

California Alternate Performance Assessment (CAPA) Science Alignment Study: Results

Hilary L. Campbell
Xiaofan Cai
Leslie R. Taylor
Arthur A. Thacker
Wade W. Buckland

Prepared for: California Department of Education
1430 N Street
Sacramento, CA 95814

Date: 14 November 2008

Executive Summary

The San Joaquin County Office of Education (SJCOE) subcontracted with the Human Resources Research Organization (HumRRO) on behalf of the California Department of Education (CDE) to conduct an external alignment study of the science portion of the California Alternate Performance Assessment (CAPA). Results of this independent alignment study are reported as part of California’s Peer Review submission to the United States Department of Education (US ED) under the requirements of the No Child Left Behind (NCLB) Act of 2001.

HumRRO used an adaptation of the Webb method (Webb, 1997; 1999; 2005) to study the alignment of the CAPA science test. In late Fall 2006, we adapted the Webb method to evaluate the CAPA math and English Language Arts tests (Taylor et al., 2007). The resulting report was submitted to and approved by the US ED. We have applied the same basic rules and methodology to analyze the alignment of the CAPA science test. The Webb method includes four alignment indicators: (a) **categorical concurrence**, which broadly addresses the extent to which the assessment and the content strands cover the same content categories; (b) **depth-of-knowledge consistency**, which evaluates the extent to which the assessment and the content expectations require the same level of cognitive complexity; (c) **range-of-knowledge correspondence**, which explores the breadth of knowledge required for the content strands and the assessment; and (d) **balance of representation**, which examines the distribution of assessment tasks across content objectives.

Summary alignment results for all four levels of the CAPA science test are presented below. Results are classified using the following scheme:

- Fully aligned—assessments align to all content strands (100%);
- Highly aligned—assessments align to the majority of strands (70–99%);
- Partially aligned—assessments align well to some strands (50–69%); and
- Weakly aligned—assessments align to less than half the strands (< 50%).

Summary of Webb Alignment Analyses across All Levels of the CAPA Science Test

CAPA Level	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Level I	Fully aligned (4 of 4)	Highly Aligned (3 of 4)	Weakly aligned (1 of 4)	Fully aligned (4 of 4)
Level III	Fully aligned (4 of 4)	Fully aligned (4 of 4)	Weakly aligned (0 of 4)	Fully aligned (4 of 4)
Level IV	Fully aligned (3 of 3)	Fully aligned (3 of 3)	Partially aligned (2 of 3)	Fully aligned (3 of 3)
Level V	Fully aligned (5 of 5)	Fully aligned (5 of 5)	Weakly aligned (1 of 5)	Fully aligned (5 of 5)

All four levels of the CAPA science test were fully or highly aligned across all Webb criteria except range-of-knowledge correspondence, where Level IV was partially aligned and Levels I, III, and V were weakly aligned. This weak range-of-knowledge correspondence likely arises because of the large number of content foci for each level of the CAPA compared with the number of performance tasks. To address this weakness, California may wish to review the number of content foci for each level. Should California judge all the content foci to be essential, it should consider increasing the number of performance tasks to enable broader coverage of the content strands. Given the number of foci (ranging from 18 for Level IV to 29 for Level V) and the number of operational performance tasks (8 at each level of the CAPA), California would need to add quite a few tasks to obtain adequate coverage of the content foci.

Results from the categorical concurrence, DOK consistency, and balance of representation analyses indicate the alignment between the CAPA science content strands and the test is generally strong across all levels—tasks are covering the right general content, are appropriately complex, and are balanced across content foci. If California addresses the lone weakness of range-of-knowledge correspondence by considering the number of content foci or performance tasks, alignment of the CAPA science assessment will be quite strong.

CAPA Science Alignment Report: Results

Table of Contents

Chapter 1: Introduction..... 1
 Report Terminology 1
 Organization and Contents of the Report 2
 Chapter 2: Methodology 3
 The Webb Alignment Method 3
 Alignment Criteria..... 3
 Expert Panelists..... 5
 Interrater Reliability 5
 Procedures 6
 Level I CAPA Science Strands and Assessment..... 8
 Categorical Concurrence 8
 Depth-of-Knowledge Consistency..... 9
 Range-of-Knowledge Correspondence..... 10
 Balance of Representation 11
 Level I Summary..... 11
 Chapter 4: Level III Results 13
 Level III CAPA Science Strands and Assessment..... 13
 Categorical Concurrence 13
 Depth-of-Knowledge Consistency..... 14
 Range-of-Knowledge Correspondence..... 14
 Balance of Representation 15
 Level III Summary..... 16
 Chapter 5: Level IV Results..... 17
 Level IV CAPA Science Strands and Assessment 17
 Categorical Concurrence 17
 Depth-of-Knowledge Consistency..... 18
 Range-of-Knowledge Correspondence..... 18
 Balance of Representation 19
 Level IV Summary 19
 Chapter 6: Level V Results..... 21
 Level V CAPA Science Strands and Assessment 21
 Categorical Concurrence 21
 Depth-of-Knowledge Consistency..... 22
 Range-of-Knowledge Correspondence..... 22
 Balance of Representation 23
 Level V Summary 24
 Chapter 7: Summary and Recommendations 25
 Appendix A Workshop Materials 28
 Depth-of-Knowledge Level Descriptions..... 28
 CAPA Science Strands DOK Rating Sheets..... 29
 CAPA Science Performance Task Rating Sheets 30
 CAPA Science Coding Sheet 31

Appendix B Performance Task Quality..... 32
 Descriptive Alignment-Type Data 34
 Task Quality Feedback and Panelist Comments 36

List of Tables

Table 2.1. DOK Descriptions for the CAPA Science Alignment Study 6
Table 3.1. Number of Content Foci and Performance Tasks for each CAPA Science
 Level I Content Strand 8
Table 3.2. Categorical Concurrence for Level I: Mean Number of Performance Tasks
 per Strand 9
Table 3.3. DOK Consistency for Level I: Mean Percentage of Performance Tasks with
 DOK Below, At, and Above DOK Level of Objectives 10
Table 3.4. Range-of-Knowledge for Level I: Mean Percentage of Content Foci per
 Strand Linked with Performance Tasks 11
Table 3.5. Balance of Representation for Level I: Mean Balance Index per Strand 11
Table 3.6. Summary of CAPA Science Level I Alignment Results 12
Table 4.1. Number of Content Foci and Performance Tasks for each CAPA Science
 Level III Content Strand 13
Table 4.2. Categorical Concurrence for Level III: Mean Number of Performance Tasks
 per Strand 14
Table 4.3. DOK Consistency for Level III: Mean Percentage of Performance Tasks with
 DOK Below, At, and Above DOK Level of Objectives 14
Table 4.4. Range-of-Knowledge for Level III: Mean Percentage of Content Foci per
 Strand Linked with Performance Tasks 15
Table 4.5. Balance of Representation for Level III: Mean Balance Index per Strand 16
Table 4.6. Summary of CAPA Science Level III Alignment Results 16
Table 5.1. Number of Content Foci and Performance Tasks for each Level IV CAPA
 Science Content Strand 17
Table 5.2. Categorical Concurrence for Level IV: Mean Number of Performance Tasks
 per Strand 18
Table 5.3. DOK Consistency for Level IV: Mean Percentage of Performance Tasks with
 DOK Below, At, and Above DOK Level of Objectives 18
Table 5.4. Range-of-Knowledge for Level IV: Mean Percentage of Content Foci per
 Strand Linked with Performance Tasks 19
Table 5.5. Balance of Representation for Level IV: Mean Balance Index per Strand 19
Table 5.6. Summary of CAPA Science Level IV Alignment Results 20
Table 6.1. Number of Content Foci and Performance Tasks for each Level V CAPA
 Science Content Strand 21
Table 6.2. Categorical Concurrence for Level V: Mean Number of Performance Tasks
 per Strand 22
Table 6.3. DOK Consistency for Level V: Mean Percentage of Performance Tasks with
 DOK Below, At, and Above DOK Level of Objectives 22
Table 6.4. Range-of-Knowledge for Level V: Mean Percentage of Content Foci per
 Strand Linked with Performance Tasks 23

Table 6.5. Balance of Representation for Level V: Mean Balance Index per Strand.....	23
Table 6.6. Summary of CAPA Science Level V Alignment Results.....	24
Table 7.1. Summary of Webb Alignment Analyses across All Levels of the CAPA Science Assessment.....	25
Table B.1. Item Quality Rating Codes	32
Table B.2. Distribution of Panelists' Task Quality Ratings.....	32
Table C.1. Distribution of Field Test Task DOK Ratings across CAPA Science Levels	34
Table C.2. Distribution of Field Test Task Linkages across CAPA Science Levels.....	35
Table C.3. Distribution of Field Test Task Alignment Ratings across CAPA Science Levels	36
Table C.4. Distribution of Panelists' Field Test Task Quality Ratings.....	36

Chapter 1: Introduction

The San Joaquin County Office of Education (SJCOE) subcontracted with the Human Resources Research Organization (HumRRO) on behalf of the California Department of Education (CDE) to conduct an external alignment study of the science portion of the California Alternate Performance Assessment (CAPA). Results of this independent alignment study are reported as part of California’s Peer Review submission to the United States Department of Education (US ED) under the requirements of the No Child Left Behind (NCLB) Act of 2001.

