Initial Science Stakeholder Meetings and Online Survey Report

Contract #5417

Initial report on the 2014 CAASPP Science Stakeholder Meetings and online survey regarding recommendations for the new California science assessment aligned to the Next Generation Science Standards.

Prepared for the California Department of Education by Educational Testing Service

Final Presented October 17, 2014

Listening. Learning. Leading.®

Table of Contents

Section 1: Executive Summary....1
1A. Overview and Background....1
1B. Findings....1
Section 2: Introduction and Background....3
2A. Historical Context of the Next Generation Science Standards (NGSS)....3
2B. NGSS Architecture....4
2C. Legislation....4
Section 3: Methodology....6
3A. Stakeholder Recruiting Processes....6
3B. Meeting Processes....8
3C. Stakeholders....10
3D. Methods Used to Analyze the Data....10
Section 4: Discussion and Feedback Pursuant to *Education Code* (*EC*) 60640(b)....11
4A. Grade Level....12
4B. Type of Assessment....19
Section 5: Discussion and Feedback Pursuant to *EC* Section 60640(c)....21
5A. Grade Levels....21
5B. Type of Assessment....27
5C. NGSS Consortium–Developed Assessments....28
5D. Various Item Types....29
5E. Online Testing....29
Section 6: Results from CAASPP Science Stakeholders Meetings....30
6A. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(b)....30
6B. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of Non–ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(c)....32
6C. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of Alternate NGSS Assessments Implemented Beyond ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(c)....33
Section 7: Results from the Online Survey....34
7A. Survey Background....34
7B. Characteristics of Survey Respondents....35
7C. Summary of Part 1 Responses on ESEA-mandated CAASPP Assessments....38
7D. Summary of Part 2 Responses on Additional CAASPP Assessments....51
7E. Summary of Part 3 Responses on Measurement Considerations....56
7F. Summary of Part 4 Responses on Overall Feedback on the Future Science Assessment System....61
Section 8: Suggestions for Interpretation and Development of Recommendations....63
8A. Suggestions for Federally Mandated ESEA Testing....63
8B. Suggestions for Non-ESEA Testing....64
8C. Suggestions for Administering Alternate Assessments....64
8D. Conclusion....64
Appendix A: Organizations Contacted for Participant Recruitment....65
Appendix B: Transcript of the Participant Application....66
Appendix C: Recommendations Outside the Scope of the Meetings....77
Appendix D: General Session PowerPoint....78
Appendix E: Group Session PowerPoint and Handouts....79
Appendix F: Group Discussion Questions....80
Appendix G: NGSS Architecture....81
Appendix H: Acronyms, Initialisms, and Definitions....82
Appendix I: Transcript of Online Survey....89
Appendix J: Summary of Responses to Part 1 (ESEA Assessments) of Online Survey for All Grade Levels....114
Appendix K: Summary of Science Stakeholder Meeting Evaluations....116
Appendix L: Codes for Describing Online Survey Responses....118
References....124

