2000). However, this research has been conducted almost exclusively at the college and graduate level with educated, adult, English-as-a-Second-Language learners. The K – 12 arena appears to be ignored perhaps due to the difficulty in articulating norms for both first- and second-language-development in school-age children.

What empirical research has been conducted suggests that there are language demands at all levels of use. At the lexical level, Stevens, Butler, and Castellon-Wellington (2000) and others (e.g., Scarcella & Zimmerman, 1998) identify three categories of words: (a) high-frequency general words, or words used regularly in everyday or social contexts; (b) nonspecialized academic words, or words that are used in academic settings across content areas; and (c) specialized content-area words, or academic words unique to specific content areas (i.e., Math and Social Science). Specialized, content-specific language includes the conceptual terminology of an academic discipline (e.g., *sedimentary*, *tectonic*, and *alluvium* as part of a large lexicon of geologic terms). Non-specialized language that cuts across content areas is a form of general academic language not specific to any one content area. For example, words such as *examine* and *cause* contrast with equivalent everyday words such as *look at* and *make* but could be encountered in ELA, Mathematics, Science, or Social Studies classroom contexts (Cunningham & Moore, 1993).

At the syntactic level, there are features of English grammar that are highly stylized or encountered primarily in Writing (Schleppegrell, 2004). These demands can be in the form of grammatical structures used in academic contexts to convey certain types of information (e.g., the comparative *greater/less than*, Spanos, Rhodes, Dale & Crandall, 1988) and may even occur more frequently in some content areas than others. At the discourse level, (i.e., the functional or organizing level of written and spoken text), academic language functions in classroom talk and textbooks, such as *explanation* and *description*, are found to differ in frequency of occurrence across content areas (Bailey,

Butler, LaFramenta, & Ong, 2001/2004; Bailey, Butler, Stevens, & Lord, in press; Butler, Bailey, Stevens, Huang, & Lord, 2004). Moreover, several language functions can play a supporting role to an overarching language function; for example, *explanations* may contain within them *definitions* and *descriptions* and may have associated syntactic structures for conveying the language function (Butler, et al., 2004).

In sum, language demands at the lexical, syntactic, and discourse levels exist in academic programs. Given that little evidence exists to describe methods for evaluating test-to-standards alignment in terms of language demands and given a lack of information on appropriate methods and expectations for establishing linkage between different sets of standards, the following pilot study was conducted.

Pilot Study Summary

As a first step in the work reported here, the CDE sponsored the pilot study to examine the effectiveness of proposed methods in yielding evidence to meet the NCLB requirements. Specifically, the pilot study was an initial attempt to systematically examine the relationships among language demands of the ELD test, ELD standards, and content standards. The pilot reviewed the standards, frameworks, tests, and selected lesson plans, using protocols specifically designed to highlight linkages and alignments. Analyses were conducted, using fifth-grade ELA Reading and Science standards, Grades 3 – 5 ELD standards, and ELD test materials. The research questions follow:

1. To what degree are the state ELD standards linked to state academic content and achievement standards in order to ensure that English Learners attain the level or range of English language proficiency needed to facilitate progress in the content areas?
2. To what degree does the ELD test reflect the full range and depth of the ELD standards?
3. To what degree is each of the three protocols used in this pilot study effective in yielding evidence that will help states meet the

requirements of the NCLB Act? What considerations and refinements are needed, if any?

The pilot methodology, which was adapted with minor differences for the current study, is discussed below. In brief, researchers used protocols that allowed for identification of language characteristics of the content standards, ELD standards, and ELD test to (a) assess linkages between ELD standards and ELA Reading and Science standards on the dimensions of *ratability* (identification of language demands), *breadth* (range of language demands), and *depth* (range of complexity of language demands); (b) evaluate the linkage between the ELD standards and instructional Science materials (a proxy for the state Science standards) for *listening* and *speaking* language demands²; and (c) determine ELD test to ELD standards alignment for ratability, breadth, depth, and *weighting* (distribution of items on the test; i.e., a standard may have greater weight or emphasis on the test if there are more test items aligned to it than to other standards). Bailey, et al., (2005) anticipated from prior work on standards (Bailey & Butler, 2002/2003) and language demands (Bailey, et al., 2001/2004, in press b; Butler, et al., 2004) that the content standards and frameworks may not always contain information about language that is sufficiently explicit for rating language-demand type and complexity. The pilot study, therefore, also created a second exploratory protocol for evaluating the nature and degree of linkage between the language demands expressed in the ELD standards and *standards-based* instructional materials (i.e., Science lesson plans) in the anticipation that more explicit language demands would be articulated in such material (a report of these findings is provided in Bailey, et al., in press a).

Four key findings from the pilot study follow (Sato, et al., 2005):

1. Linkage was inconsistent across the pairings in the pilot study (i.e., every ELA Reading standard could be coded to at least one language demand yielding 100% ratability, whereas only 34% of

² Since only 34% of the fifth-grade science standards were ratable, an alternative approach to linkage was explored using instructional materials.

the Science standards were ratable, making it difficult to assess linkage with the ELD standards).

2. The ELD standards were characterized by a lower level of complexity when compared with the ELA standards or the language demands found in the Science standards.
3. Alignment between ELD standards and the ELD test varied by proficiency level, in that some levels of proficiency received more attention than others.
4. In a given ELD test section (Listening, Speaking, Reading, or Writing), items did not necessarily correspond to the section's name when evaluated in terms of language-demand modality. For example, an item from the Writing section may be coded as the *reading* modality.

Figure 2a is a schematic overview of the linkage and alignment evaluation process created for the pilot study (Sato, et al., 2005) and developed in the current study (see Figure 2b). The process was organized into four main steps: (a) *materials and protocols*, (b) *reviews and ratings*, (c) *data reduction*, and (d) *evaluating matches*. Each step required analyses of source material that resulted in new information as indicated by arrows *within* boxes. This information was then fed into the subsequent step as the source material for analysis at the next stage, as indicated by arrows *between* the boxes. At each step, *decision rules* were created (indicated by arrows from this box to all steps; see Appendix B). These rules helped ensure accurate and consistent application of the protocols. In this way, any important modifications to the general process were carefully documented, often required input and, in some instances, ratification by the State (e.g., which standards documents, instructional materials, test blueprints, etc., to include for review).

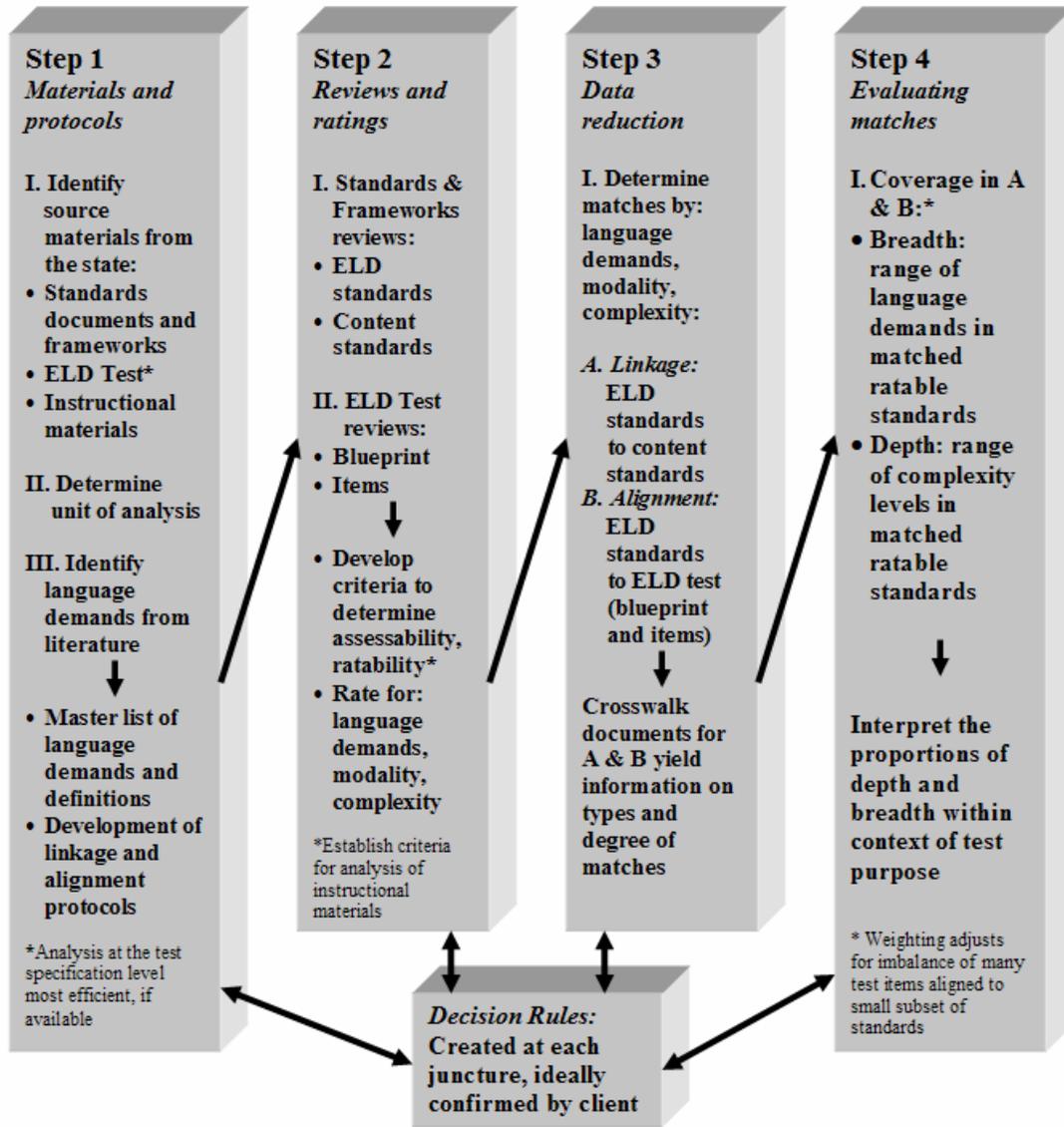


Figure 2a Schema Representing the Pilot Study Process for Evaluating Linkage and Alignment

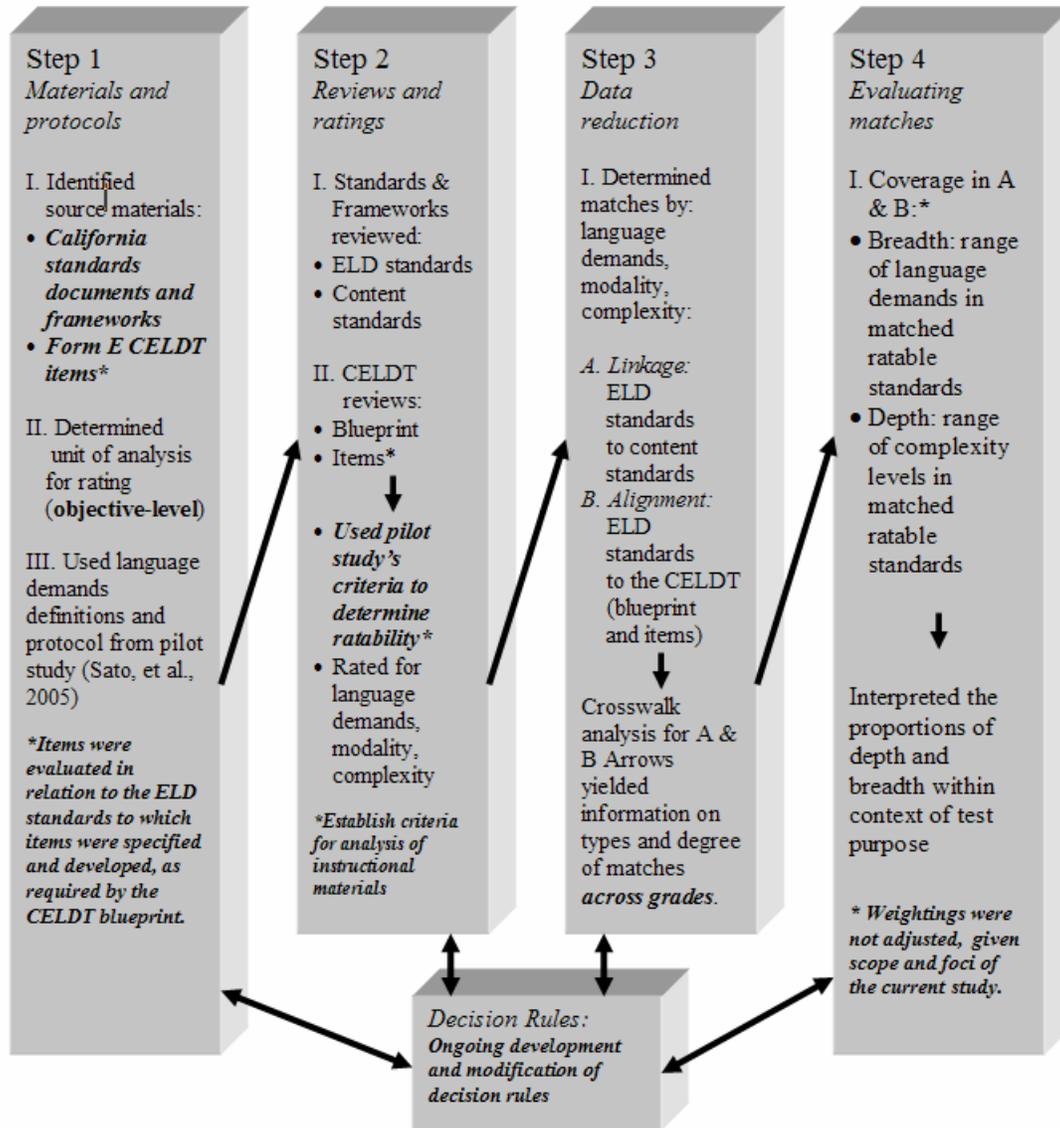
The initial findings from the pilot study showed that, while evidence of linkage and alignment exists to some degree between all of the pairs analyzed at the fifth-grade level, continuation of the analyses at additional grade levels is essential to provide a more complete picture. Based on the fifth-grade evidence, future work is justified in order to bring the entire Title III system (see Figure 1) to a level that will assure compliance with NCLB and will best serve EL students across California.

Methods for evaluating linkage and alignment have traditionally not used a language-demand approach when establishing relationships between standards and ELD assessments. However, language demands have been shown to be a powerful tool for analyzing the relationships between content standards, ELD standards, and all elements of assessment programs (Sato, et al., 2005). The current study used the language-demand approach from the pilot to evaluate alignment between ELD and content standards as well as the linkage between ELD standards and the CELDT. The approach was applied with some refinement across four grade levels, three content areas, and the ELD standards at each grade span. (Figures 2a and 2b illustrate each method.) The study also aimed to refine and further develop protocols and the analytical framework to better apply ratability, language-demand types, modalities, and complexity levels to standards and assessment data (see Appendix A for operational definitions of language demands and complexity levels).

Methods & Analysis

This linkage and alignment evaluation study used a two-part, document-review methodology to identify and analyze (a) linkage between content and ELD standards and (b) alignment between ELD standards and the CELDT. The methodology built on the work of Sato, et al., (2005, see Figure 2a) to broaden the scope of the pilot by analyzing data from four grade levels and three content areas. By increasing the number of grade levels and content areas, this study allowed for cross-grade analyses of linkage and alignment and for further evaluation of exploratory linkage and alignment protocols. The study investigated how and to what degree (a) the state ELD standards reflect the language demands of state academic content standards, (b) the CELDT reflects the language demands of state ELD standards, and (c) the evaluative process is effective in providing evidence to help the state meet requirements of NCLB. Finally, the method was designed to provide suggestions for the improvement of both the standards and the CELDT.

Figure 2b illustrates the steps taken to evaluate linkage and alignment. In Step 1, materials and protocol were identified, including the California academic standards and frameworks for analysis, and the level of analysis was determined. Next, standards materials and the CELDT Form E items were reviewed using the pilot protocol and definitions. Data were summarized to evaluate the matches between language demands, modality, and complexity. Finally, matches were evaluated in terms of breadth and depth. Steps 2 – 4 were conducted with some alterations from the pilot study (see Appendix A for more on rating protocol and definitions): unlike the pilot, the blueprint was not a focus of evaluation and was assumed fixed for the purposes of this study. Ongoing analyses and development of decision rules were documented (see Appendix B for decision rules developed during the current study). Next, linkage and alignment were determined, using the crosswalk strategy whereby language demand ratings were cross-tabulated with content and ELD standards (see Appendix C for example crosswalk table). Because this study explored a cross-section of grade levels, further analyses were conducted in which crosswalk data were inspected across grades (see Sato, et al., 2005, Appendix Q, for more on the crosswalk strategy).



Note: Changes made to Figure 2a to represent the current study's method are boldfaced.

Figure 2b Schema Representing the Process for Evaluating Linkage and Alignment in the Current Study

Materials & Protocol

The study gathered and analyzed data in two phases: first standards-to-standards linkage and then standards-to-test alignment (see Figure 2b, Step 2).

Materials

Linkage data were collected by analyzing state standards and their frameworks for ELA, Mathematics, and Science in Grades 2, 5, 7, and 9 and the ELD standards in all grade level ranges (K – 2, 3 – 5, 6 – 8, and 9 – 12; see Appendix D for linkage results; California Department of Education, 2005, 2003, 2002, 1999, 1998a, 1998b, 1997).³ The standards describe what students should know and be able to do for specific facets of a given area of content. Accompanying frameworks develop these standards by describing what teachers can and should do to assist students' attainment of the standards. Frameworks materials were included to more fully model the language demands placed on students in classrooms. Unlike the pilot, lesson plans were not evaluated.

California standards use a range of formats to articulate content standards and objectives. For example, Grade 2 Science standards are grouped into broad categories, such as Physical Sciences and Life Sciences, numbered by standard, and objectives are assigned letters. However, the ELA standards at the same grade are organized by modality (e.g., *reading*) and then by subcategories (e.g., Word Analysis, Fluency, and Systematic Vocabulary Development; Reading Comprehension, etc.) before being numbered at two or three levels of detail. The ELD standards are grouped by language development category and by proficiency level before they are subdivided into three levels. For this study, the following standard definitions were used: *content area* (level 1), *sub-area* (level 2), *standard statement* (level 3), and *objective* (level 4). For example, the content area could be *science*, the sub-area *physical sciences*, the standard statement *standard 1*, with subsequent *objectives a, b, and c* of the standard statement. Standards were evaluated to the objective level; therefore, ratings were conducted to the smallest unit possible (objective or sub-objective levels) in order to promote consistency in the approach across all data.