Under NCLB, all states accepting Title I funds from the federal government must demonstrate evidence of a valid state assessment system based on rigorous academic standards. One aspect of peer review requirements is how states’ assessments align with the content standards they are designed to measure. Comprehensive alignment studies should examine such facets as range, depth, and level of cognitive complexity. This requirement extends to alternate assessments such as the CAPA—assessments designed for students with only the most severe cognitive disabilities. Although alternate assessments are based on adapted or extended standards and may have reduced complexity or scope (compared with general assessments), they should still align with rigorous academic content standards linked to the state content standards (United States Department of Education, 2005).

HumRRO used an adaptation of the Webb method (Webb, 1997; 1999; 2005) to study the alignment of the CAPA science test. In late Fall 2006, we adapted the Webb method to evaluate the CAPA math and English Language Arts tests (Taylor et al., 2007). The resulting report was submitted to and approved by the US ED. We have applied the same basic rules and methodology to analyze the alignment of the CAPA science test. We describe the details of this methodology in Chapter 2 of this report.

Report Terminology

Although alignment studies across states and content areas share the general purpose of examining the extent to which assessments and content standards are aligned, different methodologies and different states have diverse lexicons. The terminology for content expectations varies considerably. The Webb alignment method refers to the broadest level of content expectations as “standards,” and the most specific level as “content objectives.” If an intermediate level of content expectations exists, Webb calls this level a “goal.”

Throughout this report, we primarily use California’s terminology to refer to various aspects of its standards and assessment. For instance, Webb refers to the broadest level of content expectations as “standards,” whereas California calls the highest level “strands.” This is the level at which alignment results are analyzed and reported. Hence, in this report, at each level of the CAPA, we call the broadest level of content expectations “strands.” For instance, at Level I, the strands are Physical Science, Life Science, Earth Science, and Investigation and Experimentation. At Level

V, the strands are Biology, Chemistry, Physics, Earth Science, and Investigation and Experimentation. In the CAPA science blueprints, the most specific level of the framework is comprised of the foci of the California content standards for the alternate assessment. Denoted with checkmarks in the CAPA science blueprints, these foci represent the extensions of content standards for students taking the alternate assessment. The terms “foci” and “objectives” are used interchangeably throughout this report to refer to the most specific level of the CAPA science content expectations, the “checkmark” level. The assessment “tasks” or “items” test whether students have learned the foci or objectives of the blueprints.

Organization and Contents of the Report

This report contains seven chapters and three appendices. Chapter 1 introduces the study topic and describes the organization of the report. Chapter 2 describes the methodology of the alignment study, including an introduction to the Webb alignment method and descriptions of the panelists and procedures for the alignment workshop. Chapters 3 through 6 discuss alignment results for the four levels of the CAPA science test. Finally, Chapter 7 briefly summarizes the results and offers recommendations.

Appendix A presents sample materials from the alignment workshop. Appendix B contains feedback about the quality of the CAPA science test performance tasks, which panelists evaluated as an ancillary task during the alignment workshop. Finally, Appendix C presents descriptive results for the CAPA science field test tasks.

Chapter 2: Methodology

This section of the report describes HumRRO’s methodology for the external alignment study of the CAPA science test. First, we introduce the Webb method and explain how it relates to the CAPA science blueprints and science assessment. Next, we describe the panelists who provided expert ratings at the alignment workshop and investigate the reliability of the ratings they provided. Finally, we describe the workshop itself and detail how data were collected.

The Webb Alignment Method

For this alignment study, we used the same adapted version of the Webb method (Webb, 1997; 1999; 2005) used in 2006–07 to study the alignment of the CAPA math and ELA tests. This version of the Webb alignment method enables us to evaluate how the alternate assessment’s performance tasks align with the content foci for these students, as required under NCLB. Webb (2005) broadly defines alignment as, “the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do” (p. 2).

Because alternate assessments typically differ from general assessments in terms of cognitive complexity, scope, and design, slight modifications to the Webb method were necessary. These included changes in the definitions of complexity levels for tasks and content objectives and minor modifications in criteria for what constitutes adequate alignment. This modified Webb method (Almond, Filbin, Hall, & Tindal, 2005; Browder et al., 2005; Tindal, 2005) has been used successfully to evaluate alignment in at least seven other states (Roach, Elliott, & Webb, 2005) as well as for the math and ELA portions of the CAPA (Taylor et al., 2007). The CAPA consists of performance tasks scored using a rubric. Because performance tasks are more complex and time-consuming than typical general assessment items, these slight changes to the traditional alignment method are appropriate.

Alignment Criteria

The Webb method produces four main analyses to summarize alignment results (Webb, 2005). Collectively, these alignment criteria convey the extent to which the assessment is aligned with the content expectations in terms of breadth and depth of coverage. Alignment analyses are not “all-or-none” endeavors—the results do not indicate, “Yes, this assessment is aligned,” or “No, this assessment is not aligned.” Rather, alignment is a matter of degree, and we expect to see a pattern of relative strengths and weaknesses in the extent of alignment across these four criteria. We explain each alignment analysis below (as described in Webb, 2005). We also describe the target value for acceptable alignment for each criterion as established during our analysis of the math and ELA portions of the CAPA (Taylor et al., 2007).

The first alignment criterion, ***categorical concurrence***, broadly addresses the extent to which the assessment and the content strands cover the same content categories. If the panelists find the performance tasks of the CAPA science test address all of the content strands for that level, categorical concurrence will be high. Weaker categorical concurrence would suggest panelists did not link performance tasks from the CAPA to each of the content strands listed in the CAPA science blueprints. The target criterion for acceptable categorical concurrence for general assessments is six items per strand. For alternate assessments such as the CAPA, the target is one item for each content strand. This alternate criterion reflects the fact that performance tasks tend to take longer to complete and contain more information in their responses than items in general assessments.

The second alignment criterion, ***depth-of-knowledge consistency***, evaluates the extent to which the assessment and the content expectations require the same level of cognitive complexity. Depth-of-knowledge (DOK) indicates the amount of cognitive demand or level of cognitive processing required for each content objective and performance task. The purpose of the DOK consistency criterion is to ensure knowledge is not assessed at a lower cognitive level than the level targeted in the content expectations. According to Webb, adequate consistency in general assessments requires at least 50 percent of items to be at the same or higher DOK level than the strand they are designed to assess. For this evaluation of the CAPA science test, the criterion is the same as for general assessments—at least half of the performance tasks should be at the same DOK level as, or higher, than the content foci they were designed to assess.

The third alignment criterion, ***range-of-knowledge correspondence***, explores the breadth of knowledge required for the content strands and the assessment. For high range-of-knowledge correspondence results, assessments and the strands they are designed to measure should cover a comparable scope of knowledge. To evaluate range-of-knowledge correspondence within each strand, analyses examine the number of objectives linked to at least one assessment item. For the CAPA science alignment study, this alignment index determines how many foci of the alternate assessments (the “checkmark” level of the CAPA science blueprints) are covered with at least one performance task on the CAPA science test. For alternate assessments, as for general assessments, the target for acceptable range-of-knowledge correspondence is for at least half of the foci within each strand to be linked with a performance task. Because of the limited number of performance tasks relative to the number of content objectives (foci) in the blueprint, this criterion can be particularly difficult to meet in alternate performance assessments. However, alternate assessments can increase the breadth of content covered without increasing the number of performance tasks by linking performance tasks with multiple content strands.

The fourth and final alignment criterion, ***balance of representation***, examines the distribution of assessment tasks across content objectives. Whereas range-of-knowledge correspondence examines the percentage of foci covered within a content strand, balance of representation goes one step further to explore how performance

tasks are distributed across the foci within a content strand. Put another way, balance of representation evaluates the extent to which some foci are emphasized or covered more heavily than others within a given content strand. Analysis results for the balance of representation criterion are conveyed through a *balance index*¹, an index scaled from 0 to 100, which is computed for each content strand based on the distribution of performance tasks across foci within that strand. Balance index values greater than 70 indicate an adequate distribution of tasks within a strand.