List of Tables

Table 3.1 Target Representation of Stakeholder Groups 6
Table 3.2 Stakeholder Groups Represented at the Meetings....6
Table 4.1 Summary of Stakeholder Recommendations for Grade Level, Grade Span Three to Five 12
Table 4.2 Summary of Stakeholder Recommendations for Content Area, Grade Span Three to Five 14
Table 4.3 Summary of Stakeholder Recommendations for Grade Level, Grade Span Six to Nine 15
Table 4.4 Summary of Stakeholder Recommendations for Content Area, Grade Span Six to Nine....16
Table 4.5 Summary of Stakeholder Recommendations for Grade Level, Grade Span Ten to Twelve 17
Table 4.6 Summary of Stakeholder Recommendations for Content Area, Grade Span Ten to Twelve....18
Table 4.7 Summary of Stakeholder Recommendations for Assessment Options, ESEA-mandated Tests, All Grade Spans 19
Table 4.8 Summary of Stakeholder Recommendations for Assessment Types, ESEA-mandated Tests 20
Table 5.1 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Three to Five 21
Table 5.2 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Three to Five 22
Table 5.3 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Six to Nine....23
Table 5.4 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Six to Nine 24
Table 5.5 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Ten to Twelve 25
Table 5.6 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Ten to Twelve 26
Table 5.7 Summary of Stakeholder Recommendations for Assessment Options, Non–ESEA-mandated Tests....28
Table 5.8 Summary of Stakeholder Recommendations for Assessment Types, Non–ESEA-mandated Tests 28
Table 6.1 Summary of Stakeholder Recommendations for Grade-level Assessments....31
Table 6.2 Summary of Stakeholder Recommendations for Additional Assessments 32
Table 7.1 Breakdown of Primary Stakeholder Roles of Survey Respondents 35
Table 7.2 Primary Science Stakeholder Role of Survey Respondents by Gender 36
Table 7.3 Primary Science Stakeholder Role of Survey Respondents by Ethnic Background 37
Table 7.4 Preferences of Content Domains for Respondents Who Selected Grade Five 40
Table 7.5 Preferences of Test Types for Respondents Who Selected Grade Five 41
Table 7.6 Preferences of Item Types for Respondents Who Selected Grade Five....42
Table 7.7 Preferences of Content Domains for Respondents Who Selected Grade Eight 44
Table 7.8 Preferences of Test Types for Respondents Who Selected Grade Eight....46
Table 7.9 Preferences of Item Types for Respondents Who Selected Grade Eight 46
Table 7.10 Preferences of Content Domains for Respondents Who Selected Grade Eleven 48
Table 7.11 Preferences of Test Types for Respondents Who Selected Grade Eleven 49
Table 7.12 Preferences of Item Types for Respondents Who Selected Grade Eleven 50
Table 7.13 Summary of Selections of Non-ESEA Grade-level Tests....51
Table 7.14 Summary of Selections for How Content Should Be Assessed in Non-ESEA Tests....52
Table 7.15 Summary of Selections for What Content Domain(s) Should Be Assessed in Non-ESEA Tests 53
Table 7.16 Summary of Selections for Which Type(s) of Assessments Should Be Administered in Each Non-ESEA Grade-level Test 54
Table 7.17 Summary of Selections of Grade Levels to Test Students with Severe Cognitive Disabilities for Non-ESEA Purposes....55
Table 7.18 Summary of Selections for How Content Should Be Assessed for Students with Severe Cognitive Disabilities on Additional, Non-ESEA Grade-level Tests 56
Table 7.19 Summary of Selections of Important Considerations for the Future CAASPP Science Assessments 61
Table J.1 Summary of Which Science Content Domain(s) Should Be Targeted for Assessment in the Selected ESEA Grade Test....114
Table J.2 Summary of Which Type(s) of Assessments Should Be Available for Administration in the Selected ESEA Grade Test....114
Table J.3 Summary of Which Item Type(s) Should Be Administered in the Selected ESEA Grade Test 115
Table K.1 Summary of Ratings for Aspects of the Science Stakeholder Meetings 116
Table K.2 Summary of Feedback Evaluations of the Science Stakeholder Meetings....116

List of Figures

Figure 2.1 Development of the NGSS: Building on the Past; Preparing for the Future....3
Figure 3.1 Stakeholder Meeting Hierarchy....8
Figure 7.1 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Three to Five Span....38
Figure 7.2 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Six to Nine Span....43
Figure 7.3 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Ten to Twelve Span....47
Figure 7.4 Barplots Showing Breakdown to Responses to Survey Questions on Using Matrix Sampling in CAASPP Science Assessments (Panel A) or Population Sampling (Panel B)....57
Figure 7.5 Barplot Showing Breakdown of Preferred Scoring Choice for Open-ended Test Items....59
Figure G.1 Example of How to Analyze an NGSS Box....81
Figure H.1 Example of a Census Administration....84
Figure H.2 Example of Matrix Sampling....86
Figure H.3 Example of a Two-stage CAT....87
Figure H.4 Example of Partial Matrix Sampling....87
Figure H.5 Example of Population Sampling....87

Section 1: Executive Summary

1A. Overview and Background

California *Education Code (EC)* Section 60640 set forth the requirement that the State Superintendent of Public Instruction (SSPI) provide the Legislature with recommendations, including the grade level, content, type of assessment, and a timeline for implementation, for the development of an assessment aligned to the Next Generation Science Standards (NGSS) adopted pursuant to *EC* Section 60605.85. In developing the recommendations, the SSPI was required to consult with specific science stakeholders and consider the inclusion of a variety of specific features in the new science assessment system.

In two meetings hosted on behalf of the California Department of Education (CDE) and conducted by Educational Testing Service (ETS) in Sacramento, California, from July 15–18, 2014, 130 science stakeholders from across California provided input regarding what a new California science assessment system aligned to the NGSS should look like. Additionally, an online survey was administered in August 2014 to meeting participants, applicants who did not attend the meetings, and stakeholder organizations. This report summarizes the results from these meetings and the survey.

Section 2 of this report provides background on the NGSS and related state legislation and federal requirements that led to these stakeholder meetings. Section 3 outlines the overall meeting design methodologies used, including the participant recruiting process that was undertaken and the participation targets and final counts for various stakeholder groups. The recommendations and rationales for each stakeholder group regarding the assessments needed to meet the requirements of California *EC* Section 60640(b) are described in Section 4. Section 5 provides recommendations and rationales for each stakeholder group regarding additional assessments beyond those recommended for *EC* Section 60640(b). Overall summaries of the groups' recommendations are presented in Section 6. Section 7 provides the results from the online survey, and Section 8 synthesizes the individual recommendations collected in the survey and the stakeholder group recommendations collected at the in-person meetings.

1B. Findings

For the federally mandated (Elementary and Secondary Education Act [ESEA]) testing in science for the three grade spans—grades three to five, six to nine, and ten to twelve—meeting groups and survey respondents most frequently recommended grade levels of five, eight, and eleven within each grade span, respectively.