One goal was to broaden the scope of work done in the pilot study and to apply the language-demand protocol to a range of grades and more of the academic content standards. The time-intensive nature of the protocol precluded a comprehensive study of

³ While all ELD standards and CELDT items were considered at the K – 2 level, only the second-grade students take the CELDT Reading and Writing sections.

all kindergarten to Grade 12 standards. Thus, a representative sample of Grades (2, 5, 7, and 9) was selected for investigation to represent (a) primary, middle, and secondary levels; (b) all ELD grade spans (K – 2, 3 – 5, 6 – 8, and 9 – 12); and (c) a range of developmental levels. For all study grades, except Grade 9 Math and Science, academic content standards are described by grade level. However, all Grade 9 Math and Science standards are included within the Grades 9-12 span. Therefore, Algebra I and Biology standards were analyzed, since these content areas are most commonly taught in Grade 9 throughout the State of California schools (F. Dobb, personal communication, June 5, 2006).

The second phase of data collection was to establish alignment of the ELD test and the ELD standards using the items from Form E of the CELDT (CTB/McGraw-Hill, 2005) and the ELD standards used to develop these items (California Department of Education, 2002). The CELDT Form E was designed in four sections (Listening, Speaking, Reading, and Writing) with items written to the appropriate ELD standards, as defined by the CELDT blueprint. Given the various grouping categories for the different bodies of material (ELD standards and test items), results were reported, using a combination of the CELDT sections and the ELD standards categories (Listening and Speaking, Reading, and Writing; see Tables E.1-E.24) as well as proficiency levels (see Tables E.2, E.8, E.14, and E.20, Appendix E, for results).

The pilot ratings for Grade 5 Science and ELA Reading (Sato, et al., 2005) were used as exemplars and training materials to calibrate raters. These ratings were also included to increase the size of the dataset and to make the dataset consistent across the selected grade levels.

Protocol

The protocol used for both phases of the study was originally developed in the pilot study and further refined (see Appendix A, Figures A.1 – A.4 from Bailey & Butler (2004) and Appendix B). The protocol established four types of decisions to be made by raters: ratability, modality, complexity, and language demands. Ratability was a dichotomous determination of whether a standard or a test item contained enough information to establish the language demands required and was coded as either *sufficient*

or *indeterminate*. If a standard was deemed sufficient to rate, raters determined the specific modalities required: *listening, speaking, reading, or writing*. Using the *Levels of Complexity for Rating Language Demands* (see Appendix A, Figure A.2) and further developed decision rules (Appendix B), a complexity code was assigned to reflect the number of and types of relationships between elements required to meet the standard, where elements included modalities, types of activities, inferences, references, etc. Language demands were then determined using *Language Demands and Definitions* (Sato, et al., 2005, p. 89-92) to reflect the linguistic skills, such as *phonemes* and *writing convention*, and academic language functions, such as *labeling, prediction, and analysis* (see Appendix A for more detail). Seven linguistic function codes and 14 academic language skill codes were used (Figure A.1 provides operational definitions of each code). Raters also qualitatively noted questions, comments, and suggestions for refinement of the protocol or procedure (examples of notes are provided in Appendix F).

Reviews & Ratings

The review of material was carried out by raters who meet specific background criteria. A description of the raters and a discussion of the process they used follows.

Rater Participants

Raters for the study were selected for their (a) expertise in a given content area, (b) experience teaching EL students, (c) linguistics expertise, and/or (d) knowledge of the CELDT. Groups of participants were recruited from within CTB/McGraw-Hill for evaluating linkage and for rating the ELD standards (4 – 5 people per group; see Appendix G for more information about the rater participants and groups). At least one group member overlapped between each group to promote the development of consistent decision rules.

Four groups of educators from across California were recruited (see Appendix G for more information about the raters) to evaluate alignment of the CELDT items by grade span (K – 2, 3 – 5, 6 – 8, and 9 – 12; see Appendix H for workshop agenda). These raters were first trained to understand the concept and definition of language demands, the established decision rules, the process of refining and elaborating on decision rules,

and the process of reaching consensus. After completing ratings, state educators provided feedback on the procedure and protocol.

Rating Procedure

Each standard and item rating was determined through a process of consensus, whereby 3 – 5 raters jointly agreed upon the codes assigned. In the early stages of the study or with new raters, 4 – 5 people were required in each group to attain optimal inter-rater reliability and to develop decision rules. Once decision rules were firmly established, 2 – 3 raters coded data with confirmatory checks by linguistic experts and other trained raters.

To reach consensus, raters first independently completed Linkage or Alignment Rating Sheets (Appendix A, Figures A.3 and A.4). Next, one member of the rating group for a given set of standards or items would gather and review all rating sheets to identify discrepant codes. Due to the amount of data and the time-consuming nature of consensus-building, this second step was developed to increase efficiency. Finally, each group met to discuss and resolve discrepancies while a scribe/timekeeper recorded decisions and notes (see Appendices I for scribe/timekeeper requirements). For alignment ratings of the CELDT items, raters were blind to ELD standard proficiency levels during the rating process.

Data Summarization & Evaluation

As described above, the linkage and alignment protocol involved ongoing standard- and item-level analyses; ratings data were then compiled and analyzed to address ratability, breadth, and depth of the CELDT, ELD standards, and content standards. Analyses included the calculation of frequencies and percentages of sufficiently ratable standards and items, as well as the range of language demands and complexity of those demands.

Linkage Analysis

Using frequency tables and crosswalk tables as defined by Sato, et al., (2005; see Appendix C for crosswalk table design and use), data were analyzed for areas of strong and weak linkage (correspondence between ELD and content standards). This analysis

assigned the standard codes to each cell in a cross-tabulation whereby standards at the objective level populated cells when their assigned language demands corresponded to those of the content or ELD standards. If an ELD standard and a content standard were present for a given language-demand dimension, the standards were considered linked on that dimension. For example, if an ELD standard and a Mathematics standard were both coded with *identification* at a *low* complexity, they were considered linked on both language-demand code and complexity. A cross-grade analysis combined grade-level crosswalk tables across modalities to provide a broader picture of standards-to-standards linkage. This extension allowed for the identification of strengths and weaknesses in language-demand coverage across grades and standards.

Alignment Analysis

Alignment (correspondence between ELD standards and the CELDT items) was determined by comparing items and standards at each rating decision. For ratability and complexity, items and standards were considered aligned if they shared the same code. For modality and language demands, the number of aligned (equivalent) codes and percent alignment of those codes were determined. Specifically, the total difference in language demands (code by code) was computed and divided by the total number of codes of the standard. For example, if an ELD standard were coded with *phoneme* and *identification* and a CELDT item written to that standard were coded as *phoneme*, this pairing received 1 out of 2, or 50%, alignment for language-demand code. If the same items were rated as using *listening* and *writing* modalities and the paired standards rated as *reading* and *writing*, this item-standard pair also received 1 out of 2, or 50%, alignment for modality. Item-by-item analyses are reported in Appendix E.

Once alignment was determined at the item level, the percent alignment was computed by dimension (ratability, modality, complexity, and language demand) for a given grade span. Using overall percent alignment by test section and grade span, a frequency distribution was developed to evaluate relative strength of alignment (see Table E.25).

Protocol Evaluation

For the purpose of refining the protocol and methodology developed in the pilot study, qualitative notes recorded throughout the rating process were compiled and reviewed for relevant themes. Notes from the analyses of ELD and content standards and the CELDT items were sorted and themes identified. Themes were validated and elaborated upon during a group discussion with educator participants after the completion of alignment ratings. Examples of these notes are reported in Appendix F by standard content area or CELDT item, and information from these notes regarding strengths and recommendations are summarized in Tables 15 and 16.

Results

Results of the study showed varying degrees of linkage and alignment across grades, modalities, and subject matter. This section begins with a description of the overall frequencies of assessed standards and CELDT items as well as percent of those standards and items that were ratable. Then, an overview of linkage and alignment results is reported. For ELD and content standards, Appendix D provides complete linkage results by ratability, modality, complexity (Tables D.1, D.3, D.5, and D.7), and language demand frequencies (Tables D.2, D.4, D.6, and D.8). Appendix E reports alignments results by grade span for language demands (Tables E.1, E.7, E.13, and E.19); language demands by proficiency level (Tables E.2, E.8, E.14, and E.20); and item-by-item alignments by test section modality (Tables E.3 – E.6, E.9 – E.12, E.15 – E.18, and E.21 – E.24). Finally, findings from the evaluation of the protocol and process are summarized (see Tables 5 and 6 for strengths and recommendations; Appendix F provides examples of qualitative notes taken during rating in Tables F.1 – F.5).

Overall

A total of 1,381 content and ELD standards at the objective level were analyzed with an overall ratability of 74.0%. Few standards in Mathematics and Science provided enough information to allow for a rating on a language dimension (21.1% and 21.9%, respectively), while most ELD and ELA standards were ratable (96.2% and 95.8%, respectively). A total of 396 CELDT Form E items were analyzed with 100% ratability. These 396 items, 99 per grade span, represented 20 ELD standards in grade K – 2, 25 in grades 3 – 5, 21 in grades 6 – 8, and 19 in grades 9 – 12. In grades K – 2, 11 items were paired with unratable ELD standards; otherwise, all items and ELD standards could be evaluated for alignment.

Linkage

The purpose of linkage analysis was to determine the ways and degree to which ELD standards correspond to the content standards. Linkage results are reported by grade level for each modality (*listening, speaking, reading, writing*) as well as across sampled grade levels (2, 5, 7, and 9). Appendix D provides comprehensive linkage results.

Table 1 summarizes the frequency of linguistic skills codes versus academic language function codes for the two sets of standards. Since any standard could have been assigned one or more language-demand codes and since the number of standards varied across content and grades, it was important to convert raw frequencies into a simple ratio (frequency of language demands in the content standards divided by the frequency of language demands in the ELD standards). The proportion reported in the fourth column of Table 1 reflects the relation between sets of standards. In general, ELD standards represented relatively more linguistic skills (ratio=0.17 – 0.36) than content standards. Academic language functions were most often represented in the content standards across grades and modalities (ratio=0.80 – 1.15).

Table 1 Linkage Analysis Language-Demand Code Frequencies

Grade	ELD Standards	Content Standards	Proportion Content to ELD Standards*
Linguistic Skills**			
2	123	42	0.34
5	138	26	0.19
7	133	23	0.17
9	157	56	0.36
Academic Language Functions***			
2	65	75	1.15
5	127	101	0.80
7	170	139	0.82
9	190	159	0.84

Notes. *(frequency of content standard codes) ÷ (frequency of academic language function codes) **7 possible linguistic skill codes ***14 possible academic language function codes

Table 2 reports the percent ratability of each set of standards and an overall ratability by grade. On the whole, ELD standards were more ratable (92.6% – 97.5%) than content standards (53.5% – 56.9%).

Table 2 Linkage Analysis Percent Ratability of Standards

Grade	ELD Standards (%)	Content Standards (%)	All Standards (%)
2	92.6	56.9	73.7
5	97.5	53.9	76.1
7	97.1	53.5	73.5
9	96.8	55.0	73.1

The relative portion of standards coded for each modality of *listening*, *speaking*, *reading*, and *writing* is reported in Table 3. Because a standard could be coded for one or more modalities, percentages are reported rather than raw frequencies. The analysis of the content and ELD standards revealed similar percents for *listening*; however, roughly two thirds of ELD standards required *speaking* and *writing* while only one third of the content standards required these modalities. *Reading* ranged from 23.0% to 34.4% in the ELD standards while the content standards showed a broader range across the grades (15.3% – 42.9%).

Table 3 Linkage Analysis at Percent Each Modality Rating

Grade	ELD Standards (%)	Content Standards (%)	All Standards (%)
<i>Listening</i>			
2	13.1	13.9	13.5
5	12.5	15.6	14.0
7	7.6	8.5	8.1
9	5.8	6.0	5.9
<i>Speaking</i>			
2	59.8	45.3	52.1
5	65.6	26.6	46.5
7	55.3	30.0	41.6
9	58.2	33.7	44.3
<i>Reading</i>			
2	23.0	20.4	21.6
5	26.9	42.9	34.7
7	32.9	18.0	24.9
9	34.4	15.3	23.5
<i>Writing</i>			
2	49.2	40.1	44.4
5	56.9	27.3	42.4
7	58.2	33.0	44.6
9	58.2	46.2	51.4

Complexity results are summarized in Table 4. Across the grades, the ELD standards were coded at lower complexity levels than the content standards. For example, in Grade 2, over half of the ELD standards were considered *low* complexity while about one third of the content standards were considered *low* complexity. Almost 4% of the Grade 2 content standards were rated as *high* compared to less than 1% of the ELD standards.

Table 4 Linkage Analysis Percent at Each Complexity Rating

Grade	ELD Standards (%)	Content Standards (%)	All Standards (%)
<i>Low</i>			
2	56.6	32.8	44.0
5	53.1	5.2	29.6
7	35.9	12.5	23.2
9	33.9	14.1	22.6
<i>Medium</i>			
2	34.4	16.1	24.7
5	36.9	18.8	28.0
7	46.5	13.5	28.6
9	43.4	19.7	29.9
<i>High</i>			
2	0.8	3.6	2.3
5	6.9	26.0	16.2
7	14.7	25.0	20.3
9	19.6	18.1	18.7
<i>Range (Low-High)*</i>			
2	0.8	4.4	2.7
5	0.6	3.9	2.2
7	0.0	2.5	1.4
9	0.0	3.2	1.8

Note. * see Appendix B, Decision Rules

Crosswalk analyses, described more fully in Appendix C, revealed that Grades 2, 7, and 9 showed the strongest linkages across all language demands at most levels of complexity. Grade 5 showed particular weakness in linguistic skills for the content areas. The *listening* modality was represented less often than *speaking*, *reading*, and *writing* across the grades, although standards at Grades 2 and 5 showed more reliance on *listening* than in Grades 7 and 9. *Listening* also showed the weakest overall linkage across the standards, often mismatched by complexity. The *reading* modality showed a particular weakness in the academic language function of *definition* for both sets of standards. Not surprisingly, *writing* was weakest in the linguistic skills of *phonemes* and *syllables* across all standards (see Appendix D).

Grade 2

Across the Grade 2 standards, over 73% were ratable with almost all ELD standards considered ratable (92.6%). More than 52% of all standards were coded as relevant to the *speaking* modality: ELD standards emphasized *speaking* (59.8%) while a more even distribution across the four modalities was required by the content standards. More than half the ELD standards were considered of *low* in complexity (56.0%) and less than 1% rated high in complexity; in contrast, a third of the content standards were rated *low* (32.8%), 16% *medium*, and 8% either *high* or a range of complexity.

ELD and content standards showed some linkage for linguistic skills across the modalities in Grade 2. Also, both sets of standards included the language demand *identification* across all modalities. Weaker linkage was evident for the other academic language functions: ELD standards showed fewer of these language demands than content standards, and, of these, more were of lower complexity. For example, the code for *labeling and enumeration* was well represented in the content standards but not in the ELD standards.

Grade 5

Similar to Grade 2, about three quarters of the Grade 5 standards were ratable with the ELD standards highly ratable (97.5%). *Speaking* and *writing* modalities were required by 65.6% and 56.9% of the ELD standards, respectively while the four modality codes were more evenly distributed across the content standards. Again, more than half of the ELD standards displayed *low* complexity while most content-area standards were either *medium* or *high* in complexity (44.8%).

Grade 5 showed weakness in linkage. Content area standards represented few linguistic skill language demands while the ELD standards covered many of the linguistic skills. Conversely, content standards covered a number of academic language functions while ELD standards covered few. This pattern reflected varying complexity as well, with greater complexity represented by the content standards than the ELD standards. For example, content standards showed strength in the language-demand code for *labeling and enumeration* while the ELD standards did not. An exception to this pattern was *description* which showed strong linkage.

Grade 7

Again, over 97% of the ELD standards were ratable while only 53.5% of the content standards could be rated (with most unratable standards in Math, 74.6%, and Science, 80.4%). More than half of the ELD standards were relevant to *writing* and *speaking* while the content standards displayed a more even distribution across the modalities. As in the lower grades, the dimension of depth was weakly linked with most ELD standards (82.4% with *low* or *medium* complexity) while, in the content areas, a quarter of the standards were rated as highly complex.

As with Grade 5, Grade 7 showed some weak linkages across the linguistic skills and academic language functions, although in a slightly different pattern. In general, Grade 7 ELD standards better represented linguistic skills than the content standards. At the lower levels of complexity, neither ELD nor content standards covered the academic language functions well. For example, for the *speaking* modality, *labeling and enumeration* as well as *critique and evaluation* were represented well in the content standards but not as well by the ELD standards. However, in the case of *writing*, there was a strong linkage between the sets of standards. *Identification*, for instance, was strongly linked.

Grade 9

As in the other grades, around 73% of all Grade 9 standards were considered ratable; most Math and Science standards were considered unratable (84.4% and 86.8%, respectively). Again, the ELD standards emphasized the modalities of *speaking* and *writing*; the content standards, while showing the same pattern, displayed a more evenly distributed array of modalities.

Results from Grade 9 also revealed the overall trend whereby (a) ELD standards cover linguistic skills better than content standards, and (b) ELD standards have lower levels of complexity than content standards. However, Grade 9 *speaking*, *reading*, and *writing* modalities showed strong linkage for the academic language functions, and *writing* especially showed strong linkage overall and at all complexities.

Alignment

Overall, the CELDT and ELD standards showed varying degrees of alignment across the grade spans and standards. Appendix E provides detailed alignment results including item-by-item results. Presented by grade span, Table 5 reports the number of CELDT items per section, the number of those items that were ratable, and the numbers of items aligned by complexity, modality, and language-demand ratings.