Expert Panelists

With assistance from SJCOE, we assembled a panel of nine content area experts to provide alignment ratings for the CAPA science test. SJCOE contacted potential panelists and obtained background information before selecting a panel of nine experts with approval from the CDE. Seven panelists were special education teachers; two held special education specialist or administrative positions. They had an average of 8.3 years' experience in their current positions. All nine panelists were female; six were Caucasian, one was African American, one was Hispanic/Latino, and one was Asian/Pacific Islander. In terms of geographic dispersion across California, four were from the central region, two were from the northern region, two from the southern region, and one from the coastal region.

Interrater Reliability

Because all alignment ratings are treated as averages across panelists, it is useful to determine the extent to which panelists agreed with one another. Webb (2005) recommends using the intraclass correlation coefficient (ICC) method of Shrout and Fleiss (1979) to examine rater agreement. This coefficient measures the amount of variability in the ratings that can be attributed to differences among panelists rather than differences among items. ICC values range from zero to one. Higher ICCs, which indicate a larger proportion of variance linked to differences among tasks rather than among panelists, are better. ICC values greater than 0.70 are considered adequate, and values greater than 0.80 are considered good.

To obtain a picture of rater agreement for the CAPA science alignment study, we computed ICCs for panelists' task DOK ratings for each level. We used the task DOK ratings as our measure of reliability because they are distributed across a defined scale (1 to 4; Webb, 2005). ICC values for Levels I, III, IV, and V² were 0.85, 0.88, 0.76, and 0.64, respectively. Reliability ratings for Levels I and III were good, and the ICC for Level IV was adequate. The ICC for Level V was slightly lower than desired; this may be attributed to rater fatigue or the added complexity of the Level V material. Regardless, the Level V coefficient was close to the target of 0.70 and is not low enough to be of concern.

¹ For more information about the balance index, please see Webb (2005).

² The CAPA science test does not have a Level II because it is first administered at grade 5; the math and ELA portions of the CAPA have a Level II which covers content for grades 2 and 3.

Procedures

HumRRO conducted the alignment workshop July 30–31, 2008 in Sacramento, CA. Paula Carroll of the SJCOE welcomed the panelists, thanked them for their participation, and explained the importance of the alignment study. Everyone participated in group introductions. Next, panelists read and signed affidavits of nondisclosure for the secure CAPA science test materials. HumRRO staff presented a detailed description of alignment studies and briefly explained the tasks the panelists would complete during the alignment workshop. Samples of workshop materials can be found in Appendix A.

The first task for the panel was rating the depth-of-knowledge (DOK) of the CAPA science content strands. HumRRO staff explained what DOK ratings are, introduced the DOK levels for the CAPA science alignment study, and facilitated a session for panelists to practice making DOK ratings for sample content foci. For each sample objective, panelists made individual DOK ratings, then discussed the ratings as a group. Although HumRRO staff facilitated conversation about ratings, we emphasized that we were there to act *only* as facilitators, not to provide direct input on ratings; as content area experts, panelists were responsible for making ratings. Once panelists were comfortable assigning DOK ratings, each individual rated the CAPA science content objectives (foci). When all panelists finished their individual ratings, HumRRO staff facilitated discussion to establish group consensus for the DOK level of each focus of the alternate assessment. Although complete agreement was desired and often achieved, majority agreement was also acceptable to determine the group DOK rating. The DOK descriptions (adapted from Webb, 2005) used for this study are presented in Table 2.1. Though adapted for science content, they closely resemble the descriptions used in HumRRO’s prior alignment study of the ELA and math portions of the CAPA.

Table 2.1. DOK Descriptions for the CAPA Science Alignment Study

DOK Level	Description
Level 1	Requires students to recall or observe facts, definitions, terms. Involves simple one-step procedures. May involve simple calculations. Keywords: identify, recall, recognize, measure
Level 2	This level includes the engagement of some mental processing beyond a habitual response. The item requires students to make some decisions as to how to approach a problem or activity. Keywords: classify, organize, estimate, make observations, collect and display data, compare data
Level 3	A multiple step ‘behavioral event’ is executed in more than one context. Requires reasoning, planning, or use of evidence to solve problem. May involve activity with more than one possible answer. Requires conjecture or restructuring of problems. Involves drawing conclusions from observations, citing evidence and developing logical arguments for concepts. Uses concepts to solve non-routine problems.
Level 4	The ‘behavioral event’ reflects an approach (of many) to completing the task. May require complex reasoning, planning, developing, and thinking. Typically requires extended time to complete problem, but time spent not on repetitive tasks. Requires students to make several connections and apply one approach among many to solve the problem. Involves complex restructuring of data, establishing and evaluating criteria to solve problems.

HumRRO staff conducted a second training session to teach panelists how to rate the CAPA science performance tasks. Panelists were responsible for several ratings for each task. First, they rated the DOK level for each performance task. Next, panelists indicated the most closely related content focus from the CAPA science content strands, or the focus they believed the item is designed to assess. Panelists were allowed to indicate a second content focus if they believed two foci truly matched the performance task equally well. Third, panelists rated the overall alignment of the task, indicating how closely aligned they believed the task is to the content focus to which they matched it. Finally, panelists completed the ancillary task of evaluating the quality of the performance tasks. For each task, panelists rated overall quality and had an opportunity to make comments. Panelists were encouraged to make item quality comments sparingly—only when really warranted—and make the alignment ratings the major focus of their time. Results of these ancillary task quality ratings can be found in Appendix B. Panelists conducted all task ratings independently, although HumRRO staff facilitated group practice sessions for each rating using sample items before panelists began individual ratings.

Chapter 3: Level I Results

This chapter presents complete alignment results for Level I of the CAPA science test. Panelists rated all tasks for one complete version of the Level I CAPA science test (Version 2) which contained all eight operational performance tasks as well as four field test tasks. They also reviewed the remaining Level I field test tasks in Version 3. This chapter contains results for the operational items only; descriptive results for the field test tasks can be found in Appendix C.

Level I CAPA Science Strands and Assessment

Level I of the CAPA is administered to the most severely cognitively disabled students taking the CAPA regardless of their grade level. It consists of eight performance tasks, scored using a five-point rubric. Four content strands are assessed for the Level I CAPA science test: Life Science, Physical Science, Earth Science, and Investigation and Experimentation. Content foci are drawn from the general science strands for kindergarten and first and second grades. Across the four major content strands, a total of 21 content foci are identified for students taking the Level I CAPA science exam. Table 3.1 lists the number of content foci and CAPA performance tasks for each content strand.

Table 3.1. Number of Content Foci and Performance Tasks for each CAPA Science Level I Content Strand

Content Strand	Number of Content Foci	Target Number of CAPA Performance Tasks
Life Science	5	2
Physical Science	9	3
Earth Science	2	2
Investigation & Experimentation	5	1
Total	21	8

Categorical Concurrence

Categorical concurrence is a general measure of the extent of alignment between the content strands and the performance assessment. For alternate assessments, Webb’s target for categorical concurrence is at least one performance task linked with each content strand. As Table 3.2 demonstrates, every Level I content strand met Webb’s target criterion of at least one performance task matched to each content strand. In fact, the average number of tasks panelists assigned to each content strand very closely matched the intended number of performance tasks specified in the CAPA science blueprint. Because they were able to link more than one content focus to each performance task, the mean number of tasks matched to foci by the panelists slightly exceeds the number of actual performance tasks on the CAPA science assessment. Given the limited number of performance tasks on the assessment,

linkage with multiple content strands is beneficial because it enables broader coverage of the content strands with a smaller number of tasks.

Table 3.2. Categorical Concurrence for Level I: Mean Number of Performance Tasks per Strand

Title of Strand	Number of Tasks per Strand			At Least One Task per Strand
	Target # Items from blueprint	Mean Tasks Matched*	Standard Deviation	
Life Science	2	2.11	0.33	Y
Physical Science	3	2.89	1.05	Y
Earth Science	2	2.00	0.00	Y
Investigation & Experimentation	1	1.56	0.88	Y
	8	8.56		
Percent of strands with at least one task				100%

*Because panelists can link tasks to multiple strands, total may be greater than number of target items.

Depth-of-Knowledge Consistency

Depth-of-knowledge (DOK) consistency measures the extent to which performance tasks on the CAPA science test are at the same complexity level as, or higher than, that of the content foci they are designed to assess. Two steps were required for this analysis. First, panelists determined consensus DOK ratings for each content focus. Next, they individually assigned DOK ratings for each performance task. Their average task ratings for each performance task were compared with the consensus DOK rating for the content focus to which it was linked. Within each strand, Webb’s criterion for adequate alignment is that at least 50 percent of the performance tasks should be at least as complex as the content foci they are designed to measure.