In general, stakeholder groups at the meetings and individual survey respondents both preferred computer-delivered assessments over paper-pencil tests. Specifically, the meeting groups showed a strong preference for computer-adaptive testing for providing potentially shorter tests and more precise scores. To best assess the three dimensions of the NGSS—science and engineering practices, crosscutting concepts, and disciplinary core ideas—meeting groups generally favored performance-based "hands-on" and "virtual" tasks with limited use of discrete multiple-choice items. Survey respondents also expressed an interest in such performance-based tasks and de-emphasized including items that only require memorization of facts.

Overall, California science stakeholder meeting groups and individual survey respondents often expressed similar preferences for a new California science assessment system. In addition, the meeting discussions and survey respondent rationales typically touched on several of the same underlying reasons for particular preferences. Given that only 18 percent (74 out of 422) of the survey respondents also attended one of the meetings, the common recommendations from these two events are not simply due to shared experiences but, rather, reflect the primary considerations and values of a large portion of the California science stakeholder community.

Appendixes in this report contain the following:

Appendix A	List of stakeholder organizations that were contacted to recruit meeting participants
Appendix B	Transcript of the online science stakeholder application for meeting participation
Appendix C	List of recommendations from meeting participants that were beyond the scope of the meetings
Appendix D	PowerPoint slides presented at the general session of each meeting
Appendix E	PowerPoint slides and handouts presented in each group session
Appendix F	Group discussion questions
Appendix G	Documents describing the NGSS architecture
Appendix H	Acronyms, initialisms, and definitions of terms
Appendix I	Transcript of the online survey
Appendix J	Summaries of responses to Part 1 of the online survey for all grade levels
Appendix K	Summary of Science Stakeholder Meeting evaluations submitted by meeting participants
Appendix L	Codes describing online survey responses

Section 2: Introduction and Background

2A. Historical Context of the Next Generation Science Standards (NGSS)

Development and Adoption

The adoption of the NGSS in California was preceded by several important development phases at both the national and state levels. Figure 2.1 (Pruitt, 2013) illustrates a brief overview of the NGSS development at the national level, which began when 26 states and the National Research Council (NRC) worked with Achieve, Inc., to develop the NGSS.[1] It shows the historical development of NGSS and the founding research conducted by NRC and America's Lab Report.

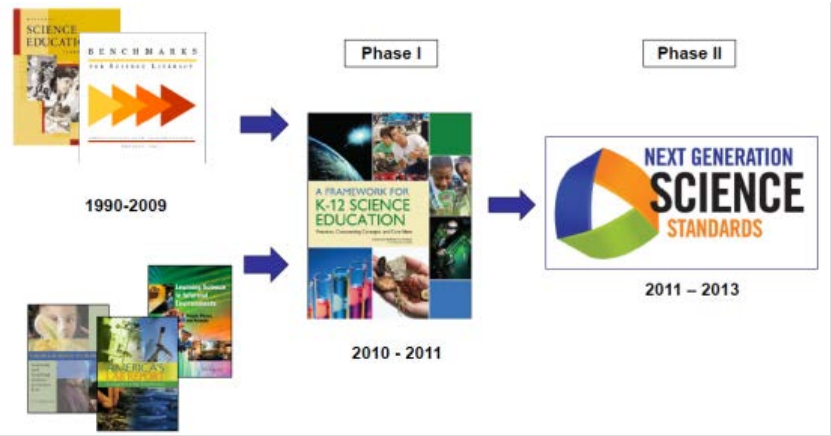

Figure 2.1 Development of the NGSS: Building on the Past; Preparing for the Future

California participated in the national development of the NGSS via the involvement of 80 members of the Strategic Leadership Team (SLT). The final public review of the NGSS occurred in January 2013, and a final version was released in April 2013. One of the SEP's roles was to review the NGSS and feedback from public forums and surveys on the NGSS, including the thousands of comments submitted to Achieve during the reviews of the draft versions of national NGSS.

Following release of the final version of NGSS, the state initiated the process of its adaptation for use in California by selecting 27 Science Expert Panel (SEP) members. This panel's members included K–12 teachers, County Office of Education science leaders, institution of higher education faculty, business and industry professionals, informal science center staff, and science advisors. The SEP provided recommendations for modifications of the NGSS to the State Superintendent of Public Instruction (SSPI). Based on California public feedback, the SEP made the following adaptations:

[1] Achieve is an independent, nonpartisan, nonprofit education reform organization dedicated to working with states to raise academic standards and graduation requirements, improve assessments, and strengthen accountability. Achieve managed the process of writing the NGSS (NGSS, 2014).

- Modification of performance expectations (PEs) clarification statements (For details about specific modifications, see http://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp.)
- Reorganization of the NGSS structure
- Development and application of criteria to redesign PEs and Learning Progression for the middle grades
- Current development of implementation recommendations for the California NGSS

California adopted the California NGSS on September 4, 2013 (California Department of Education, 2013).

For more information about the development process, refer to item 9 on the CDE's NGSS Frequently Asked Questions Web page at http://www.cde.ca.gov/pd/ca/sc/ngssfaq.asp#e9.