Table 5 Alignment Analysis Item Frequencies by Grade Span

Grade Span	Test Section	N Items in Section	n Ratable Items in Section	n Items Aligned*		
				by complexity	by modality	by language demands
K-2	Listening	20	9**	5	6	6
	Speaking	20	20	17	20	4
	Reading	35	35	25	19	35
	Writing	24	24	10	5	23
3-5	Listening	20	20	7	9	7
	Speaking	20	20	15	20	20
	Reading	35	35	24	24	26
	Writing	24	24	2	7	24
6-8	Listening	20	20	11	9	4
	Speaking	20	20	8	16	9
	Reading	35	35	14	20	16
	Writing	24	24	19	24	3
9-12	Listening	20	20	10	9	6
	Speaking	20	20	10	20	16
	Reading	35	35	16	28	26
	Writing	24	24	10	6	24

Note. * % aligned includes partial to full alignment (see Appendix E for complete percent alignment results). ** Listening standards for items 1 – 11 were rated insufficiently ratable.

Frequencies of overall alignment percentages were computed to evaluate relative strength of alignment and are presented in Table E.25. The mean percent alignment was 55.54%, the median was 55.28%, and the mode was 100.00%. Evaluation of relative alignment could then be defined at three levels, each representing a third of the

distribution: less than 45% alignment, between 45 and 71% alignment, and greater than 71% termed *weak*, *moderate*, and *strong*, respectively. Tables 6 – 9 summarize alignment findings by test section.

Overall, Speaking and Reading items displayed the strongest alignment across the grade spans and dimensions. Items in Listening sections showed the weakest alignment overall, and Writing items showed both strong and weak alignment. Strengths and weaknesses of overall alignment are summarized in Table 14.

Listening

Listening items were the most weakly aligned of the four test sections (20.0% – 66.7%). Specifically, all grade spans showed moderate alignment for modality, and all but grades 3 – 5 (35.0%) were moderately aligned for complexity. However, only Grades K-2 were moderately aligned for language demands (66.7%).

Table 6 Alignment Analysis of CELDT Listening

Grade Span	Complexity	Modality	Language Demands
K – 2	55.6	66.7	66.7
3 – 5	35.0	45.0	35.0
6 – 8	55.0	45.0	20.0
9 – 12	50.0	45.0	30.0

Note. n items = 20

Speaking

Speaking items were strongly aligned by modality in all Grades (80.0% – 100.0%), and two out of the four grade spans were strongly aligned in terms of complexity and language demands. Grades 6 – 8 language demands (45.0%) and Grades 9 – 12 complexity (50.0%) were moderately aligned, while Grades 6 – 8 complexity (40.0%) and K – 2 language demands (20.0%) were weakly aligned.

Table 7 Alignment Analysis of CELDT Speaking

Grade Span	Complexity	Modality	Language Demands
K – 2	85.0	100.0	20.0
3 – 5	75.0	100.0	100.0
6 – 8	40.0	80.0	45.0
9 – 12	50.0	100.0	80.0

Note. *n* items = 20

Reading

Reading items were strongly aligned by language demands for all grades (74.3% – 100.0%), except Grades 6 – 8 which had moderate alignment (45.7%). All grade spans showed at least moderate alignment by complexity and modality, except for Grades 6 – 8 complexity (40.0%).

Table 8 Alignment Analysis of CELDT Reading

Grade Span	Complexity	Modality	Language Demands
2*	71.4	54.3	100.0
3 – 5	68.6	68.6	74.3
6 – 8	40.0	57.1	45.7
9 – 12	45.7	80.0	74.3

Note. *n* items = 35 *Kindergarten and Grade 1 are not assessed in Reading or Writing on the CELDT.

Writing

Items in the CELDT Writing sections were strongly aligned by language demand type for all grades (95.8 – 100.0%) except Grades 6 – 8 (12.5%). Grades 6 – 8 showed strong alignment on complexity (79.2%) and modality (100.0%). However, Writing items were weakly aligned by complexity and modality in the other grade spans (8.3% – 41.7%).

Table 9 Alignment Analysis of CELDT Writing

Grade Span	Complexity	Modality	Language Demands
2*	41.7	20.8	95.8
3 – 5	8.3	29.2	100.0
6 – 8	79.2	100.0	12.5
9 – 12	41.7	25.0	100.0

Note. n items = 20 *Kindergarten and Grade 1 are not assessed in Reading or Writing on the CELDT.

Proficiency Levels

The Tables 10 – 13 report the modality and complexity ratings for all CELDT items by proficiency level for each grade span. For more detail on proficiency level analysis, including language demand ratings by proficiency level (see Appendix E Tables E.2, E.8, E.14, and E.20). Overall, most CELDT items were written to *beginning* proficiency level standards and were of *low* complexity. For example, in the K – 2 grade span, over a third of the items were based on the *beginning* proficiency level while the other proficiency levels were represented between 8.1 and 21.2% of the time. In 6 – 8 and 9 – 12 grade spans, only one item represented *advanced* proficiency on each test.

Table 10 Percent Items by Proficiency Level Coded by Modality & Complexity: K – 2

	N	Proficiency Level*				
		B	EI	I	EA	A
<i>Sufficient</i>	99**	36.4	8.1	21.2	15.2	19.2
Modality						
<i>Listening</i>	42	69.0	4.8	16.7	9.5	0.0
<i>Speaking</i>	20	85.0	10.0	0.0	5.0	0.0
<i>Reading</i>	54	14.8	3.7	25.9	20.4	35.2
<i>Writing</i>	5	0.0	80.0	20.0	0.0	0.0
Complexity						
<i>Low</i>	71	42.3	11.3	15.5	11.3	19.7
<i>Medium</i>	27	18.5	0.0	37.0	25.9	18.5
<i>High</i>	1	100.0	0.0	0.0	0.0	0.0
<i>Low-High</i>	0	0.0	0.0	0.0	0.0	0.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced; **100%

Table 11 Percent Items by Proficiency Level Coded by Modality & Complexity: 3 – 5

	N	Proficiency Level*				
		B	EI	I	EA	A
<i>Sufficient</i>	99**	19.2	29.3	15.2	15.2	21.2
Modality						
<i>Listening</i>	40	32.5	42.5	22.5	2.5	0.0
<i>Speaking</i>	20	65.0	30.0	0.0	5.0	0.0
<i>Reading</i>	54	11.1	14.8	11.1	25.9	37.0
<i>Writing</i>	5	0.0	80.0	0.0	0.0	20.0
Complexity						
<i>Low</i>	74	24.3	27.0	14.9	6.8	27.0
<i>Medium</i>	22	4.5	40.9	13.6	40.9	0.0
<i>High</i>	2	0.0	0.0	50.0	50.0	0.0
<i>Low-High</i>	1	0.0	0.0	0.0	0.0	100.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced; **100%

Table 12 Percent Items by Proficiency Level Coded by Modality & Complexity: 6 – 8

	N	Proficiency Level*				
		B	EI	I	EA	A
<i>Sufficient</i>	99**	20.2	19.2	31.3	28.3	1.0
Modality						
<i>Listening</i>	40	35.0	42.5	22.5	0.0	0.0
<i>Speaking</i>	20	55.0	30.0	15.0	0.0	0.0
<i>Reading</i>	61	9.8	11.5	31.1	45.9	1.6
<i>Writing</i>	24	0.0	0.0	20.8	79.2	0.0
Complexity						
<i>Low</i>	72	19.4	15.3	27.8	37.5	0.0
<i>Medium</i>	20	25.0	40.0	30.0	5.0	0.0
<i>High</i>	1	0.0	0.0	0.0	0.0	100.0
<i>Low-High</i>	6	16.7	0.0	83.3	0.0	0.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced; **100%

Table 13 Percent Items by Proficiency Level Coded by Modality & Complexity: 9 – 12

	N	Proficiency Level*				
		B	EI	I	EA	A
<i>Sufficient</i>	99**	24.2	35.4	19.2	20.2	1.0
Modality						
<i>Listening</i>	40	32.5	45.0	22.5	0.0	0.0
<i>Speaking</i>	20	65.0	35.0	0.0	0.0	0.0
<i>Reading</i>	64	17.2	20.3	29.7	31.3	1.6
<i>Writing</i>	5	0.0	100.0	0.0	0.0	0.0
Complexity						
<i>Low</i>	30	33.3	33.3	10.0	23.3	0.0
<i>Medium</i>	57	21.1	36.8	19.3	21.1	1.8
<i>High</i>	12	16.7	33.3	41.7	8.3	0.0
<i>Low-High</i>	0	0.0	0.0	0.0	0.0	0.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced; **100%

Table 14 Strengths and Weaknesses of CELDT-to-ELD Standards Alignment

Strengths
<ul style="list-style-type: none"> • Reading sections generally showed moderate to strong alignment overall, with the exception of Grades 6 – 8 complexity. • Speaking sections showed moderate to strong alignment in all grades except K – 2 language demands and 6 – 8 complexity. • Writing was strongly aligned for language demands in all grades except 6 – 8.
Weaknesses
<ul style="list-style-type: none"> • Listening sections overall showed weak alignment for complexity, modality, and language demands. • Complexity and modality alignment was weak for Writing sections except in Grades 6 – 8. • Grades 6 – 8 were the least aligned across all dimensions.

Protocol Evaluation

The language demand protocol used here showed promise for future evaluation of assessment linkage and alignment. Among other strengths (see Table 15), raters in the study agreed that language demands are a meaningful focus for analyzing curricular and assessment materials. Training in the concepts and decision-making associated with the protocol positively challenged assumptions raters held about test constructs and the process of learning English for EL students.

Table 15 Strengths of the Linkage/Alignment Protocol

Strengths
<ul style="list-style-type: none"> • Educators thought that the protocol was a useful tool for analyzing students’ performance, skills, and types of skills. • Training in rating language demands deepened educators’ understanding of test constructs and relationships between items and standards. • Training in rating language demands challenged some of the educators’ assumptions about the process of learning English for ELs. • Educators came to appreciate the distinction between item difficulty and language demands. • The protocol may provide more information for reporting student performance in the long run.

Table 16 displays recommendations for the further development and use of the language demand protocol. For example, raters recommended further development of operational definitions to clarify distinctions between and within codes. Also, the time-consuming nature of the process, especially the process of coming to consensus, prompted the question of how to improve efficiency. Steps, such as cross-checking rating sheets to first find discrepancies and then discuss only those discrepancies identified, reduced the time needed to reach consensus. Group calibration procedures could be developed to meet this aim. Finally, raters thought decision rules should be further clarified and made more comprehensible if the protocol is to be used routinely.

Table 16 Recommendations for Improvement of the Linkage/Alignment Protocol

Recommendations
<ul style="list-style-type: none"> • Protocol and decision-making were very time-consuming. Steps to reduce time should be developed (e.g., cross-checking rating sheets for discrepancies to reduce discussion time needed for consensus). • More attention is needed to clarify distinctions between language demands; sometimes definitions appear to overlap. Further clarification as to when such overlapping definitions are appropriate would assist raters. • Distinctions between modalities did not always appear operationally meaningful. Raters suggested a new code scheme for instances when modalities are interrelated. • Test section and language demand modalities were sometimes mismatched. Similar to the fourth finding from the pilot study (Sato, et al., 2005), the CELDT section title did not always match a specific item’s modality. Raters recommend a code for such items. • Suggested new language demand codes or expanded language demand definitions included the following: <ul style="list-style-type: none"> – Graphemes; fine-motor skill requirements of Writing – Development of a thesis statement (controlling idea or theme of a piece of Writing or a presentation) – Speaking conventions (those used in formal speechmaking or discourse) – Content-area vocabulary – Inclusion of all elements of the ELD Writing Convention standards in WRICON (e.g., parts of speech) • Complexity should be made more concrete by identifying "elements" within an item or standard such as the following: <ul style="list-style-type: none"> – Number of modalities (<i>listening, speaking, reading, writing</i>) – Number of design elements of the item (pictures, words, text, other visuals, etc.) – Number of distinctions between two or more things – Number of references to different parts of the item (e.g., use the <i>picture</i> and the information from the <i>passage</i> to <i>write</i> your answer) – Number of interactions between elements – Number of elements that do not apply to responding to the item • Training materials should be codified before broader use of the protocol.

Discussion

The findings confirm many of the results of the pilot study (Sato, et al., 2005). Linkage was inconsistent across groups of standards. As in the pilot, most ELD and ELA standards were ratable (greater than 90%); however, only 13% to 38% of the Math and Science standards were ratable using both the standards and accompanying frameworks documents.

The ELD standards usually represented linguistic skills (e.g., *phonemes, syllables, writing conventions*) while content standards represented relatively more of the academic language functions (e.g., *labeling and enumeration, explanation, analysis and synthesis*) across the grades. The ELD standards and content standards were often weakly linked by complexity of the language demands: ELD standards were generally of lower complexity than content standards.

This study found that the lower proficiency levels (*beginning* and *early intermediate*), were favored by the CELDT. For example in the K – 2 grade span, over a third of the items were based on the *beginning* proficiency level while the other proficiency levels were represented between 8.1 and 21.2% of the time. In 6 – 8 and 9 – 12 grade spans, only one item represented *advanced* proficiency on each test.

Finally, some test sections included items that required different language demand modalities. For instance, many items in the Writing section were rated with the *reading* modality. In such cases, findings here suggest the need to reconsider the test construct if the CELDT is to align to ELD standards on language-demand dimensions.

Raters cited a concern about consistency and accuracy of coding across groups of raters. For example, raters for different grade levels could have focused on different language demands for related standards. Multiple efforts were made to optimize consistency and accuracy. First, groups were trained, using the pilot study materials and data. At least one group member in each rating group had experience in another group. Decision rules were debated, recorded, and communicated to new group members (see Appendix B for decision rules). Finally, the consensus-building process required deliberation and defense of individual views.

Findings & Implications

As in the pilot study, we address several questions that had been developed by Fast and Hebbler (2004, p. 33) with CCSSO (see below) to guide states in the evaluation of alignment between content tests and content standards. Although the context of the current work was focused on identifying language demands in standards rather than on content such as Math, Science, etc., the questions still helped to guide the work. Specifically, the questions are used here as a basis for summarizing our findings and their implications. This approach helps provide a link between traditional alignment in the content areas and alignment in the area of language proficiency.

1. a) *Are the assessments and the standards aligned comprehensively, meaning that the assessments reflect the full range and depth of the standards?*

The study revealed variable degrees of alignment between ELD standards and the CELDT by language-demand dimensions and proficiency levels. Strongest alignment appeared in Reading and Writing items by language demands ratings. Reading and Speaking items showed mostly moderate to strong alignment across sections of the test for all dimensions, especially in the lower two grade spans. Writing items were variably aligned while Listening items showed the weakest alignment for all dimensions.

The 6 – 8 grade span showed the poorest alignment of the four spans, especially on the dimensions of depth and breadth. A possible explanation for these differences may be that, for this grade span, the *low*-complexity CELDT items were weighted more toward ELD standards at the *early advanced* proficiency level (see Table 12) as compared to other grades where lower complexity matched the lower proficiency levels.

By CELDT section, the weakest alignment was in Listening across the grades. A reason may be the structure of the ELD standards as compared with the CELDT. The ELD standards are written such that Listening and Speaking standards are interrelated and without distinction between the modalities. However, the CELDT breaks Listening and Speaking into two sections. Listening presents unique item-writing challenges given that the student's response is an indirect measure of listening skill.

Writing items were either strongly aligned (greater than 71%) or weakly aligned (less than 45%) on the three dimensions, with stronger alignment in the language

demands ratings and weaker alignment in complexity and modality. Given that items in the Writing section usually require *reading*, it follows that this group of items would not be aligned by modality. As for complexity, weak alignment could be related to the use of constructed-response items scored with a multiple-point rubric. Since these items were rated for complexity at the highest possible score point, these ratings may appear mismatched. Grades 6 – 8 presented a unique pattern compared with the other spans. These items were strongly aligned with respect to complexity and modality, but weakly aligned with respect to language demands. This finding suggests inconsistent item ratings, given that the ELD standards were rated by the same group across the spans.

b) Are the assessments as cognitively challenging as the standards?

The CELDT and the ELD standards both reflect overall *low* complexity in comparison to the DOK required by the content standards.

Implication: If the correspondence between California ELD standards and content standards is strengthened on all language dimensions, the CELDT could be better aligned to those standards. In the short term, the development of new CELDT items and form specifications could improve overall alignment of the system. CELDT items could be developed to reflect more complex academic language functions. Also, the test could move toward a more comprehensive range of proficiency levels at each grade span by increasing the number of ELD standards represented on the test while staying within the requirements of the blueprint. High-frequency words taken from the content areas could be integrated to increase word- and sentence-level academic language functions. However, the CELDT may be limited in its ability to assess discourse-level academic language functions, such as *analysis* and *argumentation*, given the present test administration parameters. Alternatively, these functions might be best integrated into formative, classroom assessments.

2. Are the assessments and the standards aligned in terms of content and process, meaning that the assessments measure what the standards state students should both know and be able to do?

Results show that the CELDT is somewhat aligned to the ELD standards with regard to language demands. Items in the Listening and Writing sections showed weaker

alignment than in Speaking and Reading. To improve alignment, the CELDT Listening and Writing sections could be further evaluated and new items written to increase complexity. The CELDT's alignment to content standards could be improved with greater emphasis on academic language functions. The ELD standards include and the CELDT now assesses only a subset of the content and process skills required by the content standards.

Implication: Item types that incorporate process skills and academic language functions could enhance the CELDT. High-frequency, content-specific vocabulary could be used in items.

3. Do the assessments reflect the same degree and pattern of emphasis on the content as are reflected in the state's academic content standards?

State content standards reflect more academic language functions and higher complexity when compared to the ELD standards and the CELDT. However, evidence of some correspondence between content standards and ELD standards exists. Across the grades, word- and sentence-level academic language functions appear in both the ELD and content standards. For example, *identification* was represented across all grades in both sets of standards. Other word- and sentence-level language demands, such as *labeling and enumeration* and *description*, were common in the content standards and could offer item developers direction when writing items that integrate the academic language functions.

Implication: The CELDT could be developed with more emphasis on academic language functions by identifying those specific functions that can be assessed in a standard fashion with valid and reliable results. These specific language demands (e.g., *identification, labeling and enumeration, description*) could provide a place to start in developing test items that align more strongly to content-area expectations for students.