Table 3.3 presents the DOK consistency results for Level I. Three of the four content strands met Webb’s target criterion that at least half of the performance tasks should meet or exceed the cognitive complexity of their content objectives. However, for the Earth Science strand, 67 percent of the tasks were judged to be at a lower level of cognitive complexity than their matching content foci. Because Level I of the CAPA is administered only to students with the most severe cognitive disabilities, the CDE may wish to examine the cognitive complexity of the Earth Science content foci to determine if they are truly appropriate for this population. If the foci are appropriately complex, the CDE may wish to consider developing more complex Earth Science items for future administrations.

Table 3.3. DOK Consistency for Level I: Mean Percentage of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	SD	M	SD	M	SD	
Life Science	2.11	0	0.0	85	22.7	15	22.7	Y
Physical Science	2.89	9	26.7	73	38.5	19	33.8	Y
Earth Science	2.00	67	35.4	33	35.4	0	0.0	N
Investigation & Experimentation	1.56	0	0.0	44	47.1	56	47.1	Y
Total	8.56							
Percent of strands with 50% of task DOK at or above objective DOK								75%

Range-of-Knowledge Correspondence

Range-of-knowledge correspondence measures alignment of the breadth of knowledge between the content strands and the assessment. Each content strand contains multiple content foci (see Table 3.1). Well-aligned assessment systems cover a broad range of these content foci on their assessments. Webb’s target for adequate alignment is coverage of at least half (50%) of the foci within a content strand—that is, 50 percent of the content foci should be matched with at least one task. As Table 3.4 displays, only one of the four Level I content strands, Earth Science, met this criterion. The Life Science strand is what Webb would consider weakly aligned, as its coverage falls between 41 and 49 percent.

The Earth Science strand has only two content foci, whereas Life Science, Physical Science, and Investigation and Experimentation have 5, 9, and 5 foci respectively. The Level I CAPA science test has eight performance tasks designed to assess a total of 21 content foci. Although performance tasks may be designed to measure more than one content focus, even an average of two foci for every task would not provide adequate coverage of all foci. If administering more performance tasks is not feasible for the population of students taking Level I of the CAPA, California may wish to review the number of content foci expected for this population and narrow it to include only the truly essential content foci.

Table 3.4. Range-of-Knowledge for Level I: Mean Percentage of Content Foci per Strand Linked with Performance Tasks

Title of Strand	Number of Content Foci	Mean Tasks per Strand	Range of Content Foci			Range-of-Knowledge Target Met
			Content Foci with At Least One Task		% of Total Content Foci per Strand	
			M	SD	M	
Life Science	5	2.11	2.11	0.33	42	N
Physical Science	9	2.89	2.22	0.67	25	N
Earth Science	2	2.00	1.89	0.33	94	Y
Investigation & Experimentation	5	1.56	1.44	0.73	29	N
Total	21	8.56				
Percentage of strands with 50% of Content Foci linked to at least one task						25%

Balance of Representation

Balance of representation measures the distribution of tasks across content foci. To obtain good balance, tasks should be evenly distributed among foci. Content balance is evaluated using Webb’s balance index³. This index takes into account only those foci panelists linked with performance tasks—it does not consider distribution to foci to which panelists did not match tasks. Webb’s criterion for adequate alignment is a balance index score of at least 70. As Table 3.5 demonstrates, all four content strands for Level I reflected adequate balance, suggesting performance tasks were evenly distributed across the content foci panelists judged to be represented on the CAPA science assessment.

Table 3.5. Balance of Representation for Level I: Mean Balance Index per Strand

Title of Strand	Balance of Representation						Balance Index Target Met
	Number of Content Foci	Mean Content Foci Linked with Tasks	Mean Tasks per Strand	Mean % of Tasks Linked to Strand	Mean Balance Index		
					M	SD	
Life Science	5	2.11	2.11	25	100	0.0	Y
Physical Science	9	2.22	2.89	33	91	8.5	Y
Earth Science	2	1.89	2.00	24	100	0.0	Y
Investigation & Experimentation	5	1.44	1.56	18	98	5.6	Y
Total	21	7.66	8.56				
Percentage of strands with a balance of representation index of 70 or greater							100%

Level I Summary

The alignment study of Level I of the CAPA science test demonstrated areas of strength and weakness. All four content strands met the criteria for adequate alignment

³ See Webb (2005) for details on computing the balance index.

for categorical concurrence and balance of representation. Three of the four content strands (Earth Science was the exception) met the criterion for DOK consistency. For range-of-knowledge correspondence, only Earth Science met the target. To address this weakness in range, California may wish to consider the number of content foci targeted to Level I students or review the number of performance tasks administered. Table 3.6 summarizes alignment results for Level I of the CAPA science assessment across all four Webb alignment criteria.

Table 3.6. Summary of CAPA Science Level I Alignment Results

Content Strand	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Life Science	Y	Y	N	Y
Physical Science	Y	Y	N	Y
Earth Science	Y	N	Y	Y
Investigation & Experimentation	Y	Y	N	Y

Chapter 4: Level III Results

This chapter presents complete alignment results for Level III of the CAPA science assessment. Panelists rated all tasks for one complete version of the Level III CAPA science test (Version 2), which contained eight operational performance tasks as well as four field test tasks. They also reviewed the remaining Level III field test tasks in Version 3. This chapter contains results for the operational items only; descriptive results for the field test tasks can be found in Appendix C.

Level III CAPA Science Strands and Assessment

Level III of the CAPA is administered in Grade 5 to students with the most severe cognitive disabilities. It consists of eight performance tasks scored with a four-point rubric. Four content strands are assessed for the Level III CAPA science assessment: Life Science, Physical Science, Earth Science, and Investigation and Experimentation. Content foci are drawn from the general science content expectations for grades four and five. Across the four major content strands, the Level III CAPA science blueprint identifies 28 content foci. Table 4.1 lists the number of content foci and Level III CAPA science performance tasks for each content strand.

Table 4.1. Number of Content Foci and Performance Tasks for each CAPA Science Level III Content Strand

Content Strand	Number of Content Foci	Target Number of CAPA Performance Tasks
Life Science	9	2
Physical Science	4	2
Earth Science	8	2
Investigation & Experimentation	7	2
Total	28	8

Categorical Concurrence

Categorical concurrence measures the amount of overlap between the content strands and the assessment. For an alternate assessment such as the CAPA, Webb's target for adequate alignment is at least one performance task per content strand. As Table 4.2 displays, all four content strands for Level III met Webb's criterion. In Level III, panelists frequently linked performance tasks with more than one content focus. Although panelists rated only eight tasks, the mean number of tasks matched to each content strand totaled over 12, and the average number of tasks matched to each content strand exceeded the number of tasks targeted to measure it in the blueprint. Because alternate assessments have so few tasks, it is advantageous when tasks can be linked to multiple content strands.

Table 4.2. Categorical Concurrence for Level III: Mean Number of Performance Tasks per Strand

Title of Strand	Number of Tasks per Strand			At Least One Task per Strand
	Target # Items from blueprint	Mean Tasks Matched*	Standard Deviation	
Life Science	2	3.56	0.73	Y
Physical Science	2	2.11	0.60	Y
Earth Science	2	3.56	0.88	Y
Investigation & Experimentation	2	3.00	1.83	Y
Total	8	12.23		
Percent of strands with at least one task				100%

*Because panelists can link tasks to multiple strands, total may be greater than number of target items.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency determines whether the performance tasks demanded the same or higher level of cognitive complexity as the content foci they are designed to measure. As Table 4.3 indicates, all four content strands met Webb’s target of at least half the tasks meeting or exceeding the DOK level of their content foci. Although it meets Webb’s criterion, one possible area to monitor is the level of cognitive complexity for the Life Science tasks; panelists rated 97 percent of the Life Science tasks as being of greater complexity than the content foci they are designed to measure. California may wish to ensure that the Life Science tasks are not too complex compared with their content foci.