2B. NGSS Architecture

The NGSS are structured to emphasize the intertwining nature of three dimensions—science and engineering practices, crosscutting concepts, and disciplinary core ideas—and are written as performance expectations. The NGSS require contextual application of the three dimensions by students, with a focus on how and why, as well as what. For instance, the National Science Teachers Association describes the NGSS as follows:

> NGSS differs [sic] from prior science standards in that they integrate three dimensions (science and engineering practices, disciplinary core ideas, and crosscutting concepts) into a single performance expectation and have intentional connections between performance expectations. The system architecture of [the] NGSS highlights the performance expectations as well as each of the three integral dimensions and connections to other grade [spans] and subjects. The architecture involves a table with three main sections.
>
> A performance expectation describes what students should be able to do at the end of instruction and incorporates a practice, a disciplinary core idea, and a crosscutting concept from the foundation box. Performance expectations are intended to guide the development of assessments. Groupings of performance expectations do not imply a preferred ordering for instruction—nor should all performance expectations under one topic necessarily be taught in one course.
>
> During instruction, teachers will need to have students use multiple practices to help students understand the core ideas. Most topical groupings of performance expectations emphasize only a few practices or crosscutting concepts; however, all are emphasized within a grade band. (Willard, 2013)

Please see the embedded PDF resources and Figure G.1 in Appendix G for further description of the NGSS architecture.

2C. Legislation

AB 484, chaptered into California *EC* Section 60640(b), establishes the California Assessment of Student Performance and Progress (CAASPP), commencing with the 2013–14 school year, as the statewide assessment program for specified pupils. *EC* Section 60640(b) addresses components that impact science assessment, including assessments beyond the 2013–

14 school year, assessment development, assessments needed to meet the requirements of the federal ESEA, and additional (non-ESEA) assessments that are aligned to the NGSS.

EC Section 60640(b) also provides direction to the State Board of Education (SBE), the SSPI, and the CDE on the administration and transition of California's assessment system to the CAASPP System. In addition, *EC* Section 60640(b) outlines the assessments that are to be part of CAASPP—some of which were used previously as part of the Standardized Testing and Reporting (STAR) Program—and suspends non-ESEA tests.

The stakeholder meetings held from July 15–18, 2014, were necessitated by California legislation meant to address aspects of the Transition and Implementation phases[2] of NGSS assessments. *EC* Section 60640 outlines the requirements regarding: (1) the development and implementation of grade-level statewide science assessments aligned to the newly adopted NGSS; and (2) the expansion of science assessments to augment these grade-level tests. Senate Bill 300, chaptered into *EC* Section 60640(2)(B), permits the development of a new science curriculum framework based on the NGSS with anticipated adoption of this framework in 2016.

Aspects Related to Development of California Science Assessments

Testing After the 2013–14 School Year:

In accordance with *EC* Section 60640(b)(2)(A), until a successor assessment aligned to the NGSS is developed and implemented, the California Standards Tests (CSTs), California Modified Assessment (CMA), and California Alternate Performance Assessment (CAPA) for science in grades five, eight, and ten (Life Science [LS]) will be administered. End-of-course (EOC) CSTs in Biology, Chemistry, Physics, and Integrated Science 1–4 will continue to be available for purchase through ETS, but are not ESEA-required (*EC* Section 60640[d]).

Meeting Requirements of the Federal ESEA:

In accordance with *EC* Section 60640(b)(2)(B), in order to meet federal ESEA requirements, stakeholders were asked to make recommendations regarding what type of assessments should be developed to align to the California NGSS. Stakeholders were also asked to recommend what science content should be assessed and the grade levels at which the assessment should be administered. See Section 4 for summaries of recommendations by participants at the Science Stakeholder Meetings to the CDE. (See Section 7 for summaries of survey respondents' recommendations on ESEA science testing.)

Assessing Beyond Federal ESEA Requirements:

EC Section 60640(c) allows for the expansion of the CAASPP to include additional (non-ESEA) assessments for grade levels K–12 that are aligned to the NGSS and beyond the scope of those assessments specified in *EC* Section 60640(b). Stakeholders were asked to make recommendations for additional assessments, while also considering assessments that are already being administered or planned based on *EC* Section 60640(b) and the use of consortium-developed assessments. See Sections 5 and 6 for summaries of recommendations by participants at the Science Stakeholder Meetings to the CDE. (See Section 7 for summaries of survey respondents' recommendations on non-ESEA science testing.)

[2] See question 26 on the CDE's NGSS Frequently Asked Questions Web page at http://www.cde.ca.gov/pd/ca/sc/ngssfaq.asp#e26 for more information about NGSS implementation phases.

Section 3: Methodology

3A. Stakeholder Recruiting Processes

EC Section 60640(b) requires the SSPI to consult with stakeholders in developing recommendations for science assessments aligned to the NGSS. ETS, in collaboration with the CDE, recruited stakeholders representative of California's diverse population. A complete list of organizations contacted by ETS to recruit participants can be found in Appendix A.