4. Do the assessments yield scores that reflect the full range of achievement implied by academic achievement standards?

Considering the lack of correspondence in the complexity of the content area standards as compared to the ELD standards and the CELDT, it is apparent that scores on the CELDT do not reflect the range of achievement expected for all students.

Implication: The results suggest that item development focus on the dimension of complexity in future forms of the CELDT. An emphasis on developing depth of knowledge is relevant if academic language functions are to be assessed. Test developers should add complexity to the test blueprint.

The misalignment on the dimension of complexity has a consequence for the interpretation of CELDT results as a measure of proficiency for the purpose of reclassification or initial identification of English learners as fluent. To the extent that students are prematurely reclassified or initially identified as fluent, they will not receive the support they need to succeed in school and beyond.

5. Do the assessments measure the knowledge and skills described in [the] academic content standards and not knowledge, skills, or other characteristics that are not specified in the academic content standards?

The results of the analyses of language demands required by the content standards compared with the ELD standards and the CELDT suggest that the test measures some of the relevant knowledge and skills, though not all.

Implication: The analyses of linkage indicate that each content area has standards with unique traditions, lexicons, expectations, and assumptions. An evaluation of these standards in relation to the ELD standards reveals a need for greater linkage. Students learning the English language while they develop their knowledge of Math, Science, and Language Arts present unique learning needs, depending on their level of linguistic and cognitive development. Given that certain content areas will be more or less accessible to students depending on their level of language proficiency and cognitive abilities, ELA standards that appear only at the lower grades would be relevant for older EL students as they tackle higher levels of Math or Science. In such instances, it may be appropriate that skills and knowledge *not* specified in the content standards appear on the CELDT.

6. How are gaps and weaknesses identified, and what is done to improve the alignment of its standards and assessments?

The study identified gaps in the linkage and alignment of the CELDT and the ELD standards.

Implication: The linkage and alignment protocol used for this study promises to be a useful tool both for item development and for development of curricular materials. Further refinements could improve the efficiency and consistency of the procedures (see Table 16 for specific recommendations). The approach taken here will aid test developers and state assessment decision makers by providing an opportunity to refine the processes that guide item development in order to monitor and reduce gaps in linkage and alignment.

Conclusion

The work reported here has allowed for a serious examination of linkage and alignment relationships between the California ELD, content standards, and the state's ELD assessment. (See Figure 1 for a visual representation of the various relationships involved in the assessment system which includes Title III and Title I.) The findings show that while there are areas of close correspondence on the dimension of language, important gaps do exist across grade spans, content areas, language modalities, and language proficiency levels. As mentioned above, to ensure that students have the necessary language skills in English to handle material across content areas, closer linkage and alignment are desirable.

In the short term, the development of CELDT items that reflect more academic language functions and higher levels of complexity would improve overall alignment of the system. The CELDT item development could target the following for the purpose of strengthening alignment: (a) items that incorporate academic language functions in the form of high-frequency, nonspecific, academic vocabulary; (b) items that address specific language demands at the word and sentence level (e.g., *identification, labeling and enumeration, description*); and (c) items that reflect greater depth (i.e., complexity). Given that the ELD standards and the CELDT tend toward lower complexity ratings, an emphasis on higher complexity in item development could make the test more discriminating at the higher proficiency levels. Academic language functions at the discourse level may not lend themselves to guiding item development for the standardized testing administration of the CELDT. Instead, classroom and formative assessments may be more suitable for assessing such language functions.

Findings here indicate that each content area has produced standards with unique traditions, lexicons, expectations, and assumptions. Further work to understand these unique aspects of content standards and how they are related on the language-demand dimensions could help further the goals of aligning the Title III and Title I systems (Figure 1).

Finally, the methodology used for this study promises to be a useful tool for future evaluations of linkage and alignment. Other benefits of the protocol and language-

demand definitions could include (a) assisting educators in the development of curricular materials and (b) test developers in the design of well-aligned items. Future evaluation studies could continue to define the parameters necessary for bringing all aspects of the assessment program into alignment and help in the interpretation and practice of standards-based curriculum and accountability.

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., & Butler, F. (2004) *Linkage and alignment protocol*. Unpublished manuscript.
- 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*. (CSE Tech. 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., Butler, F., & Sato, E. (in press a). Standards-to-standards linkage under Title III: Exploring common language demands in ELD and Science standards. *Applied Measurement of Education*. (CSE Tech Report 667). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing: http://cresst.org/products/reports_set.htm
- Bailey, A. L., Butler, F. A., LaFrumenta, C., & Ong, C. (2001/2004). *Towards the characterization of academic language in upper elementary science classrooms*. (OERI Contract No. R305B960002) (CSE Tech. Rep. No. 621). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- 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*. (CSE Tech. Report 664). Los Angeles: University of California, National Center for

Research on Evaluation, Standards, and Student Testing. Retrieved May 15, 2006, from http://cresst.org/products/reports_set.htm

- Bailey A. L., Butler, F. A., Stevens, R., & Lord, C. (in press b). 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.
- Bibian, G. (2006). *An ethnographic case study of a reconstituted urban middle school and the factors that contribute to improved ELL redesignation rates*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Butler, F. A., Bailey A. L., Stevens, R., Huang, B. & Lord, C. (2004). *Academic English in fifth-grade mathematics, science, and social studies textbooks*. (IES, Contract No. R305B960002). (CSE Tech. Rep. No. 642). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Butler, F.A., Lord, C., Stevens, R., Borrego, M., & Bailey, A.L. (2003/2004). *An approach to operationalizing academic language for language test development purposes: Evidence from fifth-grade science and math* (IES, Contract No. R305B960002). (CSE Tech. Rep. No. 626). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- California Department of Education. (2005). *Mathematics framework for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/ci/ma/cf/index.asp>
- California Department of Education. (2003). *Science framework for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/re/pn/fd/sci-frame-dwnld.asp> [Note, the preceding Web address is no longer valid.]

- California Department of Education. (2002). *English language development standards for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/re/pn/fd/documents/englangdev-stnd.pdf>
- California Department of Education. (1999). *Reading language arts draft framework for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/ci/rl/cf/index.asp>
- California Department of Education. (1998a). *Mathematics academic content standards for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/be/st/ss/mthmain.asp> [Note, the preceding Web address is no longer valid.]
- California Department of Education. (1998b). *Science academic content standards for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/re/pn/fd/documents/sci-stnd.pdf>
- California Department of Education. (1997). *English language arts academic content standards for California public schools: Kindergarten through grade twelve*. Sacramento: California Department of Education. Retrieved October 1, 2005, from <http://www.cde.ca.gov/be/st/ss/engmain.asp> [Note, the preceding Web address is no longer valid.]
- CTB/McGraw-Hill. (2005). *California English Language Development Test. Form E*. Monterey, CA: Author.
- Cunningham, J. W., & Moore, D. W. (1993). The contribution of understanding academic vocabulary to answering comprehension questions. *Journal of Reading Behaviors*, 25, 171–80.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.

- 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 of Education*.
- Fast, E. F. & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC. Council of Chief State School Officers.
- 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: A Case Study. *Applied Measurement of Education*.
- Larson-Freeman, D. (2000). Second language acquisition and applied linguistics. *Annual Review of Applied Linguistics*, 20, 165–181.
- No Child Left Behind (2001a). *No Child Left Behind, Title III: Language instruction for limited English proficient and immigrant students*. 107th Congress, 1st session, December 13, 2001. (Printed version prepared by the National Clearinghouse for Bilingual Education). Washington, DC: George Washington University, National Clearinghouse for Bilingual Education.
- No Child Left Behind (2001b). *No Child left Behind, Title I: Improving the academic achievement of the disadvantaged*. 107th Congress, 1st Session, December 13, 2001. (Printed version prepared by the National Clearinghouse for Bilingual Education, Washington, DC: George Washington University.
- 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., Smithson, J. Blank, R., & Zeidner, T. (in press). Alignment as a teacher variable. *Applied Measurement of 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, NY: Teachers College.
- Sato, E., Lagunoff, R., Worth, P. Bailey A.L., & Butler, F. A. (2005). ELD standards linkage and test alignment under Title III: A pilot study of the CELDT and the California ELD and content standards. Final report (June) to the California Department of Education, Sacramento, CA.
- Scarcella, R. & Zimmerman, C. (1998). Academic words and gender: ESL student performance on a test of academic lexicon. *Studies in second language acquisition*, 20(1), 27-49.
- Schleppegrell, M. J. (2004). *Language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Spanos, G., Rhodes, N. C., Dale, T. C., and Crandall, J. (1988). Linguistic features of mathematical problem solving: Insights and applications. In R. R. Cocking and J. P. Mestre (Eds.), *Linguistic and cultural influences on mathematics* (pp. 221–240). Hillsdale, NJ: Erlbaum.
- Stevens, R., Butler, F.A., and Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of ELLs* (CSE Tech. Rep. No. 552). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing. Retrieved May 15, 2006, from, http://cresst.org/products/reports_set.htm
- 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 to assessments. *Applied Measurement of Education*.

Webb, N. L., Horton, M., & O'Neal, S. (2002). *An analysis of the alignment between language arts standards and assessments for four states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

WestEd. (2003). *Alignment Study: Louisiana Content Standards and the Graduation Exit Examination for the 21st Century (GEE 21)*. San Francisco: Author.

WestEd. (2004a). *Alignment Study: Louisiana Content Standards and the Louisiana Educational Assessment Program (LEAP 21)*. San Francisco: Author.

WestEd. (2004b). *A Comparison of New England Compact Grade-Level Expectancies to NAEP Content, NCTM/NCTE, and Vermont Content Standards*. San Francisco: Author.

WestEd. (2004c). *Alignment Study: Louisiana Content Standards and the Louisiana Alternate Assessment (LAA)*. San Francisco: Author.

WestEd. (2003). *Alignment Study: Louisiana Content Standards and the Louisiana Alternate Assessment (LAA)*. San Francisco: Author.

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

Wong-Fillmore, L., & Snow, C. (2000). *What teachers need to know about language*.
ERIC Clearinghouse on Languages and Linguistics. Retrieved August 1, 2000,
from <http://www.cal.org/ericcll> [Note, the preceding Web address is no longer
valid.]

Appendices

Appendix A

Linkage & Alignment Protocol

The linkage and alignment protocol was taken from the pilot study conducted by Sato, et al.(2005) with some differences, as reflected in the decision rules, throughout ratings and analyses (see Appendix B for additional, specific decision rules developed in the current study). While the pilot researchers supplemented data with classroom materials such as lesson plans, in instances when standards did not meaningfully reflect language demands, the scope of the current study precluded this step.

Data for the protocol was collected using the following process from Sato, et al. (2005, pp.89 – 93):

1. Determine ratability of standard: Rate each standard as either *sufficient* (i.e., providing sufficient detail about potential language demands) or *indeterminate* (i.e., insufficient detail to extract language demands from content standard).
2. Identify language demands: Identify the language demand(s) reflected in the standard and code... Note whether the language demand is categorized as linguistic skill or academic language function [using *Language Demands and Definitions*, Figure A.1].
3. Determine level of complexity: Using the *Levels of Complexity* document, rate the complexity of the language demand addressed in the standard.
4. Record additional notes: If applicable, record in the “Notes” column whether linguistic demands are not grade/age appropriate, or demonstrate a lack of equity or fairness of language (p. 73 – 74).

Figure A.1 Language Demands and Definitions

Appendix C: Language Demand and Definitions

Linguistic Skills

Related Language Modalities: Listening (L); Speaking (S); Reading (R); Writing (W)

Language Demand	Code	Definition/Application	Notes
<i>The item or standard requires students to show knowledge of:</i>		<i>The ability to:</i>	
Phonemes	PHONEM	identify, manipulate, and produce the individual sounds that make up spoken words. (L, S)	phonology
Syllables	SYLLAB	identify the division of words into the smallest units of sequential speech sounds, composed of a vowel sound or a vowel-consonant combination. (L, R)	phonology
Morphemes	MORPHE	identify and distinguish the smallest unit of meaningful sound in language (i.e., words, roots, or affixes). (L, R)	morphology
Vocabulary Words	VOCWOR	identify and determine meaning of spoken or written words or short phrases in context; produce spoken or written words relevant to a particular context. (L, S, R, W)	lexicon
Phrases & Sentences	PHRSEN	determine meaning of spoken and written phrases and sentences; generate original phrases and sentences using grammatical forms. (L, S, R, W)	syntax
Sound-Symbol Correspondences	SOUSYM	identify the relationship between letters of written language (graphemes) and the individual sounds (phonemes). (R, W)	orthography
Written English Conventions	WRICON	recognize and apply written English conventions [i.e., punctuation, capitalization, spelling, paragraph structure, format (including text features)]. (R, W)	writing conventions

Academic Language Functions

Language Demand	Code	Definition/Application	Notes
The item or standard requires students to show ability to use:		The ability to:	
<i>Minimum language demand: word-level</i>			
Identification	IDENTI	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	LABENU	produce the term corresponding to a given definition.	In labeling a picture, the picture may be interpreted as a definition.
Enumeration		name things separately, one by one.	
Classifying	CLASEQ	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).	
<i>Minimum language demand: phrase or sentence-level</i>			
Definition	DEFINI	say what the meaning of something, especially a word, is.	
Interpretation	INTERP	determine and/or demonstrate understanding of the intended meaning of something, as distinct from the literal meaning.	
Comparison/Contrast	COMCON	examine or look for differences and/or similarities between two or more things.	
Explanation	EXPLAN	offer reasons for or a cause.	Includes supporting details. Used to code standards requiring expository writing.
Description	DESCRI	say or write what someone or something is like.	Used to code standards requiring narrative writing.
Inquiry	INQUIR	seek information by forming questions.	
Prediction	PREGEN	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.	

continued

Language Demand	Code	Definition/Application	Notes
The item or standard requires students to show ability to use:		The ability to:	
Inference Hypothesis	PREGEN (continued)	reason from circumstance; surmise. form an idea or explanation for something that is based on known facts but has not yet been proved.	
Retelling Summary	RETSUM	relate or tell again, possibly in a different form. express the most important facts or ideas about something or someone in a short and clear form.	Includes restating in own words.
Minimum language demand: discourse-level			
Analysis Synthesis	ANASYN	identify the parts of a whole and their relationship to one another. identify the relationships among two or more ideas or other textual elements.	
Argument Negotiation	ARGNEG	discuss a point of view with the purpose of creating agreement around a position or conviction. engage in a discussion with the point of creating mutual agreement from two or more different views.	
Persuasion Critique	CRIEVA	convince others of something. 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.	

Note. from Sato, et al. (2005) p. 94

Figure A.2 Levels of Complexity for Rating Language Demands

Appendix D: Levels of Complexity

CELDT Linkage/Alignment Pilot

Levels of Complexity for Rating Language Demands

Complexity Levels	Low	Medium	High
Complexity rating is determined at the highest level of each skill required in the standard or item. Higher levels subsume lower levels.	<i>Language Demand</i> (in text and/or response) <ul style="list-style-type: none"> requires recognition, identification, comprehension, or classification of information requires use of simple or familiar language requires simple construction, level of detail, and depth of content to present explicit information 	<i>Language Demand</i> (in text and/or response) <ul style="list-style-type: none"> requires interpretation, application, or simple analysis of information requires use of increasingly complex and varied language requires increasingly complex construction, level of detail, and depth of content to present primarily explicit information 	<i>Language Demand</i> (in text and/or response) <ul style="list-style-type: none"> requires more complex analysis, synthesis, or evaluation of information, including problem solving tasks requires use of predominantly complex and varied language requires predominantly complex construction, level of detail, and depth of content to present both explicit and implicit information
Examples			
Language Demand: Vocabulary Words	recognize and comprehend sight words in isolation or with basic contextual clues (e.g., pictures or simple sentences)	read and comprehend complex vocabulary in isolation or with increasingly complex contextual clues (e.g., detailed sentences or textual features)	determine the meaning of unfamiliar vocabulary in isolation or with more complex contextual clues (e.g., sentence)
Language Demand: Phrases and Sentences	understand basic sentences and tenses at grade-level complexity	understand intermediate-level sentences and tenses at grade-level complexity	understand advanced sentences and tenses at grade-level complexity
Language Demand: Retelling and Summarizing	retell in own words the main idea and supporting details of a text	retell with increasing complexity the main ideas and supporting details of a text	summarize with more complexity, including analysis of main ideas and supporting details

NOTE: If complexity level is indeterminate based on the wording of the standard, code as "I."

Note. from Sato, et al. (2005) p. 94

Figure A.3 Linkage Rating Sheet

Rater:			Date:			
Content Area:						
Content Standard Code	Standard	Ratability (S=sufficient; I=interdeterminate)	Language Demands Covered	Language Modality (L,S,R,W)	Complexity of language demand	Notes (e.g., age/grade appropriateness, equity, and fairness, or emphasis)
(e.g., WA10)	(e.g., 1.0 Word Analysis, Fluency, and Systematic Vocabulary Development)					

Note. from Bailey, A., & Butler, F. (2004)

Figure A.4 Alignment Rating Sheet

Rater:			Date:					
Grade Span:								
CELDT Page #	CELDT Section	CELDT Item Booklet #	ELD Standard	Ratability (S = sufficient; I = interdeterminate)	Language Demands Covered	Language Modality (L,S,R,W)	Complexity of language demand	Notes (e.g., age/grade appropriateness, equity, and fairness, or emphasis)
(e.g., 1)	(e.g., Listening)	(e.g., 1)	(e.g., 1.04.10 Restate and execute multi-step oral directions)					

Note. from Bailey, A., & Butler, F. (2004)

Appendix B

Decision Rules Developed

During individual and consensus ratings of CELDT items, ELD standards, and content standards, the following decision rules were used to ensure consistent, accurate data:

1. **Conservative Interpretation:** A *conservative interpretation* of language demands will be used. In other words, an item should be rated *only* if there is specific evidence of a language demand in the text of the item or item’s examiner’s instruction.
2. **Implied Language Demands:** If no specific evidence of a language demand exists and participants agree that a language demand is *implied*, they should note this in the “Notes” column.
3. **Complexity Ranges:** Complexity may be rated as a single code (e.g., “M”) or as a range (e.g., “L–M”). A range should only be used when raters can give examples of the item’s demands both at the high and at the low ends of that range.
4. **Consensus Goals:** Consensus can be considered achieved if at least 100% of table’s participants agree on a specific decision. If less than 100% of the group agree on a decision within the time allotted per item, the item is coded “NOCON.”
5. **Word Interpretations:** If the participants decide upon a specific interpretation of a given word in an item or examiner’s instruction, they should note this in the “Notes” column (e.g., “‘tell time’ is interpreted here to mean orally report the time”). This is important when two or more people have different interpretations of the same word.
6. **Sample Items:** No sample items will be coded.
7. **Other Notes:** Notes should be taken in the “Notes” column when

- a. no language-demand code can be applied to an item when raters believe a language demand exists;
 - b. raters believe there is overlapping meaning between two or more codes;
 - c. raters have any other relevant comments or questions.
8. **Regarding Complexity:** Evaluate item in terms of the number of elements. Elements *relevant to the item's language demands* could be
- a. number of modalities (*listening, speaking, reading, and/or writing*)
 - b. number of design elements of the item (pictures, words, text, other visuals, etc.);
 - c. number of distinctions between two or more things;
 - d. number of references to different parts of the item (e.g., use the *picture* and the information from the *passage* to *write* your answer);
 - e. number of interactions between elements ;
 - f. number of irrelevant elements (elements that don't apply to responding to the item).
9. **Remember:** Language demands are different than item difficulty. Language demands are not dependent on the child; rather, they are dependent on what language the item (or standard) demands of the child.
10. **Language-Demand Definitions:** *Please* note the item if you think the code does not fit in the appropriate category (Linguistic vs. Academic).