Table 4.3. DOK Consistency for Level III: Mean Percentage of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	SD	M	SD	M	SD	
Life Science	3.56	0	0.0	3	8.3	97	8.3	Y
Physical Science	2.11	26	35.5	26	25.2	48	19.4	Y
Earth Science	3.56	11	18.2	54	22.2	35	24.1	Y
Investigation & Experimentation	3.00	6	12.5	76	20.6	18	23.6	Y
Percent of strands with 50% of task DOK at or above objective DOK								100%

Range-of-Knowledge Correspondence

Range-of-knowledge correspondence measures the coverage of content foci within each content strand. Webb’s target criterion for adequate alignment is that at least half of the content foci within each content strand be linked with performance tasks. As Table 4.4 demonstrates, none of the content strands for Level III met this criterion. The Physical Science strand, with only four content foci, was weakly aligned,

nearly meeting the target. As in Level I, the failure of the Level III content strands to meet the range-of-knowledge criterion is likely attributable to the large number of content foci (28) the eight performance tasks are supposed to measure. Again, California may wish to examine the content foci for Level III CAPA students and determine if a smaller number of foci could be maintained as the core emphases of the curriculum and assessment. Alternatively, California could consider administering more performance tasks to obtain broader coverage of the content foci. Because the panelists already evaluated many of these tasks as measuring more than one content focus, it may be challenging to design more multi-dimensional tasks.

Table 4.4. Range-of-Knowledge for Level III: Mean Percentage of Content Foci per Strand Linked with Performance Tasks

Title of Strand	Number of Content Foci	Mean Tasks per Strand	Range of Content Foci			Range-of-Knowledge Target Met
			Content Foci with At Least One Task		% of Total Content Foci per Strand	
			M	SD	M	
Life Science	9	3.56	3.56	0.73	40	N
Physical Science	4	2.11	1.89	0.33	47	N
Earth Science	8	3.56	3.00	0.50	38	N
Investigation & Experimentation	7	3.00	1.75	0.96	25	N
Total	28	12.23				
Percentage of strands with 50% of Content Foci linked to at least one task						0%

Balance of Representation

Balance of representation explores the distribution of the performance tasks across the content foci to which the panelists linked them. Webb's criterion for adequate alignment is a balance index score of at least 70. All four Level III content strands met Webb's criterion for balance of representation (see Table 4.5). Although the range-of-knowledge results indicated inadequate coverage of the content foci within each content strand, balance of representation results show the performance tasks are evenly distributed across the content foci that were linked to those items.

Table 4.5. Balance of Representation for Level III: Mean Balance Index per Strand

Title of Strand	Balance of Representation						Balance Index Target Met
	Number of Content Foci	Mean Content Foci Linked with Tasks	Mean Tasks per Strand	Mean % of Tasks Linked to Strand	Mean Balance Index		
		M	M	M	M	SD	
Life Science	9	3.56	3.56	34	100	0.0	Y
Physical Science	4	1.89	2.11	20	96	7.3	Y
Earth Science	8	3.00	3.56	34	91	8.6	Y
Investigation & Experimentation	7	1.75	3.00	26	90	12.0	Y
Total	28	10.20	12.23				
Percentage of strands with a balance of representation index of 70 or greater							100%

Level III Summary

As in any alignment study, alignment of Level III of the CAPA science assessment demonstrated areas of strength and weakness. All four content strands met the criterion for adequate alignment for categorical concurrence, DOK consistency, and balance of representation. For range-of-knowledge correspondence, none of the four content strands met the target. To address this weakness, California may wish to consider the number of content foci targeted to Level III students or review the number or performance tasks administered. Table 4.6 summarizes alignment results for Level III of the CAPA science assessment across all four Webb alignment criteria.

Table 4.6. Summary of CAPA Science Level III Alignment Results

Content Strand	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Life Science	Y	Y	N	Y
Physical Science	Y	Y	N	Y
Earth Science	Y	Y	N	Y
Investigation & Experimentation	Y	Y	N	Y

Chapter 5: Level IV Results

This chapter presents complete alignment results for Level IV of the CAPA science assessment. Panelists rated all tasks for one complete version of the Level IV CAPA science test (Version 2), which contained eight operational performance tasks as well as four field test tasks. They also reviewed the remaining Level IV field test tasks in Version 3. This chapter contains results for the operational items only; descriptive results for the field test tasks can be found in Appendix C.

Level IV CAPA Science Strands and Assessment

Level IV of the CAPA is administered in grade eight to students with the most severe cognitive disabilities. It consists of eight performance tasks scored using a four-point rubric. Three content strands are assessed for the Level IV CAPA science assessment: Physical Science, Earth Science, and Investigation and Experimentation. Content foci classified under the Physical Science strand for this study are listed in the CAPA science blueprint under their respective sub-strands (Motion, Forces, Structure of Matter, Reactions, Periodic Table, and Density and Buoyancy). Content foci are drawn from grade eight, which focuses primarily on Physical Science. Across the three major content strands, the Level IV CAPA science test covers a total of 18 content foci. Table 5.1 lists the number of content foci and CAPA performance tasks for each content strand.

Table 5.1. Number of Content Foci and Performance Tasks for each Level IV CAPA Science Content Strand

Content Strand	Number of Content Foci	Target Number of CAPA Performance Tasks
Physical Science	10	6
Earth Science	3	1
Investigation & Experimentation	5	1
Total	18	8

Categorical Concurrence

Table 5.2 displays the categorical concurrence results for Level IV. All three content strands met Webb's target of at least one task matched to each content strand. Indeed, the number of tasks matched to each strand exceeds the targeted number from the CAPA science blueprints because panelists frequently indicated that performance tasks assessed more than one content focus.

Table 5.2. Categorical Concurrence for Level IV: Mean Number of Performance Tasks per Strand

Title of Strand	Number of Tasks per Strand			At Least One Task per Strand
	Target # Items from blueprint	Mean Tasks Matched*	Standard Deviation	
Physical Science**	6	6.22	0.44	Y
Earth Science	1	1.78	0.44	Y
Investigation & Experimentation	1	1.22	0.44	Y
	8	9.22		
Percent of strands with at least one task				100%

*Because panelists can link tasks to multiple strands, total may be greater than number of target items.

** In the CAPA science blueprints, Physical Science objectives are listed by sub-strand.

Depth-of-Knowledge Consistency

Results for DOK consistency are presented in Table 5.3. All three content strands met Webb’s alignment criterion of at least half the performance tasks meeting or exceeding the level of complexity of the content foci to which panelists matched them. As Table 5.3 demonstrates, the complexity of these tasks was well distributed. For instance, panelists judged that approximately half the tasks in every strand were at a higher level of complexity than the foci they measure.

Table 5.3. DOK Consistency for Level IV: Mean Percentage of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	SD	M	SD	M	SD	
Physical Science**	6.22	2	5.6	52	14.1	47	15.0	Y
Earth Science	1.78	0	0.0	44	52.7	56	52.7	Y
Investigation & Experimentation	1.22	17	35.4	28	44.1	56	52.7	Y
	9.22							
Percent of strands with 50% of item DOK at or above objective DOK								100%

** In the CAPA science blueprints, Physical Science objectives are listed by sub-strand.

Range-of-Knowledge Correspondence

Range-of-knowledge results are presented in Table 5.4. In Level IV, two of the three content strands (Physical Science and Earth Science) met Webb’s target of matching tasks to at least 50 percent of its content foci. The third strand, Investigation and Experimentation, did not meet the criterion. In the CAPA science blueprint, the Investigation and Experimentation strand includes five content foci but targets only one task, making it difficult to cover at least half of its content foci. To obtain more coverage

of this strand, California may wish to consider allotting it another task or designing an item for another strand that might also measure Investigation and Experimentation.

Table 5.4. Range-of-Knowledge for Level IV: Mean Percentage of Content Foci per Strand Linked with Performance Tasks

Title of Strand	Number of Content Foci	Mean Tasks per Strand	Range of Content Foci			Range-of-Knowledge Target Met
			Content Foci with At Least One Task		% of Total Content Foci per Strand	
			M	SD	M	
Physical Science**	10	6.22	6.22	0.44	62	Y
Earth Science	3	1.78	1.78	0.44	59	Y
Investigation & Experimentation	5	1.22	1.22	0.44	24	N
	18	9.22				
Percentage of strands with 50% of Content Foci linked to at least one task						67%

** In the CAPA science blueprints, Physical Science objectives are listed by sub-strand.

Balance of Representation

Table 5.5 displays the balance of representation results for Level IV. All three content strands have a perfect balance index score of 100, thereby meeting Webb’s target of at least a 70 on the balance index and indicating even distribution of tasks across the content foci to which they were linked.