Table 3.1 shows the target stakeholder representation, by percentage of the total, of each meeting. The target representation guided the participant recruitment and selection processes:

Table 3.1 Target Representation of Stakeholder Groups

Percent	Stakeholder Group
50%	California K–12 science teachers and administrators
10%	Experts in assessing English learners (ELs)
10%	Experts in assessing students with disabilities
10%	Parents/guardians
10%	Higher education experts
10%	Other professionals (i.e., scientists, researchers, business professionals)

To recruit meeting participants, ETS distributed an online application to the aforementioned stakeholder groups and to local educational agencies (LEAs). A transcript of the application can be found in Appendix B. Representatives of the organizations and LEAs circulated the application, and interested individuals applied to participate in a meeting. Each application was carefully considered and reviewed by the CDE and ETS. Selection of the participants was based on an applicant's relevant experience, expertise, and representation of the specific demographics and/or stakeholder group. Table 3.2 shows the counts of meeting participants representing particular groups.

Table 3.2 Stakeholder Groups Represented at the Meetings

Stakeholder	Number of Participants	
	Meeting 1 (July 15–16)	Meeting 2 (July 17–18)
K–12 administrators	6	4
K–5 teachers	12	12
Middle school teachers	11	11
High school teachers	11	12
Experts in assessing ELs	4	6
Experts in assessing students with disabilities	6	6
Parents/Guardians	4	5
Higher education experts	4	5
Other professionals (*i.e., scientists, researchers, business professionals*)	6	5
Total	**64**	**66**

The 130 participants at the meetings represented the following organizational affiliations; note that participants represented more organizations than those initially sought out for recruitment and that are listed in Appendix A.

- Academy of Nutrition and Dietetics (AND), formerly American Dietetic Association (ADA)
- American Association for the Advancement of Science (AAAS)
- American Association of Diabetes Educators (AADE)
- American Association of Physics Teachers (AAPT)
- American Chemical Society (ACS)
- American Public Health Association (APHA)
- American School Counselor Association (ACSA)
- Association for Science Teacher Education (ASTE)
- Association for Supervision and Curriculum Development (ASCD)
- Bechtel
- California Association for the Gifted (CAG)
- California Association of Bilingual Educators (CABE)
- California Association of Resource Specialists (CARS+)
- California Department of Education (CDE)
- California Educational Research Association (CERA)
- California English Language Development Test (CELDT) District and Site Coordinators
- California Science Project (CSP)
- California Science Teacher Association (CSTA)
- Chevron
- Computer-Using Educators (CUE)
- Curriculum and Instruction Steering Committee (CISC)
- Global Legislators Organization for a Balanced Environment (GLOBE)
- Monterey Bay Aquarium Educator Programs
- National Association for Research in Science Teaching (NARST)
- National Association of Biology Teachers (NABT)
- National Association of Geoscience Teachers (NAGT)
- National Board Certified Teachers (NBCT)
- National Commission for Health Education Credentialing (NCHEC)
- National Earth Science Teachers Association (NESTA)
- National Middle Level Science Teachers Association (NMLSTA)
- National Science Teachers Association (NSTA)
- Parent Teacher Association (PTA)
- Project Lead the Way (PLTW)
- San Diego Science Alliance (SDSA)
- San Diego Science Educators Association (SDSEA)
- Science Expert Panel (SEP)
- Southern California Association of Science Specialists ($SCAS^2$)
- Special Education Local Plan Area (SELPA)
- Technical Advisory Group (TAG)
- Technology and Telecommunications Steering Committee (TTSC)

3B. Meeting Processes

Introduction

The task of science stakeholders invited to the meetings was to provide input on the shape and form of new California science assessments aligned to the NGSS. The meetings were open for public observation. Participants provided input through in-depth group discussions on different aspects of new science assessments, including but not limited to, assessments mandated by federal or state laws and regulations.

Prior to the Meetings

After the list of meeting participants was finalized, each participant was assigned to one of two sessions and within each session, to a room that adhered to the desired makeup of each stakeholder group shown in Table 3.1. Each room was then divided into two heterogeneous small groups proportionate to the makeup of the room. The hierarchy for each session is shown in Figure 3.1.

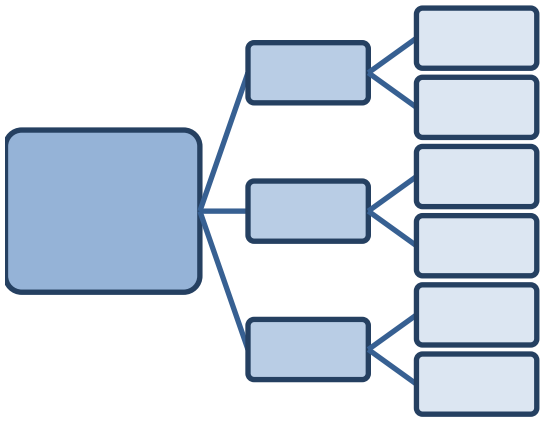

Figure 3.1 Stakeholder Meeting Hierarchy

Prior to the meetings, stakeholders were instructed to watch a recorded Webcast that provided an overview of the NGSS. The Webcast included background information on the NGSS, assessment design, and information and guiding questions based on legislative requirements of *EC* Sections 60640(a)(1)(B) and 60640(c) to assist in the development of ideas and recommendations.