Appendix C

Crosswalk Analysis Procedure

Using data gathered from rating the ELD and content standards, standards' codes were entered into Crosswalk Tables (see Figure C.1) for each modality. The Crosswalk Tables cross-tabulate the standards with language demands. For example, a standard number coded as showing evidence of a particular language demand (linguistic skill or academic language function) is entered into the table in the cell that reflect that language demand and the standards complexity. Once crosswalk tables were created by grade span for each modality, crosswalk tables were compiled to examine patterns across grade spans. An example image of a Crosswalk Table is depicted in Figure C.2.

Crosswalk Tables

Figure C.1 Crosswalk Table Design

Language Demand (Linguistic Skills)	Complexity	ELD Listening & Speaking	ELD Reading	ELD Writing	English Language Arts	Math	Science
<i>Phonemes</i>	<i>Low</i>	<i>e.g., 212, 214</i>			<i>e.g., WA1</i>		
	<i>Medium</i>						
	<i>High*</i>						
<i>Syllables</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Morphemes</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Vocabulary Words</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Phrases & Sentences</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Sound-Symbol Correspondences</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Written English Conventions</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
	<i>Low-High</i>						

continued

Language Demand (Academic Language Functions)	Complexity	ELD Listening & Speaking	ELD Reading	ELD Writing	English Language Arts	Math	Science
<i>Identification</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Labeling & Enumeration</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Classification & Sequencing</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Definition</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Interpretation</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Comparison/ Contrast</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Explanation</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Description</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Inquiry</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Prediction, Generalization, Inference, & Hypothesis</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Retelling & Summary</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						

continued

<i>Analysis & Synthesis</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Argument, Negotiation, & Persuasion</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
<i>Critique & Evaluation</i>	<i>Low</i>						
	<i>Medium</i>						
	<i>High</i>						
	<i>Low-High</i>						

Note. *Notes were taken regarding a range of complexity and work considered for crosswalk analyses.

Figure C.2 Example Image of Crosswalk Analysis

1	A	B	C	D	E	F	G	H
2	Grade 2 Speaking	Reading	Writing	Writing	Writing	Content-Based Speaking	Content-Based Speaking	Content-Based Speaking
3	Phonemes	181	212					
4		182	213					
5		183	214					
6		184	215					
7		185	216					
8		186	217					
9		187	218					
10		188	219					
11		189	220					
12		190	221					
13		191	222					
14		192	223					
15		193	224					
16		194	225					
17		195	226					
18		196	227					
19		197	228					
20		198	229					
21		199	230					
22		200	231					
23		201	232					
24		202	233					
25		203	234					
26		204	235					
27		205	236					
28		206	237					
29		207	238					
30		208	239					
31		209	240					
32		210	241					
33		211	242					
34		212	243					
35		213	244					
36		214	245					
37		215	246					
38		216	247					
39		217	248					
40		218	249					
41		219	250					
42		220	251					
43		221	252					
44		222	253					
45		223	254					
46		224	255					
47		225	256					
48		226	257					
49		227	258					
50		228	259					
51		229	260					
52		230	261					
53		231	262					
54		232	263					
55		233	264					
56		234	265					
57		235	266					
58		236	267					
59		237	268					
60		238	269					
61		239	270					
62		240	271					
63		241	272					
64		242	273					
65		243	274					
66		244	275					
67		245	276					
68		246	277					
69		247	278					
70		248	279					
71		249	280					
72		250	281					
73		251	282					
74		252	283					
75		253	284					
76		254	285					
77		255	286					
78		256	287					
79		257	288					
80		258	289					
81		259	290					
82		260	291					
83		261	292					
84		262	293					
85		263	294					
86		264	295					
87		265	296					
88		266	297					
89		267	298					
90		268	299					
91		269	300					
92		270	301					
93		271	302					
94		272	303					
95		273	304					
96		274	305					
97		275	306					
98		276	307					
99		277	308					
100		278	309					
101		279	310					
102		280	311					
103		281	312					
104		282	313					
105		283	314					
106		284	315					
107		285	316					
108		286	317					
109		287	318					
110		288	319					
111		289	320					
112		290	321					
113		291	322					
114		292	323					
115		293	324					
116		294	325					
117		295	326					
118		296	327					
119		297	328					
120		298	329					
121		299	330					
122		300	331					
123		301	332					
124		302	333					
125		303	334					
126		304	335					
127		305	336					
128		306	337					
129		307	338					
130		308	339					
131		309	340					
132		310	341					
133		311	342					
134		312	343					
135		313	344					
136		314	345					
137		315	346					
138		316	347					
139		317	348					
140		318	349					
141		319	350					
142		320	351					
143		321	352					
144		322	353					
145		323	354					
146		324	355					
147		325	356					
148		326	357					
149		327	358					
150		328	359					
151		329	360					
152		330	361					
153		331	362					
154		332	363					
155		333	364					
156		334	365					
157		335	366					
158		336	367					
159		337	368					
160		338	369					
161		339	370					
162		340	371					
163		341	372					
164		342	373					
165		343	374					
166		344	375					
167		345	376					
168		346	377					
169		347	378					
170		348	379					
171		349	380					
172		350	381					
173		351	382					
174		352	383					
175		353	384					
176		354	385					
177		355	386					
178		356	387					
179		357	388					
180		358	389					
181		359	390					
182		360	391					
183		361	392					
184		362	393					
185		363	394					
186		364	395					
187		365	396					
188		366	397					
189		367	398					
190		368	399					
191								

Appendix D

Linkage Results

Tables presented in Appendix D are organized by grade (2, 5, 7, and 9). For each grade, a first table reports overall ratability, modality, and complexity results of all standards and for each content area (ELD Listening & Speaking, ELD Reading, ELD Writing, ELA, Mathematics, and Science). A second table for each grade presents the results for language-demand codes organized by linguistic skills and academic language functions.

Grade 2

Table D.1 Grade 2 Linkage Rating Results: Ratability, Modality, & Complexity

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Math	Science	Subtotal	%
Total Standards	259	100.0	26	65	31	122	100.0	59	49	29	137	100.0
Ratability												
<i>Insufficient</i>	68	26.3	1	6	2	9	7.4	3	38	18	59	43.1
<i>Sufficient</i>	191	73.7	25	59	29	113	92.6	56	11	11	78	56.9
Modality												
<i>Listening</i>	35	13.5	2	11	3	16	13.1*	14	2	3	19	13.9**
<i>Speaking</i>	135	52.1	25	46	2	73	59.8	43	11	8	62	45.3
<i>Reading</i>	56	21.6	0	27	1	28	23.0	24	2	2	28	20.4
<i>Writing</i>	115	44.4	10	21	29	60	49.2	34	11	10	55	40.1
Complexity												
<i>Low</i>	114	44.0	7	44	18	69	56.6*	28	9	8	45	32.8**
<i>Medium</i>	64	24.7	18	13	11	42	34.4	20	1	1	22	16.1
<i>High</i>	6	2.3	0	1	0	1	0.8	3	1	1	5	3.6
<i>Low-High</i>	7	2.7	0	1	0	1	0.8	5	0	1	6	4.4

Note. * Percent of all ELD standards ($n = 122$); ** Percent of all content standards ($n = 137$).

Table D.2 Grade 2 Linkage Rating Results: Language Demands

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Math	Science	Subtotal	%
Linguistic Skills												
<i>Phonemes</i>	15	5.8	6	6	1	13	10.7*	2	0	0	2	1.5**
<i>Syllables</i>	9	3.5	5	1	0	6	4.9	3	0	0	3	2.2
<i>Morphemes</i>	12	4.6	0	5	0	5	4.1	7	0	0	7	5.1
<i>Vocabulary Words</i>	27	10.4	0	18	4	22	18.0	4	1	0	5	3.6
<i>Phrases & Sentences</i>	56	21.6	15	19	12	46	37.7	8	0	2	10	7.3
<i>Sound-Symbol Correspondences</i>	10	3.9	0	7	0	7	5.7	3	0	0	3	2.2
<i>Written English Conventions</i>	36	13.9	4	8	12	24	19.7	11	1		12	8.8
Academic Language Functions												
<i>Identification</i>	29	11.2	3	19	0	22	18.0	6	0	1	7	5.1
<i>Labeling & Enumeration</i>	11	4.2	0	2	0	2	1.6	1	5	3	9	6.6
<i>Classification & Sequencing</i>	14	5.4	0	2	3	5	4.1	5	3	1	9	6.6
<i>Definition</i>	3	1.2	0	2	0	2	1.6	1	0	0	1	0.7
<i>Interpretation</i>	8	3.1	1	2	0	3	2.5	4	1	0	5	3.6
<i>Comparison/Contrast</i>	8	3.1	0	1	0	1	0.8	4	2	1	7	5.1
<i>Explanation</i>	4	1.5	1	0	0	1	0.8	3	0	0	3	2.2
<i>Description</i>	27	10.4	0	1	9	10	8.2	7	4	6	17	12.4
<i>Inquiry</i>	12	4.6	6	1	0	7	5.7	3	1	1	5	3.6
<i>Prediction, Generalization, Inference, & Hypothesis</i>	5	1.9	0	3	0	3	2.5	2	0	0	2	1.5
<i>Retelling & Summary</i>	16	6.2	5	4	0	9	7.4	7	0	0	7	5.1
<i>Analysis & Synthesis</i>	0	0.0	0	0	0	0	0.0	0	0	0	0	0.0
<i>Argument, Negotiation, & Persuasion</i>	1	0.4	0	0	0	0	0.0	0	0	1	1	0.7
<i>Critique & Evaluation</i>	2	0.8	0	0	0	0	0.0	2	0	0	2	1.5

Note. * Percent of all ELD standards (n = 122); ** Percent of all content standards (n = 137).

Grade 5

Table D.3 Grade 5 Linkage Rating Results: Ratability, Modality, & Complexity

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics	Science	Subtotal	%
Total Standards	314	100.0	28	93	39	160	100.0	64	46	44	154	100.0
Ratability												
<i>Insufficient</i>	75	23.9	0	4	0	4	2.5	3	38	30	71	46.1
<i>Sufficient</i>	239	76.1	28	89	39	156	97.5	61	8	14	83	53.9
Modality												
<i>Listening</i>	44	14.0	3	15	2	20	12.5*	9	2	13	24	15.6**
<i>Speaking</i>	146	46.5	28	74	3	105	65.6	21	7	13	41	26.6
<i>Reading</i>	109	34.7	0	41	2	43	26.9	50	2	14	66	42.9
<i>Writing</i>	133	42.4	13	39	39	91	56.9	22	7	13	42	27.3
Complexity												
<i>Low</i>	93	29.6	9	54	22	85	53.1*	4	2	2	8	5.2**
<i>Medium</i>	88	28.0	19	25	15	59	36.9	21	3	5	29	18.8
<i>High</i>	51	16.2	0	9	2	11	6.9	30	3	7	40	26.0
<i>Low-High</i>	7	2.2	0	1	0	1	0.6	6	0	0	6	3.9

Note. * Percent of all ELD standards ($n = 160$); ** Percent of all content standards ($n = 154$).

Table D.4 Grade 5 Linkage Rating Results: Language Demands

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics	Science	Subtotal	%
Linguistic Skills												
<i>Phonemes</i>	9	2.9	6	3	0	9	5.6*	0	0	0	0	0.0**
<i>Syllables</i>	6	1.9	5	1	0	6	3.8	0	0	0	0	0.0
<i>Morphemes</i>	12	3.8	0	10	0	10	6.3	2	0	0	2	1.3
<i>Vocabulary Words</i>	27	8.6	0	17	3	20	12.5	6	0	1	7	4.5
<i>Phrases & Sentences</i>	60	19.1	17	26	11	54	33.8	3	0	3	6	3.9
<i>Sound-Symbol Correspondences</i>	5	1.6	0	5	0	5	3.1	0	0	0	0	0.0
<i>Written English Conventions</i>	45	14.3	5	13	16	34	21.3	9	0	2	11	7.1
Academic Language Functions												
<i>Identification</i>	48	15.3	5	36	2	43	26.9	3	1	1	5	3.2
<i>Labeling & Enumeration</i>	5	1.6	0	1	1	2	1.3	0	0	3	3	1.9
<i>Classification & Sequencing</i>	17	5.4	0	5	4	9	5.6	2	0	6	8	5.2
<i>Definition</i>	5	1.6	0	5	0	5	3.1	0	0	0	0	0.0
<i>Interpretation</i>	17	5.4	2	9	0	11	6.9	4	2	0	6	3.9
<i>Comparison/Contrast</i>	6	1.9	0	4	0	4	2.5	1	1	0	2	1.3
<i>Explanation</i>	21	6.7	1	1	0	2	1.3	14	4	1	19	12.3
<i>Description</i>	37	11.8	1	10	14	25	15.6	9	2	1	12	7.8
<i>Inquiry</i>	12	3.8	6	1	0	7	4.4	3	0	2	5	3.2
<i>Prediction, Generalization, Inference, & Hypothesis</i>	12	3.8	0	4	0	4	2.5	3	1	4	8	5.2
<i>Retelling & Summary</i>	10	3.2	5	2	1	8	5.0	2	0	0	2	1.3
<i>Analysis & Synthesis</i>	14	4.5	0	4	0	4	2.5	10	0	0	10	6.5
<i>Argument, Negotiation, & Persuasion</i>	7	2.2	0	0	2	2	1.3	5	0	0	5	3.2
<i>Critique & Evaluation</i>	17	5.4	0	1	0	1	0.6	15	0	1	16	10.4

Note. * Percent of all ELD standards (n = 160); ** Percent of all content standards (n = 154).

Grade 7

Table D.5 Grade 7 Linkage Rating Results: Ratability, Modality, & Complexity

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics	Science	Subtotal	%
Total Standards	370	100.0	29	89	52	170	100.0	82	67	51	200	100.0
Ratability												
<i>Insufficient</i>	98	26.5	1	3	1	5	2.9	2	50	41	93	46.5
<i>Sufficient</i>	272	73.5	28	86	51	165	97.1	80	17	10	107	53.5
Modality												
<i>Listening</i>	30	8.1	3	8	2	13	7.6*	9	3	5	17	8.5**
<i>Speaking</i>	154	41.6	28	63	3	94	55.3	36	16	8	60	30.0
<i>Reading</i>	92	24.9	0	47	9	56	32.9	25	6	5	36	18.0
<i>Writing</i>	165	44.6	16	32	51	99	58.2	39	17	10	66	33.0
Complexity												
<i>Low</i>	86	23.2	7	39	15	61	35.9*	11	8	6	25	12.5**
<i>Medium</i>	106	28.6	19	33	27	79	46.5	17	8	2	27	13.5
<i>High</i>	75	20.3	2	14	9	25	14.7	47	1	2	50	25.0
<i>Low-High</i>	5	1.4	0	0	0	0	0.0	5	0	0	5	2.5

Note. * Percent of all ELD standards ($n = 170$); ** Percent of all content standards ($n = 200$).

Table D.6 Grade 7 Linkage Rating Results: Language Demands

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics	Science	Subtotal	%
Linguistic Skills												
<i>Phonemes</i>	6	1.6	4	2	0	6	3.5*	0	0	0	0	0.0**
<i>Syllables</i>	5	1.4	5	0	0	5	2.9	0	0	0	0	0.0
<i>Morphemes</i>	8	2.2	0	7	0	7	4.1	1	0	0	1	0.5
<i>Vocabulary Words</i>	24	6.5	0	14	3	17	10.0	3	3	1	7	3.5
<i>Phrases & Sentences</i>	65	17.6	17	32	7	56	32.9	6	1	2	9	4.5
<i>Sound-Symbol Correspondences</i>	4	1.1	0	4	0	4	2.4	0	0	0	0	0.0
<i>Written English Conventions</i>	44	11.9	5	17	16	38	22.4	6	0	0	6	3.0
Academic Language Functions												
<i>Identification</i>	49	13.2	5	28	8	41	24.1	5	1	2	8	4.0
<i>Labeling & Enumeration</i>	14	3.8	0	0	5	5	2.9	2	4	3	9	4.5
<i>Classification & Sequencing</i>	25	6.8	0	7	7	14	8.2	9	2	0	11	5.5
<i>Definition</i>	7	1.9	0	6	0	6	3.5	1	0	0	1	0.5
<i>Interpretation</i>	21	5.7	2	9	3	14	8.2	6	1	0	7	3.5
<i>Comparison/Contrast</i>	14	3.8	0	8	3	11	6.5	3	0	0	3	1.5
<i>Explanation</i>	27	7.3	1	9	5	15	8.8	7	3	2	12	6.0
<i>Description</i>	52	14.1	1	7	19	27	15.9	22	2	1	25	12.5
<i>Inquiry</i>	13	3.5	7	0	0	7	4.1	5	0	1	6	3.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	10	2.7	0	2	3	5	2.9	3	2	0	5	2.5
<i>Retelling & Summary</i>	24	6.5	5	4	3	12	7.1	10	0	2	12	6.0
<i>Analysis & Synthesis</i>	22	5.9	1	7	1	9	5.3	13	0	0	13	6.5
<i>Argument, Negotiation, & Persuasion</i>	20	5.4	1		3	4	2.4	13	3	0	16	8.0
<i>Critique & Evaluation</i>	11	3.0	0	0	0	0	0.0	11	0	0	11	5.5

Note. * Percent of all ELD standards (n = 170); ** Percent of all content standards (n = 200).