Table 5.5. Balance of Representation for Level IV: Mean Balance Index per Strand

Title of Strand	Number of Content Foci	Balance-of-Knowledge Representation					Balance Index Target Met
		Mean Content Foci Linked with Tasks	Mean Tasks per Strand	Mean % of Tasks Linked to Strand	Mean Balance Index		
		M	M	M	M	SD	
Physical Science**	10	6.22	6.22	68	100	0.0	Y
Earth Science	3	1.78	1.78	19	100	0.0	Y
Investigation & Experimentation	5	1.22	1.22	13	100	0.0	Y
Total	18		9.22				
Percentage of strands with a balance of representation index of 70 or greater						100%	

** In the CAPA science blueprints, Physical Science objectives are listed by sub-strand.

Level IV Summary

Analyses of Level IV of the CAPA science assessment demonstrated areas of strength and weakness. All three content strands met the criterion for adequate alignment for categorical concurrence, DOK consistency, and balance of representation. For range-of-knowledge correspondence, two of the three content strands (Physical Science and Earth Science) met the target, and only Investigation and Experimentation did not. To address this minor weakness, California may wish to consider allotting

additional tasks to the Investigation and Experimentation strand or designing tasks for other content strands that also tap into Investigation and Experimentation. Table 5.6 summarizes alignment results for Level IV of the CAPA science assessment across all four Webb alignment criteria. As Table 5.6 demonstrates, Level IV generally has very strong alignment outcomes.

Table 5.6. Summary of CAPA Science Level IV Alignment Results

Content Strand	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Physical Science	Y	Y	Y	Y
Earth Science	Y	Y	Y	Y
Investigation & Experimentation	Y	Y	N	Y

Chapter 6: Level V Results

This chapter presents complete alignment results for Level V of the CAPA science assessment. Panelists rated all tasks for one complete version of the Level V CAPA science test (Version 2), which contained all eight operational performance tasks as well as four field test tasks. They also reviewed the remaining Level V field test tasks in Version 3. This chapter contains results for the operational items only; descriptive results for the field test tasks can be found in Appendix C.

Level V CAPA Science Strands and Assessment

Level V of the CAPA is administered in Grade 10 to students with the most severe cognitive disabilities. It consists of eight performance tasks, scored using a four-point rubric. Five content strands are assessed on the Level V CAPA science test: Earth Science, Investigation and Experimentation, Biology, Physics, and Chemistry. Content foci are drawn from the general science strands for high school science. The Level V CAPA science exam tests 29 content foci across the five major content strands. Table 6.1 lists the number of content foci and CAPA performance tasks for each content strand.

Table 6.1. Number of Content Foci and Performance Tasks for each Level V CAPA Science Content Strand

Content Strand	Number of Content Foci	Target Number of CAPA Performance Tasks
Earth Science	7	2
Investigation & Experimentation	5	1
Biology	12	3
Physics	2	1
Chemistry	3	1
Total	29	8

Categorical Concurrence

Level V results for categorical concurrence are presented in Table 6.2. All five content strands met Webb's target of being matched to at least one performance task. These results indicate the assessment and the blueprint are covering the same basic content areas. Additionally, as Table 6.2 indicates, the number of tasks panelists linked to each strand closely matches the number of tasks the CAPA science blueprints targeted for the strand.

Table 6.2. Categorical Concurrence for Level V: Mean Number of Performance Tasks per Strand

Title of Strand	Number of Tasks per Strand			At Least One Task per Strand
	Target # Items from blueprint	Mean Tasks Matched*	Standard Deviation	
Earth Science	2	2.44	0.73	Y
Investigation & Experimentation	1	1.00	0.00	Y
Biology	3	3.33	0.50	Y
Physics	1	1.00	0.00	Y
Chemistry	1	1.00	0.00	Y
Total	8	8.77		
Percent of strands with at least one task				100%

*Because panelists can link tasks to multiple strands, total may be greater than number of target items.

Depth-of-Knowledge Consistency

Table 6.3 displays DOK consistency results for Level V. All five content strands met Webb's criterion of at least half the tasks meeting or exceeding the level of complexity of their respective content foci. As Table 6.3 indicates, the distribution of task complexity relative to the complexity of the content foci is well distributed within each of the content strands.

Table 6.3. DOK Consistency for Level V: Mean Percentage of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	SD	M	SD	M	SD	
Earth Science	2.44	17	25.0	19	28.2	65	26.9	Y
Investigation & Experimentation	1.00	22	44.1	44	52.7	33	50.0	Y
Biology	3.33	7	14.7	57	32.4	35	33.8	Y
Physics	1.00	11	33.3	44	52.7	44	52.7	Y
Chemistry	1.00	22	44.1	67	50.0	11	33.3	Y
8.77								
Percent of strands with 50% of task DOK at or above objective DOK								100%

Range-of-Knowledge Correspondence

Range-of-knowledge correspondence results for Level V are presented in Table 6.4. Only one of the five content strands (Physics) met Webb's target of at least 50 percent of the content foci within a strand being matched with a task. It is not surprising that Physics was the one strand to meet the target since it has the fewest content foci to assess. The range-of-knowledge issue at Level V is the same as at other CAPA levels—the test attempts to assess a large number of content foci with a limited number

of performance tasks. The Level V CAPA has the most content foci, at 29, but still administers only eight performance tasks to measure all foci. To address weaknesses in this area, California should review whether all content foci are truly essential for high school students taking the CAPA. Alternately, California might wish to consider whether it would be appropriate to administer more tasks to the Level V CAPA test-takers.

Table 6.4. Range-of-Knowledge for Level V: Mean Percentage of Content Foci per Strand Linked with Performance Tasks

Title of Strand	Number of Content Foci	Mean Tasks per Strand	Range of Content Foci		Range-of-Knowledge Target Met	
			Content Foci with At Least One Task	% of Total Content Foci per Strand		
			M	SD		M
Earth Science	7	2.44	2.22	0.44	32	N
Investigation & Experimentation	5	1.00	1.00	0.00	20	N
Biology	12	3.33	3.33	0.50	28	N
Physics	2	1.00	1.00	0.00	50	Y
Chemistry	3	1.00	1.00	0.00	33	N
Total	29	8.77				
Percentage of strands with 50% of Content Foci linked to at least one task						20%

Balance of Representation

Table 6.5 displays balance of representation results for Level V. All five content strands met Webb’s criterion of a balance index score of at least 70. Further, all but one content strand (Earth Science) had a perfect balance index score. These results indicate performance tasks are evenly distributed across the content foci to which they were linked.

Table 6.5. Balance of Representation for Level V: Mean Balance Index per Strand

Title of Strand	Number of Content Foci	Balance of Representation					Balance Index Target Met
		Mean Content Foci Linked with Tasks	Mean Tasks per Strand	Mean % of Tasks Linked to Strand	Mean Balance Index		
		M	M	M	M	SD	
Earth Science	7	2.22	2.44	27	96	7.3	Y
Investigation & Experimentation	5	1.00	1.00	12	100	0.0	Y
Biology	12	3.33	3.33	38	100	0.0	Y
Physics	2	1.00	1.00	12	100	0.0	Y
Chemistry	3	1.00	1.00	12	100	0.0	Y
Total	29		8.77				
Percentage of strands with a balance of representation index of 70 or greater						100%	

Level V Summary

Areas of strength and weakness emerged in alignment of CAPA science Level V. All five content strands met the criteria for adequate alignment for categorical concurrence, DOK consistency, and balance of representation. For range-of-knowledge correspondence, only one of the five content strands (Physics) met the target. To address this weakness, California may wish to consider the number of content foci targeted to Level V students or review the number of performance tasks administered. Table 6.6 summarizes alignment results for Level V of the CAPA science assessment across all four Webb alignment criteria. As Table 6.6 demonstrates, the only considerable area of weakness for Level V was range-of-knowledge correspondence, or sufficient coverage of the content foci within the strand.

Table 6.6. Summary of CAPA Science Level V Alignment Results

Content Strand	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Earth Science	Y	Y	N	Y
Investigation & Experimentation	Y	Y	N	Y
Biology	Y	Y	N	Y
Physics	Y	Y	Y	Y
Chemistry	Y	Y	N	Y

Chapter 7: Summary and Recommendations

This chapter summarizes alignment results for all four levels of the CAPA science test. For each level, results for all Webb criteria are classified into the following categories:

- Fully aligned—assessments align to all content strands (100%);
- Highly aligned—assessments align to the majority of strands (70–99%);
- Partially aligned—assessments align well to some strands (50–69%); and
- Weakly aligned—assessments align to less than half the strands (below 50%).

For each level, Table 7.1 presents the alignment category and, below in parentheses, the number of content strands meeting Webb’s targets. Partially and weakly aligned categories are highlighted in yellow and red, respectively, to highlight areas that need review.