Orientation materials and discussion guidelines were developed for the meetings. Six facilitators were identified and trained on the NGSS and the goals of the stakeholder meetings.

Contractor staffing at the meetings included the following roles:

- Lead facilitator for the general session: Introduced the subject matter and meeting goals
- Assessment Development experts: Subject matter experts; facilitated group discussions and also provided guidance, answered questions, and redirected discussions as needed
- Measurement experts: Copresented in the general session and provided psychometric guidance
- Program managers: Provided general and logistical oversight, liaised between client representatives and ETS experts, and provided general oversight of proceedings
- Note-takers: Took notes on the main conversation points for groups and any feedback from the CDE and ETS
- On-site logistics coordinator: Prepared the meeting space; provided participants with supplies and expense reimbursement information

General Session

Each meeting began with a general session, during which participants were given information about factors affecting the implementation of the NGSS in California, related legislation, and meeting goals and logistics. This presentation, found in the PDF embedded in Appendix D, also included the following meeting-specific topics:

- Overview of Previous Systems
- Legislation
- NGSS
- Science Assessments for California NGSS
- Special Studies

Group Sessions

After the conclusion of the general session, stakeholders were divided into three rooms, each of which was then subdivided into two groups for a total of six groups at each meeting.

Facilitator Protocols:

Each room had two facilitators, one for each group, who asked stakeholders entering the room to choose a group. The participants were, in some cases, asked to move to a new group to balance the stakeholder representation in each group. Facilitators then presented a PowerPoint presentation that reviewed main topics from the general session in the context of the discussion questions. A PDF of the room-level presentation is found in Appendix E.

After the presentation, facilitators gave the groups several explanatory publications (see PDFs in Appendix E) and a list of discussion questions (included in Appendix F) meant to stimulate discussion among and elicit recommendations from the stakeholders. At this point, each group was asked by the lead facilitator to pick a scribe to record the major topics of discussion and recommendations of the entire group using the discussion questions handout as a guide.

Facilitators were available during stakeholder discussions to answer questions and redirect the conversation as needed, with the intent of securing feedback for each of the discussion questions. Facilitators were responsible for ensuring that all stakeholders had ample opportunity to contribute feedback.

Facilitators instructed the groups to try to reach a majority consensus for each recommendation to a discussion question. Each group provided notes of the conversations and recommendations for each discussion question that were recorded by one or two stakeholders (in each group) who volunteered to be the group's scribe; some individuals also provided their own notes. If a majority consensus was not possible, facilitators asked the scribe of the group to write down the main recommendations and rationales, along with any information pertinent as to why consensus could not be reached among the group. Any issues raised by stakeholders that were outside the scope of the group discussions were recorded by the room's facilitators until CDE staff were available to respond directly to the group. CDE staff were available to answer any policy-related questions; they were not direct participants in the group discussions. A measurement expert was also available in each room to answer questions on psychometric issues.

Documenting the Meeting

The general session and group sessions were recorded via audio and by in-person note-takers (one per room). The notes from each group's scribe, facilitator, and note-taker were used to clarify the stakeholders' recommendations during discussions held at the end of each group session and to support the recommendations recorded by facilitators. Results were tabulated at the group level, relying on any majority consensus that was recorded by the scribe for each group. Recommendations and prevailing rationales from each group can be found in Sections 4 and 5; summarized recommendations are found in Section 6.

3C. Stakeholders

The meetings involved the following stakeholders outlined in *EC* Section 60640(b):

- California science teachers
- Individuals with expertise in assessing ELs
- Individuals with expertise in assessing students with disabilities
- Parents/Guardians
- Measurement experts

California K–12 administrators, higher education experts, scientists, researchers, engineers, and business professionals were also invited to participate in the meetings.

Appendix A lists the organizations that were contacted to recruit participants. The list of organizations represented is on page 7. Table 3.2 lists the number and types of stakeholders represented at each meeting.

3D. Methods Used to Analyze the Data

The purpose of these meetings was to gather input from groups of stakeholders from different educational, industry, and business organizations—see Appendix A for list of organizations contacted for participant recruitment, and see the list on page 7 for the organizations that were represented at the meetings. Representatives from these stakeholder groups collaborated in making recommendations for each of the guiding questions given to them by room facilitators (see Appendix F for guiding questions). Stakeholders were informed at the meetings that they would have an opportunity to give their own, individual opinions in an online survey, and later received invitations to participate in this online survey. See Section 7 for survey results.