Grade 9

Table D.7 Grade 9 Linkage Rating Results: Ratability, Modality, & Complexity

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics*	Science	Subtotal	%
Total Standards	438	100.0	34	103	52	189	100.0	126	32	91	249	100.0
Ratability												
<i>Insufficient</i>	118	26.9	1	4	1	6	3.2	6	27	79	112	45.0
<i>Sufficient</i>	320	73.1	33	99	51	183	96.8	120	5	12	137	55.0
Modality												
<i>Listening</i>	26	5.9	3	7	1	11	5.8*	11	0	4	15	6.0**
<i>Speaking</i>	194	44.3	33	73	4	110	58.2	72	4	8	84	33.7
<i>Reading</i>	103	23.5		55	10	65	34.4	28	2	8	38	15.3
<i>Writing</i>	225	51.4	22	38	50	110	58.2	100	5	10	115	46.2
Complexity												
<i>Low</i>	99	22.6	8	40	16	64	33.9*	29	1	5	35	14.1**
<i>Medium</i>	131	29.9	23	36	23	82	43.4	46	0	3	49	19.7
<i>High</i>	82	18.7	2	23	12	37	19.6	38	4	3	45	18.1
<i>Low-High</i>	8	1.8	0	0	0	0	0.0	7	0	1	8	3.2

Note. * Percent of all ELD standards ($n = 189$); ** Percent of all content standards ($n = 249$).

Table D.8 Grade 9 Linkage Rating Results: Language Demands

	All Standards		ELD Standards					Content-Area Standards				
	N	%	Listening & Speaking	Reading	Writing	Subtotal	%	ELA	Mathematics	Science	Subtotal	%
Linguistic Skills												
<i>Phonemes</i>	7	1.6	5	2	0	7	3.7*	0	0	0	0	0.0**
<i>Syllables</i>	5	1.1	5	0	0	5	2.6	0	0	0	0	0.0
<i>Morphemes</i>	11	2.5	0	10	0	10	5.3	1	0	0	1	0.4
<i>Vocabulary Words</i>	28	6.4	0	19	2	21	11.1	7	0	0	7	2.8
<i>Phrases & Sentences</i>	83	18.9	19	37	6	62	32.8	19	2	0	21	8.4
<i>Sound-Symbol Correspondences</i>	4	0.9	0	4	0	4	2.1		0	0	0	0.0
<i>Written English Conventions</i>	75	17.1	7	26	15	48	25.4	27	0	0	27	10.8
Academic Language Functions												
<i>Identification</i>	62	14.2	5	31	8	44	23.3	15	0	3	18	7.2
<i>Labeling & Enumeration</i>	11	2.5	1	0	6	7	3.7	1	0	3	4	1.6
<i>Classification & Sequencing</i>	34	7.8	1	7	9	17	9.0	15	1	1	17	6.8
<i>Definition</i>	6	1.4	0	5	0	5	2.6	0	0	1	1	0.4
<i>Interpretation</i>	22	5.0	2	10	2	14	7.4	5	1	2	8	3.2
<i>Comparison/Contrast</i>	21	4.8	0	11	4	15	7.9	6	0	0	6	2.4
<i>Explanation</i>	38	8.7	3	6	5	14	7.4	20	2	2	24	9.6
<i>Description</i>	27	6.2	1	3	15	19	10.1	6	0	2	8	3.2
<i>Inquiry</i>	13	3.0	8	0	0	8	4.2	4	0	1	5	2.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	11	2.5	0	1	2	3	1.6	8	0	0	8	3.2
<i>Retelling & Summary</i>	24	5.5	5	5	4	14	7.4	10	0	0	10	4.0
<i>Analysis & Synthesis</i>	36	8.2	1	13	1	15	7.9	20	0	1	21	8.4
<i>Argument, Negotiation, & Persuasion</i>	24	5.5	1	2	7	10	5.3	13	1	0	14	5.6
<i>Critique & Evaluation</i>	20	4.6	0	5	0	5	2.6	13	2	0	15	6.0

Note. * Percent of all ELD standards (n = 189); ** Percent of all content standards (n = 249).

Appendix E

Alignment Results

Tables presented in Appendix E are organized by grade span (K – 2, 3 – 5, 6 – 8, and 9 – 12). For each grade span, the first table reports overall frequencies of ELD standards and CELDT items and ratability codes, modalities, complexity, and language-demand codes for the CELDT items. Results are grouped for greater ease of interpretation by three groupings based on both ELD standards and CELDT sections. The subsequent tables display alignment analyses by each test section: Listening, Speaking, Reading, and Writing.

Grades K – 2

Table E.1 Grades K – 2 Alignment Rating Results Summary

	Listening & Speaking	Reading	Writing	N	% of Items
Total ELD Standards Represented on CELDT	6	11	3	20	
Total CELDT Items	40	35	24	99	100.0
Ratability					
<i>Insufficient</i>	0	0	0	0	0.0
<i>Sufficient</i>	40	35	24	99	100.0
Modality					
<i>Listening</i>	40	2	0	42	42.4
<i>Speaking</i>	20	0	0	20	20.2
<i>Reading</i>	0	35	19	54	54.5
<i>Writing</i>	0	0	5	5	5.1
Complexity					
<i>Low</i>	31	22	18	71	71.7
<i>Medium</i>	8	13	6	27	27.3
<i>High</i>	1	0	0	1	1.0
<i>Low-High</i>	0	0	0	0	0.0
Language Demands					
<i>Phonemes</i>	0	5	0	5	5.1
<i>Syllables</i>	0	0	0	0	0.0
<i>Morphemes</i>	5	5	0	10	10.1
<i>Vocabulary Words</i>	33	11	0	44	44.4
<i>Phrases & Sentences</i>	22	27	24	73	73.7
<i>Sound-Symbol Correspondences</i>	0	5	0	5	5.1
<i>Written English Conventions</i>	0	2	10	12	12.1
<i>Identification</i>	23	8	0	31	31.3
<i>Labeling & Enumeration</i>	0	6	0	6	6.1
<i>Classification & Sequencing</i>	1	2	1	4	4.0
<i>Definition</i>	5	0	0	5	5.1
<i>Interpretation</i>	0	2	0	2	2.0
<i>Comparison/Contrast</i>	3	8	1	12	12.1
<i>Explanation</i>	2	0	0	2	2.0
<i>Description</i>	0	0	5	5	5.1
<i>Inquiry</i>	0	1	0	1	1.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	0	3	0	3	3.0
<i>Retelling & Summary</i>	1	1	1	3	3.0
<i>Analysis & Synthesis</i>	2	0	0	2	2.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	0	0	0.0
<i>Critique & Evaluation</i>	0	0	0	0	0.0

Table E.2 Grades K – 2 Alignment Rating Results by ELD Proficiency Level

	Proficiency Level*					N	% of Items
	B	EI	I	EA	A		
Ratability							
<i>Insufficient</i>	0	0	0	0	0	0	0.0
<i>Sufficient</i>	36	8	21	15	19	99	100.0
Modality							
<i>Listening</i>	29	2	7	4	0	42	42.4
<i>Speaking</i>	17	2	0	1	0	20	20.2
<i>Reading</i>	8	2	14	11	19	54	54.5
<i>Writing</i>	0	4	1	0	0	5	5.1
Complexity							
<i>Low</i>	30	8	11	8	14	71	71.7
<i>Medium</i>	5	0	10	7	5	27	27.3
<i>High</i>	1	0	0	0	0	1	1.0
<i>Low-High</i>	0	0	0	0	0	0	0.0
Language Demands							
<i>Phonemes</i>	3	1	1	0	0	5	5.1
<i>Syllables</i>	0	0	0	0	0	0	0.0
<i>Morphemes</i>	5	0	0	5	0	10	10.1
<i>Vocabulary Words</i>	24	0	14	6	0	44	44.4
<i>Phrases & Sentences</i>	19	8	15	12	19	73	73.7
<i>Sound-Symbol Correspondences</i>	3	1	1	0		5	5.1
<i>Written English Conventions</i>	0	5	2	0	5	12	12.1
<i>Identification</i>	22	0	7	2	0	31	31.3
<i>Labeling & Enumeration</i>	0	0	1	5	0	6	6.1
<i>Classification & Sequencing</i>	1	0	3	0	0	4	4.0
<i>Definition</i>	5	0	0	0	0	5	5.1
<i>Interpretation</i>	0	0	2	0	0	2	2.0
<i>Comparison/Contrast</i>	4	0	4	3	1	12	12.1
<i>Explanation</i>	0	2	0	0	0	2	2.0
<i>Description</i>	0	4	1	0	0	5	5.1
<i>Inquiry</i>	0	1	0	0	0	1	1.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	0	0	3	0	0	3	3.0
<i>Retelling & Summary</i>	0	0	1	2	0	3	3.0
<i>Analysis & Synthesis</i>	0	0	0	2	0	2	2.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	0	0	0	0	0.0
<i>Critique & Evaluation</i>	0	0	0	0	0	0	0.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced

Table E.3 Grades K – 2 Listening Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	3	33.3	1	100.0
2	1	1	3	33.3	1	100.0
3	1	1	3	33.3	1	100.0
4	1	.	3	33.3	1	100.0
5	1	.	3	33.3	1	100.0
6	1	1	3	33.3	1	100.0
7	1	.	1	0.0	1	0.0
8	1	1	1	0.0	1	0.0
9	1	.	1	0.0	1	0.0
10*
11
12
13
14
15
16
17
18
19
20
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	9	5	3	22.2	1	66.7

Note. * The standards for items 10-20 were unratable.

Table E.4 Grades K – 2 Speaking Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	1	100.0	3	.
2	1	1	1	100.0	3	.
3	1	1	1	100.0	3	.
4	1	1	1	100.0	3	.
5	1	1	1	100.0	3	.
6	1	1	1	100.0	3	.
7	1	1	1	100.0	3	.
8	1	.	1	100.0	3	.
9	1	1	1	100.0	3	.
10	1	1	1	100.0	3	.
11	1	1	1	100.0	3	.
12	1	1	1	100.0	3	.
13	1	1	1	100.0	3	.
14	1	1	1	100.0	3	.
15	1	1	1	100.0	3	33.3
16	1	1	1	100.0	3	.
17	1	1	1	100.0	3	.
18	1	.	2	50.0	3	33.3
19	1	.	2	50.0	3	33.3
20	1	1	2	50.0	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	17	1	92.5	3	50.0

Table E.5 Grade 2 Reading Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	2	50.0	1	100.0
2	1	1	2	50.0	1	100.0
3	1	.	2	.	1	100.0
4	1	.	2	.	1	100.0
5	1	1	2	50.0	3	33.3
6	1	.	2	50.0	3	33.3
7	1	.	2	50.0	3	33.3
8	1	.	2	50.0	3	33.3
9	1	1	2	50.0	3	33.3
10	1	1	2	.	1	100.0
11	1	1	3	33.3	2	100.0
12	1	1	3	33.3	2	50.0
13	1	1	1	100.0	2	50.0
14	1	1	3	33.3	2	100.0
15	1	.	3	33.3	2	100.0
16	1	.	3	33.3	2	100.0
17	1	.	3	33.3	2	50.0
18	1	1	3	33.3	2	100.0
19	1	1	2	50.0	1	100.0
20	1	1	3	33.3	2	50.0
21	1	1	2	50.0	1	100.0
22	1	1	2	50.0	1	100.0
23	1	1	2	.	3	33.3
24	1	1	2	.	2	50.0
25	1	1	2	.	2	50.0
26	1	.	1	.	4	25.0
27	1	1	2	.	2	50.0
28	1	.	2	.	3	33.3
29	1	1	2	.	2	50.0
30	1	1	2	.	3	33.3
31	1	1	2	.	2	50.0
32	1	1	2	.	3	33.3
33	1	1	2	.	1	100.0
34	1	1	2	.	1	100.0
35	1	1	2	.	3	66.7
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
35	35	25	2	45.6	2	66.9

Note. Grades K and 1 do not take Reading or Writing portions of the CELDT.

Table E.6 Grade 2 Writing Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	2	.	1	100.0
2	1	.	2	.	1	100.0
3	1	.	2	.	1	100.0
4	1	.	2	.	1	100.0
5	1	.	2	.	1	100.0
6	1	.	2	.	1	100.0
7	1	.	2	.	1	100.0
8	1	.	2	.	1	100.0
9	1	.	2	.	1	100.0
10	1	1	2	.	1	100.0
11	1	1	2	.	1	100.0
12	1	1	2	.	1	100.0
13	1	.	2	.	1	100.0
14	1	.	2	.	1	100.0
15	1	.	2	.	1	100.0
16	1	.	2	.	1	100.0
17	1	1	2	.	1	100.0
18	1	1	2	.	1	100.0
19	1	.	2	.	1	100.0
20	1	1	1	100.0	1	100.0
21	1	1	1	100.0	1	100.0
22	1	1	1	100.0	1	.
23	1	1	1	100.0	1	100.0
24	1	1	1	100.0	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
24	24	10	2	100.0	1	100.0

Grades 3 – 5

Table E.7 Grades 3 – 5 Alignment Rating Results Summary

	Listening & Speaking	Reading	Writing	N	% of Items
Total ELD Standards Represented on CELDT	7	14	4	25	
Total CELDT Items	40	35	24	99	100.0
Ratability					
<i>Insufficient</i>	0	0	0	0	0.0
<i>Sufficient</i>	40	35	24	99	100.0
Modality					
<i>Listening</i>	40	0	0	40	40.4
<i>Speaking</i>	20	0	0	20	20.2
<i>Reading</i>	0	35	19	54	54.5
<i>Writing</i>	0	0	5	5	5.1
Complexity					
<i>Low</i>	31	24	19	74	74.7
<i>Medium</i>	8	10	4	22	22.2
<i>High</i>	1	1	0	2	2.0
<i>Low-High</i>	0	0	1	1	1.0
Language Demands					
<i>Phonemes</i>	20	9	0	29	29.3
<i>Syllables</i>	0	2	0	2	2.0
<i>Morphemes</i>	0	2	0	2	2.0
<i>Vocabulary Words</i>	40	35	24	99	100.0
<i>Phrases & Sentences</i>	27	35	24	86	86.9
<i>Sound-Symbol Correspondences</i>	0	8	4	12	12.1
<i>Written English Conventions</i>	1	1	24	26	26.3
<i>Identification</i>	23	7	5	35	35.4
<i>Labeling & Enumeration</i>	13	7	0	20	20.2
<i>Classification & Sequencing</i>	0	0	0	0	0.0
<i>Definition</i>	3	6	0	9	9.1
<i>Interpretation</i>	0	6	0	6	6.1
<i>Comparison/Contrast</i>	0	5	0	5	5.1
<i>Explanation</i>	2	0	0	2	2.0
<i>Description</i>	1	0	5	6	6.1
<i>Inquiry</i>	4	0	0	4	4.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	1	4	0	5	5.1
<i>Retelling & Summary</i>	1	0	5	6	6.1

continued

Table E.7 Grades 3 – 5 Alignment Rating Results Summary (continued)

	Listening & Speaking	Reading	Writing	N	% of Items
<i>Analysis & Synthesis</i>	0	0	0	0	0.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	0	0	0.0
<i>Critique & Evaluation</i>	0	1	0	1	1.0

Table E.8 Grades 3 – 5 Alignment Rating Results by ELD Proficiency Level

	Proficiency Level					N	% of Items
	B	EI	I	EA	A		
<i>Insufficient</i>	0	0	0	0	0	0	0.0
<i>Sufficient</i>	19	29	15	15	21	99	100.0
Modality							
<i>Listening</i>	13	17	9	1	0	40	40.4
<i>Speaking</i>	13	6	0	1	0	20	20.2
<i>Reading</i>	6	8	6	14	20	54	54.5
<i>Writing</i>	0	4	0	0	1	5	5.1
Complexity							
<i>Low</i>	18	20	11	5	20	74	74.7
<i>Medium</i>	1	9	3	9	0	22	22.2
<i>High</i>	0	0	1	1	0	2	2.0
<i>Low-High</i>	0	0	0	0	1	1	1.0
Language Demands							
<i>Phonemes</i>	18	10	0	1	0	29	29.3
<i>Syllables</i>	0	2	0	0	0	2	2.0
<i>Morphemes</i>	0	1	0	0	1	2	2.0
<i>Vocabulary Words</i>	19	29	15	15	21	99	100.0
<i>Phrases & Sentences</i>	6	29	15	15	21	86	86.9
<i>Sound-Symbol Correspondences</i>	5	6	0	0	1	12	12.1
<i>Written English Conventions</i>	0	6	0	0	20	26	26.3
<i>Identification</i>	1	19	7	7	1	35	35.4
<i>Labeling & Enumeration</i>	11	2	4	3	0	20	20.2
<i>Classification & Sequencing</i>	0	0	0	0	0	0	0.0
<i>Definition</i>	2	1	4	2	0	9	9.1
<i>Interpretation</i>	0	1	0	5	0	6	6.1
<i>Comparison/Contrast</i>	0	1	3	1	0	5	5.1
<i>Explanation</i>	0	2	0	0	0	2	2.0
<i>Description</i>	0	4	0	1	1	6	6.1
<i>Inquiry</i>	0	4	0	0	0	4	4.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	0	0	2	3	0	5	5.1
<i>Retelling & Summary</i>	0	4	0	1	1	6	6.1
<i>Analysis & Synthesis</i>	0	0	0	0	0	0	0.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	0	0	0	0	0.0
<i>Critique & Evaluation</i>	0	0	0	1	0	1	1.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced

Table E.9 Grades 3 – 5 Listening Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	2	.	1	.
2	1	.	2	.	1	.
3	1	.	2	.	1	.
4	1	.	2	.	1	.
5	1	1	2	.	1	.
6	1	.	2	.	1	.
7	1	.	2	.	1	.
8	1	.	2	.	1	.
9	1	.	2	.	1	.
10	1	.	2	.	1	.
11	1	.	2	.	1	.
12	1	1	3	33.3	1	100.0
13	1	1	3	33.3	1	100.0
14	1	1	3	33.3	1	.
15	1	.	3	33.3	1	100.0
16	1	.	3	33.3	1	100.0
17	1	.	3	33.3	1	.
18	1	1	3	33.3	1	100.0
19	1	1	3	33.3	1	100.0
20	1	1	3	33.3	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	7	2	33.3	1	100.0

Table E.10 Grade 3 – 5 Speaking Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	1	100.0	3	33.3
2	1	1	1	100.0	3	33.3
3	1	1	1	100.0	3	33.3
4	1	1	1	100.0	3	33.3
5	1	1	1	100.0	3	33.3
6	1	1	1	100.0	3	33.3
7	1	1	1	100.0	3	33.3
8	1	1	1	100.0	1	100.0
9	1	1	1	100.0	3	33.3
10	1	1	1	100.0	3	33.3
11	1	1	1	100.0	3	33.3
12	1	1	1	100.0	1	100.0
13	1	1	1	100.0	1	100.0
14	1	.	1	100.0	1	100.0
15	1	.	1	100.0	1	100.0
16	1	.	1	100.0	1	100.0
17	1	.	1	100.0	1	100.0
18	1	.	2	50.0	3	33.3
19	1	.	2	50.0	3	33.3
20	1	1	2	50.0	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	14	1	92.5	3	60.0

Table E.11 Grades 3 – 5 Reading Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	3	33.3	1	100.0
2	1	1	3	33.3	1	100.0
3	1	1	2	50.0	3	66.7
4	1	1	2	50.0	3	66.7
5	1	1	2	50.0	2	50.0
6	1	1	2	50.0	3	33.3
7	1	1	3	33.3	1	100.0
8	1	1	3	33.3	1	100.0
9	1	1	2	50.0	2	50.0
10	1	1	3	33.3	1	100.0
11	1	1	2	50.0	1	100.0
12	1	1	1	100.0	2	100.0
13	1	1	1	100.0	2	100.0
14	1	1	1	100.0	2	.
15	1	1	1	100.0	2	.
16	1	1	1	100.0	2	.
17	1	1	2	50.0	2	100.0
18	1	1	2	50.0	2	100.0
19	1	1	1	100.0	2	100.0
20	1	1	2	50.0	2	100.0
21	1	1	2	50.0	2	100.0
22	1	.	2	50.0	3	.
23	1	1	2	50.0	2	100.0
24	1	.	3	33.3	2	100.0
25	1	.	2	.	4	25.0
26	1	.	2	.	1	.
27	1	.	2	.	1	100.0
28	1	.	2	.	2	50.0
29	1	.	2	.	2	50.0
30	1	.	2	.	1	.
31	1	.	2	.	1	100.0
32	1	1	2	.	1	.
33	1	1	2	.	1	.
34	1	.	2	.	1	100.0
35	1	.	2	.	1	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
35	35	24	2	58.3	2	84.3

Table E.12 Grades 3 – 5 Writing Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	2	.	1	100.0
2	1	.	2	50.0	1	100.0
3	1	.	2	.	1	100.0
4	1	.	2	.	1	100.0
5	1	.	2	.	1	100.0
6	1	.	2	.	1	100.0
7	1	.	2	.	1	100.0
8	1	.	2	.	1	100.0
9	1	.	2	50.0	1	100.0
10	1	.	2	.	1	100.0
11	1	.	2	.	1	100.0
12	1	.	2	.	1	100.0
13	1	.	2	.	1	100.0
14	1	.	2	.	1	100.0
15	1	.	2	.	1	100.0
16	1	.	2	.	1	100.0
17	1	.	2	.	1	100.0
18	1	.	2	.	1	100.0
19	1	.	2	.	1	100.0
20	1	.	1	100.0	1	100.0
21	1	.	1	100.0	1	100.0
22	1	.	1	100.0	1	100.0
23	1	.	1	100.0	1	100.0
24	1	.	1	100.0	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
24	24	0	2	85.7	1	100.0

Grades 6 – 8

Table E.13 Grades 6 – 8 Alignment Rating Results Summary

	Listening & Speaking	Reading	Writing	N	% of Items
Total ELD Standards Represented on CELDT	7	11	3	21	
Total CELDT Items	40	35	24	99	100.0
Ratability					
<i>Insufficient</i>	0	0	0	0	0.0
<i>Sufficient</i>	40	35	24	99	100.0
Modality					
<i>Listening</i>	39	1	0	40	40.4
<i>Speaking</i>	19	1	0	20	20.2
<i>Reading</i>	6	34	21	61	61.6
<i>Writing</i>	0	0	24	24	24.2
Complexity					
<i>Low</i>	29	24	19	72	72.7
<i>Medium</i>	11	9	0	20	20.2
<i>High</i>	0	1	0	1	1.0
<i>Low-High</i>	0	1	5	6	6.1
Language Demands					
<i>Phonemes</i>	0	0	0	0	0.0
<i>Syllables</i>	0	1	0	1	1.0
<i>Morphemes</i>	1	6	6	13	13.1
<i>Vocabulary Words</i>	13	5	0	18	18.2
<i>Phrases & Sentences</i>	26	18	19	63	63.6
<i>Sound-Symbol Correspondences</i>	0	3	0	3	3.0
<i>Written English Conventions</i>	1	2	7	10	10.1
<i>Identification</i>	5	8	0	13	13.1
<i>Labeling & Enumeration</i>	2	0	0	2	2.0
<i>Classification & Sequencing</i>	0	0	0	0	0.0
<i>Definition</i>	0	0	0	0	0.0
<i>Interpretation</i>	3	5	0	8	8.1
<i>Comparison/Contrast</i>	0	0	0	0	0.0
<i>Explanation</i>	3	0	1	4	4.0
<i>Description</i>	0	0	5	5	5.1
<i>Inquiry</i>	4	0	0	4	4.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	2	2	0	4	4.0
<i>Retelling & Summary</i>	0	1	0	1	1.0

continued

Table E.13 Grades 6 – 8 Alignment Rating Results Summary (continued)

	Listening & Speaking	Reading	Writing	N	% of Items
<i>Analysis & Synthesis</i>	0	1	0	1	1.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	1	1	1.0
<i>Critique & Evaluation</i>	0	2	0	2	2.0

Table E.14 Grades 6 – 8 Alignment Rating Results by ELD Proficiency Level

	Proficiency Level					N	% of Items
	B	EI	I	EA	A		
<i>Insufficient</i>	0	0	0	0	0	0	0.0
<i>Sufficient</i>	20	19	31	28	1	99	100.0
Modality							
<i>Listening</i>	14	17	9	0	0	40	40.4
<i>Speaking</i>	11	6	3	0	0	20	20.2
<i>Reading</i>	6	7	19	28	1	61	61.6
<i>Writing</i>	0	0	5	19	0	24	24.2
Complexity							
<i>Low</i>	14	11	20	27	0	72	72.7
<i>Medium</i>	5	8	6	1	0	20	20.2
<i>High</i>	0	0	0	0	1	1	1.0
<i>Low-High</i>	1	0	5	0	0	6	6.1
Language Demands							
<i>Phonemes</i>	0	0	0	0	0	0	0.0
<i>Syllables</i>	0	0	0	1	0	1	1.0
<i>Morphemes</i>	0	0	1	12	0	13	13.1
<i>Vocabulary Words</i>	10	0	8	0	0	18	18.2
<i>Phrases & Sentences</i>	7	18	16	21	1	63	63.6
<i>Sound-Symbol Correspondences</i>	2	1	0	0	0	3	3.0
<i>Written English Conventions</i>	0	1	7	2	0	10	10.1
<i>Identification</i>	2	1	10	0	0	13	13.1
<i>Labeling & Enumeration</i>	2	0	0	0	0	2	2.0
<i>Classification & Sequencing</i>	0	0	0	0	0	0	0.0
<i>Definition</i>	0	0	0	0	0	0	0.0
<i>Interpretation</i>	2	0	3	3	0	8	8.1
<i>Comparison/Contrast</i>	0	0	0	0	0	0	0.0
<i>Explanation</i>	1	2	1	0	0	4	4.0
<i>Description</i>	0	0	5	0	0	5	5.1
<i>Inquiry</i>	0	4	0	0	0	4	4.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	1	0	2	0	1	4	4.0
<i>Retelling & Summary</i>	1	0	0	0	0	1	1.0
<i>Analysis & Synthesis</i>	1	0	0	0	0	1	1.0
<i>Argument, Negotiation, & Persuasion</i>	0	0	1	0	0	1	1.0
<i>Critique & Evaluation</i>	1	0	1	0	0	2	2.0

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced

Table E.15 Grades 6 – 8 Listening Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	2	.	1	.
2	1	1	2	.	1	.
3	1	.	2	.	1	.
4	1	.	2	.	1	.
5	1	.	2	.	1	.
6	1	1	2	.	1	.
7	1	.	2	.	1	.
8	1	.	2	.	1	.
9	1	.	2	.	1	.
10	1	.	2	.	1	.
11	1	.	2	.	1	.
12	1	1	3	33.3	1	.
13	1	1	3	33.3	1	100.0
14	1	1	3	33.3	1	100.0
15	1	1	3	33.3	1	100.0
16	1	1	3	33.3	1	.
17	1	1	3	33.3	1	100.0
18	1	1	3	66.7	1	.
19	1	1	3	66.7	1	.
20	1	1	3	66.7	1	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	11	2	44.4	1	100.0

Table E.16 Grades 6 – 8 Speaking Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	1	.	3	33.3
2	1	.	1	.	3	33.3
3	1	.	1	.	3	33.3
4	1	1	1	100.0	3	.
5	1	.	1	100.0	2	.
6	1	.	1	100.0	2	.
7	1	1	1	100.0	3	.
8	1	1	1	100.0	3	.
9	1	1	1	100.0	3	.
10	1	1	1	100.0	3	.
11	1	1	1	100.0	3	.
12	1	.	1	100.0	2	.
13	1	.	1	100.0	2	.
14	1	.	1	100.0	1	100.0
15	1	.	1	100.0	1	100.0
16	1	.	1	100.0	1	100.0
17	1	.	1	100.0	1	100.0
18	1	1	2	50.0	3	33.3
19	1	1	2	50.0	3	33.3
20	1	.	2	.	1	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	8	1	93.8	3	63.0

Table E.17 Grades 6 – 8 Reading Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	3	66.7	1	.
2	1	1	3	33.3	1	.
3	1	1	3	33.3	1	.
4	1	1	2	50.0	1	.
5	1	.	2	50.0	3	.
6	1	.	2	50.0	3	33.3
7	1	.	2	50.0	3	33.3
8	1	.	2	50.0	3	33.3
9	1	.	2	50.0	3	33.3
10	1	.	2	50.0	3	33.3
11	1	.	1	100.0	2	50.0
12	1	.	1	100.0	2	50.0
13	1	.	1	100.0	2	50.0
14	1	.	1	100.0	2	50.0
15	1	.	1	100.0	2	50.0
16	1	.	2	50.0	2	.
17	1	1	2	50.0	2	.
18	1	1	1	100.0	2	.
19	1	.	2	50.0	2	50.0
20	1	.	1	100.0	2	.
21	1	.	2	.	2	50.0
22	1	1	2	.	2	.
23	1	1	2	.	2	.
24	1	.	2	.	1	.
25	1	.	2	.	2	.
26	1	.	2	.	2	.
27	1	1	2	.	2	.
28	1	1	2	.	1	.
29	1	.	2	.	2	50.0
30	1	1	2	.	2	.
31	1	1	2	.	2	.
32	1	.	1	.	2	50.0
33	1	1	1	.	2	100.0
34	1	1	1	.	2	100.0
35	1	1	2	.	2	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
35	35	14	2	66.7	2	51.0

Table E.18 Grades 6 – 8 Writing Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	1	100.0	2	.
2	1	1	1	100.0	2	.
3	1	1	1	100.0	2	.
4	1	1	1	100.0	2	.
5	1	1	1	100.0	2	.
6	1	1	1	100.0	2	50.0
7	1	1	1	100.0	2	.
8	1	1	1	100.0	2	.
9	1	1	1	100.0	2	.
10	1	1	1	100.0	2	.
11	1	1	1	100.0	2	.
12	1	1	1	100.0	2	.
13	1	1	1	100.0	2	.
14	1	1	1	100.0	2	.
15	1	1	1	100.0	2	.
16	1	1	1	100.0	2	.
17	1	1	1	100.0	2	.
18	1	1	1	100.0	2	50.0
19	1	1	1	100.0	2	.
20	1	.	1	100.0	1	.
21	1	.	1	100.0	1	.
22	1	.	1	100.0	1	.
23	1	.	1	100.0	1	.
24	1	.	1	100.0	2	50.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
24	24	19	1	100.0	2	50.0

Grades 9 – 12

Table E.19 Grades 9 – 12 Alignment Rating Results Summary

	Listening & Speaking	Reading	Writing	N	% of Items
Total ELD Standards Represented on CELDT	7	8	4	19	
Total CELDT Items	40	35	24	99	100.0
Ratability					
<i>Insufficient</i>	0	0	0	0	0.0
<i>Sufficient</i>	40	35	24	99	100.0
Modality					
<i>Listening</i>	40	0	0	40	40.4
<i>Speaking</i>	20	0	0	20	20.2
<i>Reading</i>	9	35	20	64	64.6
<i>Writing</i>	0	0	5	5	5.1
Complexity					
<i>Low</i>	13	8	9	30	30.3
<i>Medium</i>	26	18	13	57	57.6
<i>High</i>	1	9	2	12	12.1
<i>Low-High</i>	0	0	0	0	0.0
Language Demands					
<i>Phonemes</i>	20	0	0	20	20.2
<i>Syllables</i>	0	2	0	2	2.0
<i>Morphemes</i>	0	10	10	20	20.2
<i>Vocabulary Words</i>	9	16	15	40	40.4
<i>Phrases & Sentences</i>	30	29	24	83	83.8
<i>Sound-Symbol Correspondences</i>	0	3	0	3	3.0
<i>Written English Conventions</i>	3	2	24	29	29.3
<i>Identification</i>	17	6	0	23	23.2
<i>Labeling & Enumeration</i>	5	0	0	5	5.1
<i>Classification & Sequencing</i>	0	0	0	0	0.0
<i>Definition</i>	2	0	0	2	2.0
<i>Interpretation</i>	0	6	0	6	6.1
<i>Comparison/Contrast</i>	7	7	0	14	14.1
<i>Explanation</i>	2	1	0	3	3.0
<i>Description</i>	0	0	5	5	5.1
<i>Inquiry</i>	3	0	0	3	3.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	5	3	0	8	8.1
<i>Retelling & Summary</i>	10	7	0	17	17.2

continued

Table E.19 Grades 9 – 12 Alignment Rating Results Summary (continued)

	Listening & Speaking	Reading	Writing	N	% of Items
<i>Analysis & Synthesis</i>	1	7	0	8	8.1
<i>Argument, Negotiation, & Persuasion</i>	0	0	1	1	1.0
<i>Critique & Evaluation</i>	1	5	0	6	6.1

Table E.20 Grades 9 – 12 Alignment Rating Results by ELD Proficiency Level

	Proficiency Level					N	% of Items
	B	EI	I	EA	A		
<i>Insufficient</i>	0	0	0	0	0	0	0.0
<i>Sufficient</i>	24	35	19	20	1	99	100.0
Modality							
<i>Listening</i>	13	18	9	0	0	40	40.4
<i>Speaking</i>	13	7	0	0	0	20	20.2
<i>Reading</i>	11	13	19	20	1	64	64.6
<i>Writing</i>	0	5	0	0	0	5	5.1
Complexity							
<i>Low</i>	10	10	3	7	0	30	30.3
<i>Medium</i>	12	21	11	12	1	57	57.6
<i>High</i>	2	4	5	1	0	12	12.1
<i>Low-High</i>	0	0	0	0	0	0	0.0
Language Demands							
<i>Phonemes</i>	13	7	0	0	0	20	20.2
<i>Syllables</i>	2	0	0	0	0	2	2.0
<i>Morphemes</i>	7	0	1	11	1	20	20.2
<i>Vocabulary Words</i>	10	6	7	16	1	40	40.4
<i>Phrases & Sentences</i>	8	35	19	20	1	83	83.8
<i>Sound-Symbol Correspondences</i>	3	0	0	0	0	3	3.0
<i>Written English Conventions</i>	4	6	0	18	1	29	29.3
<i>Identification</i>	6	11	6	0	0	23	23.2
<i>Labeling & Enumeration</i>	5	0	0	0	0	5	5.1
<i>Classification & Sequencing</i>	0	0	0	0	0	0	0.0
<i>Definition</i>	2	0	0	0	0	2	2.0
<i>Interpretation</i>	1	3	2	0	0	6	6.1
<i>Comparison/Contrast</i>	5	3	4	2	0	14	14.1
<i>Explanation</i>	0	3	0	0	0	3	3.0
<i>Description</i>	0	5	0	0	0	5	5.1
<i>Inquiry</i>	0	3	0	0	0	3	3.0
<i>Prediction, Generalization, Inference, & Hypothesis</i>	1	2	5	0	0	8	8.1
<i>Retelling & Summary</i>	3	8	6	0	0	17	17.2
<i>Analysis & Synthesis</i>	0	5	3	0	0	8	8.1
<i>Argument, Negotiation, & Persuasion</i>	0	1	0	0	0	1	1.0
<i>Critique & Evaluation</i>	3	1	2	0	0	6	6.1

Note. * B = Beginning; EI = Early Intermediate; I = Intermediate; EA = Early Advanced; A = Advanced