Table 7.1. Summary of Webb Alignment Analyses across All Levels of the CAPA Science Assessment

CAPA Level	Categorical Concurrence	DOK Consistency	Range-of-Knowledge Correspondence	Balance of Representation
Level I	Fully aligned (4 of 4)	Highly Aligned (3 of 4)	Weakly aligned (1 of 4)	Fully aligned (4 of 4)
Level III	Fully aligned (4 of 4)	Fully aligned (4 of 4)	Weakly aligned (0 of 4)	Fully aligned (4 of 4)
Level IV	Fully aligned (3 of 3)	Fully aligned (3 of 3)	Partially aligned (2 of 3)	Fully aligned (3 of 3)
Level V	Fully aligned (5 of 5)	Fully aligned (5 of 5)	Weakly aligned (1 of 5)	Fully aligned (5 of 5)

As Table 7.1 indicates, all four levels of the CAPA science test were highly or fully aligned across all Webb criteria except range-of-knowledge correspondence, where Level IV was partially aligned and Levels I, III, and V were weakly aligned. As noted throughout this report, this weak alignment likely arises because of the large number of content foci for each level of the CAPA compared with the number of performance tasks. To address this weakness, California may wish to review the number of content foci for each level. Should California judge all the content foci to be essential, it may wish to consider increasing the number of performance tasks to enable broader coverage of the content strands on the assessment.

Results from the categorical concurrence, DOK consistency, and balance of representation analyses indicate the alignment between the CAPA science content strands and the test is generally strong across all levels—tasks are covering the right general content, are appropriately complex, and are balanced across content foci. If

California addresses the lone weakness of range-of-knowledge correspondence by considering the number of content foci or performance tasks, alignment of the CAPA science assessment will be quite strong.

References

- Almond, P., Filbin, J., Hall, J. and Tindal, J. (2005, June). *Alignment to Strands: Alternate Assessments, Expanded Benchmarks, Universal Design for Learning and Individualized Education Programs*. Paper presented at the meeting of the Council of Chief State School Officers, San Antonio, TX.
- Browder, D. M., Flowers, C., Ahlgrim-Dezell, L., Karvonen, M., Spooner, F., & Algozzine, R. (2004). The alignment of alternate assessment content to academic and functional curricula. *Journal of Special Education, 37*, 211-223.
- Roach, A. T., Elliott, S. N., & Webb, N. L. (2005). Alignment of an alternate assessment with state academic strands: Evidence content validity of the Wisconsin Alternate Assessment. *Journal of Special Education, 38*, 218–231.
- Taylor, L. R., Wise, L. L., Thacker, A. A., Moody, R. L., Koger, L. E., Dickinson, E. R., et al. (2007). *Independent evaluation of the California Strands Tests (CSTs) and the California Alternate Performance Assessment (CAPA)*. (TR-07-04). Alexandria, VA: Human Resources Research Organization.
- Tindal, G. (2005). *Alignment of alternate assessments using the Webb system: An abbreviated version*. Paper presented at the meeting of the Council of Chief State School Officers, San Antonio, TX.
- U.S. Department of Education. (2005, August). *Alternate achievement standards for students with the most significant cognitive disabilities*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved from <http://www.ed.gov/policy/elsec/guid/altguidance.doc>.
- Webb, N. L. (1997). *Research Monograph No. 6: Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, D.C.: Council of Chief State Schools Officers.
- Webb, N. L. (1999). *Alignment of science and mathematics strands and assessments in four states*. (Research Monograph 18). Madison, WI: National Institute for Science Education and Council of Chief State School Officers. (ERIC Document Reproduction Service No. ED440852)
- Webb, N. L. (2005). *Webb alignment tool: Training manual*. Madison, WI: Wisconsin Center for Education Research. Available: <http://www.wcer.wisc.edu/WAT/index.aspx>.

Appendix A

Workshop Materials

This Appendix contains examples of the materials panelists used during the alignment workshop. Some are truncated for space constraints, and formatting has been slightly modified to satisfy requirements for this report. Otherwise, materials are presented as the panelists saw them.

Depth-of-Knowledge Level Descriptions

DOK Level 1

Requires students to recall or observe facts, definitions, terms. Involves simple one-step procedures. May involve simple calculations.

Keywords: identify, recall, recognize, measure

DOK Level 2

This level includes the engagement of some mental processing beyond a habitual response. The item requires students to make some decisions as to how to approach a problem or activity.

Keywords: classify, organize, estimate, make observations, collect and display data, compare data

DOK Level 3

A multiple step 'behavioral event' is executed in more than one context. Requires reasoning, planning, or use of evidence to solve problem. May involve activity with more than one possible answer. Requires conjecture or restructuring of problems. Involves drawing conclusions from observations, citing evidence and developing logical arguments for concepts. Uses concepts to solve non-routine problems.

DOK Level 4

The 'behavioral event' reflects an approach (of many) to completing the task. May require complex reasoning, planning, developing, and thinking. Typically requires extended time to complete problem, but time spent not on repetitive tasks. Requires students to make several connections and apply one approach among many to solve the problem. Involves complex restructuring of data, establishing and evaluating criteria to solve problems.

CAPA Science Strands DOK Rating Sheets

CALIFORNIA CONTENT STRANDS		Code
Physical Science:	Number of Tasks: 3 Percentage of Test: 37.5%	
Kindergarten		
1. Properties of materials can be observed, measured, and predicted. As a basis for understanding this concept:		
1.a Students know objects can be described in terms of the materials they are made of (e.g., clay, cloth, paper) and their physical properties (e.g., color, size, shape, weight, texture, flexibility, attraction to magnets, floating, sinking).		
<ul style="list-style-type: none"> ✓ Identify color of object. ✓ Identify size of object. ✓ Identify texture of object. 		
1.b Students know water can be a liquid or a solid and can be made to change back and forth from one form to the other.		
<ul style="list-style-type: none"> ✓ Identify ice. ✓ Identify water. 		
Grade 2		
1. The motion of objects can be observed and measured. As a basis for understanding this concept:		
1.c Students know the way to change how something is moving is by giving it a push or a pull. The size of the change is related to the strength, or the amount of force of the push or pull.		
<ul style="list-style-type: none"> ✓ Pull an object/switch. ✓ Push an object/switch. 		
1.e Students know objects fall to the ground unless something holds them up.		
<ul style="list-style-type: none"> ✓ Explore gravity by causing different objects to fall (e.g., feather, balloon, ball, etc.). ✓ Hold object and release upon request. 		
Life Science:	Number of Tasks: 2 Percentage of Test: 25%	
Kindergarten		
2. Different types of plants and animals inhabit the earth. As a basis for understanding this concept:		
2.c Students know how to identify major structures of common plants and animals (e.g., stems, leaves, roots, arms, wings, legs).		
<ul style="list-style-type: none"> ✓ Identify body parts on self. ✓ Identify animal body parts. 		

CAPA Science Performance Task Rating Sheets

Name:			LEVEL: I		Content Area: CAPA SCIENCE		
Test Version	Item Number	Depth Of Knowledge	Content Strand/Objective 1	Content Strand/Objective 2	Overall Alignment	Overall Item Quality	Comments
	(Number Listed in Test Form)	(Enter Level 1 to 4)	(Enter Strand ID Code)	(Enter Strand ID Code)	(Enter Scale of 1 to 4)	(Enter Scale of 1 to 4)	
2	1						
2	2						
2	3						
2	4						
2	5						
2	6						
2	7						
2	8						
2	9						
2	10						
2	11						
2	12						
3	2						
3	5						
3	8						
3	11						

CAPA Science Coding Sheet

Depth of Knowledge

See Depth-of-Knowledge Level Descriptions handout for full descriptions

- 1 Recall
- 2 Skills/Concepts
- 3 Strategic Thinking
- 4 Extended Thinking

Content Strand/Objective

5-digit HumRRO code assigned to appropriate CAPA Science content strand

Overall Alignment

- 1 Not aligned to any California alternate content strand (Use **ONLY** if you did **not** assign a strand to the item)
- 2 Weakly aligned to this California alternate content strand – Not a very good example of the strands.
- 3 Highly aligned to this California alternate content strand – Good and reasonable example of the strands.
- 4 Fully aligned to the California alternate content strands – Exemplary item, clear example of strand for which it is matched.

Overall Item Quality

- 1 Item is of poor overall quality (Rating requires annotation).
- 2 Item is of good quality, but has some easily repairable flaw (Rating requires annotation).
- 3 Item is of good quality, typical of what you would expect on this and similar tests.
- 4 Item is of exceptional quality (annotations encouraged).

Appendix B

Performance Task Quality

As an ancillary task to the alignment study, panelists were given an opportunity to rate the quality of the CAPA science performance tasks. Panelists also commented on item quality, particularly if they thought tasks were problematic. Results of task quality ratings for the CAPA science operational tasks are provided in this Appendix; task quality ratings for the field test tasks are briefly addressed in Appendix C.