ETS staff then analyzed the summaries of the major discussion points of the stakeholders using these resources: systematically taken group notes, including the group scribe's notes and individual stakeholders' notes; and notes from the room's note-takers that addressed both common themes and the majority and minority opinions of stakeholders. After the meeting, ETS staff also replayed and summarized audio recordings in an outline format. These outlines were used in sections 4 and 5 to summarize both the major points discussed by stakeholders and group notes from each session of the stakeholder meetings. These summaries are a preliminary indication of the participants' recommendations and rationales.

Section 4: Discussion and Feedback Pursuant to *Education Code* (*EC*) 60640(b)

Participants were subdivided into 12 groups, each with 10–12 stakeholders, to discuss the assessments needed to meet the requirements of *EC* Section 60640(b). Table 6.2 in Section 6B contains data for recommendations in the following areas:

Grade levels (GL)	*EC* Section 60640(b) grade spans: • Elementary School—three to five • Middle School (MS)—six to nine • High School (HS)—ten to twelve *EC* Section 60649(c) grade spans: • Kindergarten to twelve
Assessed Performance Expectations	• Grade-level-specific performance expectations (grade 3, grade 4, etc.) • MS performance expectations • HS performance expectations • End-of-course (EOC) • End-of-year (EOY) • Life Science (LS)
Options	• Paper-pencil (P/P) • Computer-based testing (CBT) • Computer-adaptive testing (CAT) • Multistage CAT (MSg CAT)
Item types and development	• Technology-enhanced (TE) • Locally scored performance-based task (PT) • Portfolio • Consortium-developed • Item bank (IB) • Varied
Assessment types	• Formative (F) • Interim (I) • Summative (S)

Note that the terms “EOC” and “EOY” were used by participants in their recommendations. An EOC assessment is based on the idea that grade-level instruction in science may be course-specific and an assessment such as the CST for Chemistry may be administered near or at the completion of the course. These courses may span all or part of a school year.

An EOY assessment is based on the idea that grade-level instruction in science may be more of an integration of science domains as described by the NGSS, and as such, an assessment may be administered near or at the end of the year.

The subsections of Section 4 summarize both the major discussion points and group notes from each session of the stakeholder meetings. Notes addressed both common themes and minority opinions of meeting participants.

Table 4.1 through Table 4.8 provide the meeting dates, room number, and designations of groups. Table 4.1 through Table 4.6 also provide recommendations and rationales for grade levels or content area (by performance expectations). Table 4.7 and Table 4.8 provide recommendations and rationales for the assessment options. The information in the tables is organized horizontally.

4A. Grade Level

Grade Span Three to Five, Inclusive

A concern expressed by all groups was whether or not science is consistently taught prior to grade five. Current fifth-grade teachers noted that many of their students seem unprepared for science learning when they begin the school year, and question how well prepared the new fifth graders are. Other considerations and concerns included:

- Student literacy at lower grades
- Teachers at lower grades dedicating more time to English–language arts (ELA) and mathematics than science concepts
- Teachers at lower grades who are generalists rather than subject-specific experts

Table 4.1 summarizes stakeholder recommendations about the grade level for the ESEA-mandated test for grade span three to five, including the prevailing rationale for each group's recommendation.

Table 4.1 Summary of Stakeholder Recommendations for Grade Level, Grade Span Three to Five

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	5	• Assessment earlier than grade five is more of a reading test. • At this age, students are mature enough to provide arguments about evidence.
		B	5	• The test should be a culminating assessment for elementary school science.
	2	C	3	• The test should be administered in the third grade to promote early emphasis on science education.
		D	4	• The test should be administered in the fourth grade to emphasize need for science education to begin before fifth grade.
	3	E	5	• A test of the end-of-elementary-school span allows assessment of full range of performance expectations.
		F	3	• The test can be used as a system that measures learning progression of students over time.

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/17–7/18	1	A	3 and 5	• The test can be used to demonstrate and assess growth of students.
		B	3 and 5	• The test can show science learning progressions and emphasize early start to science education.
	2	C	4 or 5	• The test should be administered in the fourth grade to emphasize early start to science education. • The test should be administered in the fifth grade because this grade is at the end of the elementary school span and allows the best opportunity to expand on the performance expectations.
		D	5	• The test should be administered in the fifth grade to allow time for emphasis of all performance expectations in the curriculum.
	3	E	5	• Since big ideas are emphasized each year, a culminating assessment of all elementary grades maximizes the depth of each performance expectation.
		F	5	• A test in the fifth grade allows for most exposure to elementary performance expectations, including as a culmination of all elementary grades.

Most discussions about the content area assessed for ESEA-mandated tests in grades three through five focused on the fact that, at those levels, the NGSS are set as grade specific rather than in spans like middle school.

There was significant discussion about covering crosscutting concepts and whether or not it is fair to fifth-grade students and teachers to have a test covering material taught in previous grades. Stakeholders referenced experience with the current assessment system in which students are assessed at grade five over both grade four and grade five standards. There was significant group discussion (but no clear consensus) for three of the groups about how they felt, that the fifth-grade science teachers were being held accountable for the content that students should have received in fourth grade and desired an assessment system that promotes early science education and grade-level responsibility for instruction. Because NGSS performance expectations at elementary grades, through grade five, are organized at grade level, stakeholders were concerned that an assessment at grade five would mirror their current experience with instruction at lower grade levels. Despite the concern expressed in the discussion, several of these groups went on to recommend that a grade five test could serve as a culminating assessment of elementary school science instruction for all grades, three through five.