Table E.21 Grades 9 – 12 Listening Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	.	2	.	1	.
2	1	.	2	.	1	.
3	1	.	2	.	1	.
4	1	1	2	.	1	.
5	1	1	2	.	1	.
6	1	1	2	.	1	.
7	1	.	2	.	1	.
8	1	1	2	.	1	.
9	1	1	2	.	1	.
10	1	1	2	.	1	.
11	1	1	2	.	1	.
12	1	.	3	33.3	1	100.0
13	1	.	3	33.3	1	.
14	1	.	3	33.3	1	100.0
15	1	.	3	33.3	1	100.0
16	1	.	3	33.3	1	.
17	1	.	3	33.3	1	.
18	1	1	3	33.3	1	100.0
19	1	1	3	33.3	1	100.0
20	1	1	3	33.3	1	100.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	10	2	33.3	1	100.0

Table E.22 Grades 9 – 12 Speaking Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	1	100.0	3	33.3
2	1	1	1	100.0	3	33.3
3	1	.	1	100.0	3	66.7
4	1	1	1	100.0	2	.
5	1	.	1	100.0	2	.
6	1	1	1	100.0	2	50.0
7	1	.	1	100.0	2	.
8	1	1	1	100.0	2	50.0
9	1	1	1	100.0	3	33.3
10	1	1	1	100.0	3	33.3
11	1	.	1	100.0	3	33.3
12	1	.	1	100.0	3	33.3
13	1	.	1	100.0	3	33.3
14	1	.	1	100.0	1	100.0
15	1	.	1	100.0	1	100.0
16	1	.	1	100.0	1	100.0
17	1	.	1	100.0	1	100.0
18	1	1	2	50.0	3	33.3
19	1	1	2	50.0	3	33.3
20	1	1	2	50.0	1	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
20	20	10	1	92.5	3	54.2

Table E.23 Grades 9 – 12 Reading Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	2	50.0	2	50.0
2	1	1	2	50.0	2	50.0
3	1	1	2	50.0	2	50.0
4	1	.	2	50.0	2	50.0
5	1	.	2	50.0	2	50.0
6	1	1	2	50.0	2	50.0
7	1	.	2	50.0	2	100.0
8	1	1	1	100.0	2	100.0
9	1	.	1	100.0	2	100.0
10	1	1	1	100.0	2	50.0
11	1	1	2	50.0	3	33.3
12	1	1	2	50.0	3	33.3
13	1	.	1	100.0	2	50.0
14	1	.	1	100.0	2	50.0
15	1	.	1	100.0	2	50.0
16	1	.	1	100.0	2	50.0
17	1	.	1	100.0	2	50.0
18	1	.	1	100.0	2	50.0
19	1	.	1	100.0	2	50.0
20	1	.	1	100.0	1	.
21	1	1	2	50.0	2	.
22	1	1	2	50.0	2	.
23	1	.	1	.	2	50.0
24	1	1	1	.	2	50.0
25	1	.	2	.	3	33.3
26	1	1	1	.	2	50.0
27	1	.	2	50.0	2	.
28	1	.	2	50.0	2	.
29	1	.	1	.	2	50.0
30	1	1	1	.	2	50.0
31	1	1	2	50.0	2	.
32	1	1	1	100.0	1	.
33	1	.	2	.	3	33.3
34	1	.	2	50.0	2	.
35	1	1	1	100.0	2	.
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
35	35	16	1	73.2	2	53.2

Table E.24 Grades 9 – 12 Writing Alignment Results

Item Number	Ratable	Standard Complexity Aligned	N Standard Modalities	% Modalities Aligned	N Standard Language Demands	% Language Demands Aligned
1	1	1	1	.	2	100.0
2	1	1	1	.	2	50.0
3	1	1	1	.	2	50.0
4	1	1	1	.	2	50.0
5	1	1	1	.	2	100.0
6	1	.	1	.	2	50.0
7	1	.	1	.	2	100.0
8	1	.	1	.	2	100.0
9	1	.	1	.	2	100.0
10	1	.	1	.	2	100.0
11	1	.	1	.	2	100.0
12	1	.	1	.	2	100.0
13	1	1	1	.	2	100.0
14	1	.	1	.	2	100.0
15	1	1	2	50.0	1	100.0
16	1	.	1	.	2	100.0
17	1	.	1	.	2	100.0
18	1	1	1	.	2	100.0
19	1	.	1	.	2	100.0
20	1	1	1	100.0	1	100.0
21	1	1	1	100.0	1	100.0
22	1	.	1	100.0	1	100.0
23	1	.	1	100.0	1	100.0
24	1	.	1	100.0	2	50.0
Aligned Items						
N Items	n Ratable	n Aligned by Complexity	Mode N Standard Modalities	Mean % Modalities Aligned	Mode N Standard Language Demands	Mean % Language Demands Aligned
24	24	10	1	91.7	2	89.6

Percent Alignment

Table E.25 Overall Percent Alignment Frequency Distribution for All Modalities

Percent Alignment*	Frequency	Percent	Cumulative Percent
Weak Alignment			
8.33	1	2.08	2.08
12.50	1	2.08	4.17
20.00	2	4.17	8.33
20.83	1	2.08	10.42
25.00	1	2.08	12.50
29.17	1	2.08	14.58
30.00	1	2.08	16.67
35.00	2	4.17	20.83
40.00	2	4.17	25.00
41.67	2	4.17	29.17
Moderate Alignment			
45.00	4	8.33	37.50
45.71	2	4.17	41.67
50.00	2	4.17	45.83
54.29	1	2.08	47.92
55.00	1	2.08	50.00
55.56	1	2.08	52.08
57.14	1	2.08	54.17
66.67	2	4.17	58.33
68.57	2	4.17	62.50
Strong Alignment			
71.43	1	2.08	64.58
74.29	2	4.17	68.75
75.00	1	2.08	70.83
79.17	1	2.08	72.92
80.00	3	6.25	79.17
85.00	1	2.08	81.25
95.83	1	2.08	83.33
100.00	8	16.67	100.00

Note. *Mean=59.54, median=55.28, mode=100

Appendix F

Qualitative Notes

Tables in Appendix F illustrate the types of notes recorded by raters during the rating process. Because of the large number of notes taken during the study, typical examples are represented here.

Table F.1 Example Notes on Science Standards

Grade	Standard & Objective	Notes
2	ES3	Implies a deeper analysis ANASYN. "Students should be able to discuss and identify the origin of things they use in their everyday lives."
	IE4B	Suggest content knowledge, not language demand.
	IE4E	"Using appropriately labeled axes..." could imply CLASEQ.
5	IE6C	Implies ANASYN.
	IE6G	"Use" is unclear.
	IE6H	Implies ANASYN.
7	EV2D	Implies INTERP with example of interpreting a Punnett square. "Using the correct vocabulary is important..."
	IE7B	Implies INQUIR and discourse-level language demands.
	IE7C	Implies ANASYN and ARGNEG; interpret "communicate" in terms of a language demand.
9	GT3C	Pedigree charts requires a key.
	GT4A	"Expository tests"; letters (A, G, C, T, amino acids codes) required in modeling which implies LABENU.
	GT5B	Implies LABENU with modeling. "Apply rules...to explain."
	PY10D	Research implies language demand, but not clear which one.
	IE1L	"Combining and applying concepts" implies ANASYN.

Table F.2 Example Notes on Mathematics Standards

Grade	Standard & Objective	Notes
2	MG21	"Describe" requires LD, but "classify" does not necessarily require LD.
	SP10	"Bar graph" requires labeling; "Organize" suggests classification.
	MR21	"Defend the reasoning used" and "Justify" imply ARGNEG, although this could be all mathematical proof without language.
	MR30	"Note" interpreted to mean speak or write a note.
5	NS112	"Explain why" is interpreted here to mean that students use language to explain; implies INTERP.
	MG200	"Identify, describe, and classify" - classify may occur without language. Describing a relationship requires COMCON.
	SD112	"Display...in appropriate graphs and representations" implies CLASEQ or LABENU.
7	NS24	Explain = Words; Justify = Mathematical Evidence; Support = Either.
	NS25	Implies INTERP but not necessarily language demand.
	ALG15	CLASEQ and INTERP may be implicit.
	MG36	Describe could be non-language.
	MR32	"Note the method" requires generalization.
9	AL1	Intro from Frameworks says, "...students should learn to prove every statement they make."
	AL160	"Give pertinent information" about relations and functions implies COMCON and INTERP.
	AL1241	To "explain the difference between inductive and deductive reasoning" requires not only verbal and written skills but analysis of these two complex method of reasoning. There are ways to prove an idea or concept which could imply INTERP and RETSUM.
	AL1250	"To prove or disprove a statement with justification" can be done algebraically with no English demands but implies verbal and/or written language demand. Implies ARGNEG.

Table F.3 Example Notes on ELA Standards

Grade	Standard & Objective	Notes
2	WS11	Implies CLASEQ. Ideas would require command of the language.
	WS12	Need a code for graphemes.
	LS10	Need a code for Speaking conventions.
	LS14	Need a code to cover commands, following directions.
5	CON100	Standard specifying Standard American English is assumed to be high complexity. Preamble includes written and spoken English; WRICON includes parts of speech and grammatical forms.
	WS113	Includes skim/scan for all types of information needed for research papers.
	WA21B	“Show” not interpretable for language demand.
	WA22A	500–700 word response could include any of the higher level language demands.
	LS17	Media interpreted to include written media.
7	WA11	Could be L–H (assumed highest level).
	WS12	Word “claim” implies ARGNEG.
	WS13	“Impose Structure” interpreted as the main intent of the standard.
	WS17	Need a code for thesis statement/controlling idea.
9	RC26	Need a code for following directions.
	WS10	“Writing process” implies CRIEVA and WRICON.
	WS11	Need a code for thesis statement/controlling idea.
	WA21E	Should all 21 include WRICON and PHRSEN?
	SA22E	Need a code for Speaking conventions.

Table F.4 Example Notes on ELD Standards

Grade	Standard & Objective	Notes
2	10109	Seems like it should be codes but with what code?
	22701	Standard seems to contradict itself.
	22621	Very similar to 22318. Implies INTERP.
	34701	Missing category for penmanship/fine motor skill/orthography.
	35513	"M" because standard says "Use complex vocabulary and sentences."
5	22710	Used ANASYN because of "relationship."
	22810	This could just be physical pointing.
	22318	"Appropriate...expression" may imply INTERP.
	22602	"Apply knowledge..." does not specify modality.
	35010	CLASEQ used because "Follow a model".
7	22015	"Dictionary of words" implies organization.
	22523	Increased complexity over 22521 inferred from "achieve...reading".
	22615	"Use common idioms..." does not specify modality.
	35003	Students must recognize elements that are similar to their experience.
	35011	DESCRI used because "at least four sentences".
	35403	CLASEQ used because "writing process to structure drafts".
	36111	References and bibliography imply organization = CLASEQ.
9	22012	"Respond appropriately" implies all modalities.
	22236	Not sure who is producing "written text".
	22616	"Use common idioms..." does not specify modality.
	35604	INTERP because "explore significant of events".
	35908	Implies PREGEN and ARGNEG; missing tight connection.
	35916	ARGNEG may only be implied by "fit purpose and audience".
	36008	"Using rhetoric."
	36108	Support a thesis.

Table F.5 Example Notes on CELDT Items

Grade	CELDT Page #	CELT Section	CELDT Item Booklet #	Standard Code	Notes
2	2	Listening	4	10605	"sneakers" problematic word
	4	Listening	7	10801	abstract concept of tree as a friend
	4	Listening	9	10801	VOCWOR - "Pretend, imagine"
	10	Speaking	15	10101	COMCON=opposites
	12	Speaking	20	10901	This was done using a Level 4 score in the scoring rubric.
	23	Reading	23	23509	IDENTI=here means RECALL
	28	Writing	4	37601	' = punctuation/WRICON
5	6	Speaking	8	10106	possible to have PHRSEN depending on student response
	9	Reading	10	21202	internal voice
	11	Reading	12	22210	PHRSEN only for directions
	14	Reading	23	22322	opposites=COMCON
	19	Reading	33	23902	recall
	26	Writing	24	35910	may have ANASYN
7	1	Listening	1	10411	No academic language function
	1	Listening	2	10411	INTERP implied
	5	Listening	12	10607	Reading modality implied.
	15	Reading	21	23511	Student is not orally explaining
	22	Writing	1	37303	NUMBERS 1–19: Student task is not writing. Members understand why grammar is in this part of the test but modality is <i>reading</i> .
	22	Writing	4	37303	Need more explicit linguistic category
9	13	Reading	17	22324	More complex because of 3 blanks
	17	Reading	28	23104	PREGEN - create an opinion EXPLAN - reasons for cause of relaxing. Similar to above.
	18	Reading	30	22808	Medium - due to choices
	22	Writing	1	37304	WRICON includes grammatical structures - subject/verb agreement. Refer to CA English Standards: Writing Conventions 1.1, 1.2, 1.3
	23	Writing	11	37304	High due to complexity of answer choices
	26	Writing	23	35016	Changed due to complexity of vocabulary required to answer the questions

Appendix G

Rater Participants

Groups of raters were trained to assign codes by consensus to the linkage and alignment data. The groups are described in Tables G.1 and G.2. All linkage data was rated by groups of CTB/McGraw-Hill employees with expertise in specific content areas, ELD teaching methods, and/or the CELDT. A WestEd consultant, who had participated in the pilot study (Sato, et al., 2005), reviewed codes for consistency and accuracy. Alignment data was rated by groups of California educators recruited for their expertise in teaching EL students, knowledge of the ELD and content standards, and experience with the CELDT.

Table G.1 Linkage & Alignment Rater Groups

Data Type	Content	Raters' Affiliation	N
Linkage	ELA	CTB/McGraw-Hill	5
Linkage	Math & Science	CTB/McGraw-Hill	4
Linkage	ELD	CTB/McGraw-Hill	3*
Alignment	CELDT, K–2	CA Educators	4
Alignment	CELDT, 3–5	CA Educators	4
Alignment	CELDT, 6–8	CA Educators	5
Alignment	CELDT, 9–12	CA Educators	5
Linkage & Alignment	ELD, Science, ELA (Reading), CELDT	WestEd (pilot)	(see Sato, et al., 2005, for group sizes)

Note. * Selections of the ELD standards codes were also reviewed by participating linguistic experts.

Table G.2 Demographic Information for Raters

	Alignment*	Linkage**
Race/ethnicity		
Asian	1	0
Hispanic	3	1
Japanese American	1	0
White	10	8
Gender		
Female	13	8
Male	1	1
Intersexed	1	0
Language(s) Spoken		
English Only	5	7
English & French	1	1
English & Spanish	6	1
English & Farsi	1	0
English & Japanese	1	1
English & Vietnamese	1	0
Region		
North	6	0
Central	3	9
South	7	0
Community Type		
Rural	5	0
Suburban	8	9
Central Urban	1	0
Professional Role		
Administrator	1	0
Administrator/Teacher	1	0
Researcher/Teacher	1	0
Teacher	2	0
Teacher/Educator	8	0
Editor	0	6
Researcher	0	1
Manager	0	2
Other	3	0
Experience with CELDT		
1 year	0	6
2 years	1	0
3 years	2	1
4 years	3	0
5 years	7	0
6 years	3	2

continued

Table G.2 Demographic Information for Alignment Raters (continued)

	Alignment*	Linkage**
Grade Area Teaching Experience		
K – 12	5	3
K – 2	2	0
K – 5	2	0
K – 8	1	0
3 – 5	1	0
3 – 8	1	0
6 – 12	2	1
9 – 12	2	3

Note. * Alignment raters were CA educators. ** Linkage raters were CTB/McGraw-Hill employees.

Appendix H

Alignment Study Workshop Agenda

CELDT Alignment Workshop

March 27 – 28, 2006

Meeting Intent: Using the method detailed in a WestEd Pilot Study, four groups of 4-5 participants will evaluate the alignment of the CELDT Form E and the California ELD standards in grade spans of Kindergarten-2, 3-5, 6-8, and 9-12.

Meeting Location: CTB/McGraw-Hill, Mather facility, 10548 Armstrong Avenue, Mather CA, 95655

Day 1

7:45-8:00: Registration: Participants sign in, pick up materials upon arrival, sign and return confidentiality agreements.

8:00-8:15: Introductions. CDE and CTB staff

8:15-8:30: Review materials package.

8:30-9:15: Overview of Pilot Study, training on language demands and decision rules, and give overview of the Workshop plans.

9:15-9:30: Break

9:30- 11:30: Protocol Training: Participants rate a selection of items and practice consensus-building.

11:30-12:30: Lunch

12:30-1:15: Summarize findings from practice ratings, questions and answers.

1:15- 2:15: Conduct individual test ratings

2:15-2:30: Assemble in large-group room for status check and Q&A.

2:30-5:30: Complete individual test ratings and begin consensus rounds.

Day 2

8:00-10:00: Consensus rounds

10:00-10:15: Break

10:15-11:30: Complete consensus rounds

11:30-12:30: Lunch (CTB staff compile data for presentation)

12:30-4:30: Share data from consensus rounds. Discuss conclusions. Ask participants for general comments on alignment and appropriateness of the blueprint.

Appendix I

Consensus Data Collection Requirements

CELDT Linkage & Alignment Study Consensus Data Collection Requirements

1. An excel sheet will be given to each scribe; this sheet will match the hard copy rating sheets participants used for their individual ratings. Please save this sheet on your computer's hard drive.
2. Make sure you safeguard against losing data by doing the following:
 - a. Ask participants at your table to record consensus data on their own sheets as you go.
 - b. Save the data frequently.
 - c. Make sure your computer is plugged in and that power is coming from the plug (the green light on the power cord is on).
 - d. If you prefer record data manually, transfer results to excel sheet after rating.
3. The scribe is responsible for recording all participants' decisions during the consensus meetings as impartially, consistently, and accurately as possible using the study protocol. This includes the following:
 - a. **Ratability** (*S=sufficient; I=insufficient*)
 - b. **Language Demand codes** (see Protocol for specific codes)
 - c. **Modality of item** (*L=listening; S=speaking; W=writing; R=reading*)
 - d. **Complexity of item** (*L=low; M=medium; H=high. Note: a range may be indicated, e.g., L-M.*)
 - e. **Notes** (Notes should include any specific information the participants deem important. For example, include specified interpretations of words.)
4. Scribes must also reinforce decision rules (see list of Decision Rules, Appendix B), reminding the group if appropriate.
5. Scribes must keep track of time. However, some items will go quickly and others will take more deliberation. Use your best judgment.