Panelists rated performance tasks on a scale of 1 to 4 (see Table B.1), with 1 indicating poor overall quality and 4 indicating exceptional quality. We instructed panelists to comment on any tasks rated a 1 or 2, and encouraged them to comment on tasks rated a 4. Because item quality ratings were an ancillary task and not the emphasis of the alignment workshop, we encouraged panelists to comment sparingly and only address item quality if an item appeared to be truly problematic or exceptional.

Table B.1. Item Quality Rating Codes

Rating	Description
1	Item is of poor overall quality.
2	Item is of good quality, but has some easily repairable flaw.
3	Item is of good quality, typical of what you would expect on this and similar tests.
4	Item is of exceptional quality.

Table B.2 displays panelists' item quality ratings across all operational performance tasks. Panelists each rated eight tasks at each of the four levels of the CAPA, resulting in a total of 288 item quality ratings for the nine panelists. The smallest number of comments made by an individual rater across all levels was 2, while the largest number of comments made by an individual rater was 16. Panelists provided more comments for Level I (n = 30) and Level III (n = 23) than Level IV (n = 10) or Level V (n = 4). As Table B.2 demonstrates, panelists rated a total of 230 of the 288 tasks (80%) as quality level three or level four, indicating good or exceptional task quality. The majority (71%) of task ratings were "3," indicating good quality. Average ratings were similar across all four levels of the CAPA.

Table B.2. Distribution of Panelists' Task Quality Ratings

Quality Rating	Number of Tasks
1 Poor	6
2 Needs Improvement	52
3 Good	204
4 Exceptional	26
Total	288

Panelists rated approximately one-fifth of the tasks as poor or in need of improvement. However, all of the "1" ratings and a quarter of the "2" ratings were assigned by the same rater, suggesting one rater might have been particularly harsh.

To better understand the reasons for these less favorable quality ratings, we conducted a brief content analysis of raters' comments. Three themes emerged from the content analysis: unclear task instructions, concerns about task content, and challenges unique to test takers with physical disabilities and/or visual impairment.

The first theme, unclear task instructions, seemed to apply mostly to Levels I and III. Panelists expressed concern that test-takers would not understand how to perform specific assessment tasks when prompted. These comments suggested examiners' cues should be more explicit about what action the test-taker must take to complete the task properly. The second theme was concerns about task content. Panelists across all CAPA levels noted issues with task content. Because task content is confidential, we provided content-related item feedback directly to the contractor. Generally, examples included poor stimulus cards, irrelevant items and technicalities that may unduly impact students' ability to correctly respond to performance tasks. The third theme was concern about the appropriateness of specific tasks for test-takers with physical disabilities and/or visual impairment. These tasks require the test taker to move or visually examine a stimulus in a manner that would make the tasks difficult, if not impossible, for students with these disability types to complete.

Appendix C

Field Test Tasks

This Appendix presents basic results for the CAPA Science field test tasks. Panelists reviewed all eight field test tasks at each level of the CAPA science test. HumRRO staff did not tell panelists which tasks were field test tasks and which were operational tasks. These tasks were distributed across Version 2 (four items) and Version 3 (four items) of the CAPA science test. Because they are not designed to act as an intact test form, it would be illogical to conduct a true alignment study of the field test tasks. However, the same types of information examined in an alignment study can be examined in a descriptive way for the field test tasks. The first part of this Appendix looks at alignment-type data, such as depth-of-knowledge, benchmarks to which tasks are linked, and overall level of alignment. The second part of the Appendix presents task quality ratings and panelist feedback on the field test tasks; this feedback may be of particular interest since the field test tasks are still in the development stage.

Descriptive Alignment-Type Data

We explored alignment-type data for the field test tasks in a purely descriptive manner but did not compute any of the Webb alignment statistics because the tasks are not intended to act together as a test form. Although these data are less interpretable than the actual alignment study data, they provide a preliminary picture of the field test tasks. We provide descriptive summaries of task DOK ratings and distribution of linkage across content strands for each level. Then, to provide a picture of the degree of alignment in lieu of actual alignment ratings, we offer a brief summary of panelists' ratings of the extent of alignment for these tasks.

First, we computed frequencies of task DOK ratings for each level. Although we did not compare these ratings with the content strands to which the tasks were linked, this descriptive analysis gives an overall picture of the complexity level to which most of the field test tasks are targeted. Table C.1 displays task DOK rating distributions. Few of the field test tasks were rated as DOK level 4. The most common rating was 2 (36%), followed closely by 3 (33%) and 1 (26%). Because the CAPA is for students with the most severe cognitive disabilities, this distribution of DOK complexity ratings is likely appropriate. Additionally, panelists rated very few of the content foci at DOK level 4.

Table C.1. Distribution of Field Test Task DOK Ratings across CAPA Science Levels

CAPA Level	DOK Rating			
	1	2	3	4
I	44	22	6	0
III	7	41	22	2
IV	2	13	47	10
V	22	29	21	0
Total	75	105	96	12

Next, we explored the content strands to which panelists linked the field test tasks. This descriptive analysis should provide California with a general picture of the content strands the panelists believe the field test tasks are designed to measure. If panelists' ratings are not in line with expectations, California can revise tasks or create additional tasks to address areas of need.

Table C.2 presents the number of tasks linked to each general content strand for each CAPA science level. For Level I, panelists linked most field test tasks with Physical Science or Life Science, followed by Investigation and Experimentation. Panelists did not match any Level I tasks with Earth Science. At Level III, panelists matched most tasks with Life Science or Physical Science, although panelists also linked tasks to Earth Science and Investigation and Experimentation. At Level IV, panelists matched most tasks to Physical Science, followed by Investigation and Experimentation and then Earth Science. Finally, at Level V, panelists linked the majority of tasks with Biology, followed by Chemistry, Earth Science and Physics. Only one panelist linked a task with Investigation and Experimentation at Level V. Note that Table C.2 displays the number of times a panelist matched a task to the content strand, not the number of tasks matched. For instance, for Life Science at Level I, there were 27 linkages, which likely implies that all nine panelists linked 3 tasks with this strand—not that 27 tasks were matched to the strand.

Table C.2. Distribution of Field Test Task Linkages across CAPA Science Levels

Content Strand	CAPA Level*			
	I	III	IV	V
Life Science	27	36	NA	NA
Physical Science	32	36	42	NA
Earth Science	0	9	11	10
Investigation & Experimentation	16	9	28	1
Biology	NA	NA	NA	40
Physics	NA	NA	NA	9
Chemistry	NA	NA	NA	17

*NA indicates the content strand is not assessed at this CAPA level.

Finally, to obtain a picture of the overall alignment of these tasks with the content strands to which panelists matched them, we asked panelists to rate their degree of alignment. Panelists used a four-point scale to rate the amount of overall alignment for each task. In this scale, a score of 1 indicated the task was not aligned to an alternate content strand; 2 indicated weak alignment; 3 indicated high alignment; and 4 indicated full alignment. Table C.3 presents the distribution of panelists' alignment ratings across CAPA levels. Panelists rated the degree of alignment for most items at 3 or 4, indicating most (92%) field test tasks are highly or fully aligned to a content focus from the California alternate content expectations. For those items rated a 1 or 2, California may wish to examine how to link them more strongly to the content focus they were designed to measure.

Table C.3. Distribution of Field Test Task Alignment Ratings across CAPA Science Levels

CAPA Level	Alignment Rating			
	1	2	3	4
I	1	7	49	15
III	0	6	54	12
IV	0	7	58	7
V	0	2	60	10
Total	1	22	221	44

Task Quality Feedback and Panelist Comments

Task quality ratings for the field test tasks were strikingly similar to the quality ratings for the operational tasks. Table C.4 presents the distribution of panelists' quality ratings for the field test tasks, ranging from poor (1) to exceptional (4). The most frequent (75%) quality rating was 3, indicating good item quality. Approximately 17 percent of tasks were rated 1 or 2, indicating these items are of poor quality or are in need of improvement. However, one-third of these less positive ratings came from one panelist. Thus, to some extent, quality ratings might reflect the harsh rating tendencies of one rater. Qualitative exploration of rater comments reveals similar patterns to the operational tasks. The same three themes emerged: unclear task instructions, concerns about task content, and challenges unique to test takers with physical disabilities and/or visual impairment. California may wish to consider panelists' comments in completing development of these field test tasks.

Table C.4. Distribution of Panelists' Field Test Task Quality Ratings

Quality Rating	Number of Tasks
1 Poor	10
2 Needs Improvement	38
3 Good	216
4 Exceptional	24
Total	288