Table 4.2 summarizes stakeholder recommendations on the content area for an ESEA-mandated test for grade span three to five, including the prevailing rationale for each group's recommendation.

Table 4.2 Summary of Stakeholder Recommendations for Content Area, Grade Span Three to Five

Meeting Dates	Room	Group	Performance Expectations	Rationale
7/15–7/16	1	A	Grade 5	• The test should be administered in the fifth grade to allow students to demonstrate mastery of performance expectations (PEs) for all grades.
		B	Grades 3–5	• An assessment for elementary school should include PEs from a span of grade levels.
	2	C	Grade 3	• Grade-level PEs are built on previous expectations, so grade-level–specific performance expectations taught should be the PEs assessed.
		D	Grades 3–4	• PEs for both grades three and four should be covered to emphasize need for science at lower grades.
	3	E	Grades 3–5	• The test should assess all learning through elementary school grades.
		F	Grades 3–5	• The test should cover the span of PEs to assess learning progressions.
7/17–7/18	1	A	Grade 3 and Grade 5	• The test should emphasize crosscutting concepts throughout all grades, applying knowledge in practical application within and outside of science in a manner that emphasizes problem-solving and decision-making. • Practical application of knowledge and skills is more important than any specific content.
		B	Grade 3 and Grade 5	• The test should assess PEs through the grade level to emphasize early science education and demonstrate growth.
	2	C	Grades 3–4 or Grades 3–5	• The test should assess what is taught up to each grade-level assessment.
		D	Grades 3–5	• If you assess spiral performance expectations across administrations, then students will be allowed time to develop skills.
	3	E	Grades 3–5	• The test should assess all PEs.
		F	Grades 3–5	• The test should assess all PEs.

Grade Span Six to Nine, Inclusive

There was agreement among participants that it would be beneficial to have a test at the end of middle school, to both provide information about progress thus far and provide direction for high school science course selection. While grade nine was considered middle school for purposes of this discussion, teachers noted that middle school typically ends in eighth grade and that should be the cutoff for consideration of ESEA requirements.

Table 4.3 summarizes stakeholder recommendations about the grade level for the ESEA-mandated test for grade span six to nine, including the prevailing rationale for each group's recommendation.

Table 4.3 Summary of Stakeholder Recommendations for Grade Level, Grade Span Six to Nine

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	8	• The test should be an assessment of middle school science that provides direction for student science pathway in high school.
		B	8	• The test should assess all middle school science performance expectations (PEs).
	2	C	6	• The test should promote an early emphasis on science education and provide a benchmark for growth during elementary grades.
		D	8	• The test should assess all middle school science PEs.
	3	E	8	• The test should assess all middle school science PEs.
		F	6	• The test should be used as a system that measures learning progression of students over time.
7/17–7/18	1	A	8	• The test should assess all content from middle school science PEs.
		B	8	• The test should assess all middle school science PEs.
	2	C	8	• The test should assess all middle school science PEs.
		D	8	• A test at grade eight is appropriate to cover all levels. • The test should be designed in a way to hold lower grades accountable for instruction.
	3	E	8	• The test should assess all middle school science PEs.
		F	8	• The test should address large ideas from year to year by the end of grade eight.

Groups had differing opinions of the value of an assessment that covers all material at the six to eight grade span. While some felt such an assessment would provide valuable information about student growth and science literacy, others felt it was unnecessary and potentially unfair to students and teachers to test students on material introduced at earlier grades even though the curriculum is spanned.

Table 4.4 summarizes stakeholder recommendations about the content area for the ESEA mandated test for grade span six to nine, including the prevailing rationale for each group's recommendation. In this case, because the NGSS has no specific performance expectations per grade in middle school, stakeholders did not recommend performance expectations by grade level. Instead, the assumption was that the assessment would cover all middle school performance expectations.

Table 4.4 Summary of Stakeholder Recommendations for Content Area, Grade Span Six to Nine

Meeting Dates	Room	Group	Performance Expectations	Rationale
7/15–7/16	1	A	Middle School (MS)	• The test should incorporate all middle school science performance expectations (PEs) and crosscutting concepts.
		B	MS	• The test should cover PEs regardless of integrated or sequential curriculum.
	2	C	End-of-Year (EOY)	• The test should reflect content taught during the academic year.
		D	MS	• The test should incorporate all middle school science PEs.
		E	MS	• The test should incorporate all middle school science PEs.
	3	F	Grades 3–5 or EOY	• A grade five test should be used against a grade three benchmark to determine student growth in science knowledge. • The group was split between assessing PEs from elementary school grade band on a sixth grade assessment or a grade-level assessment over annual instruction at grade six.
7/17–7/18	1	A	MS	• The test should emphasize crosscutting concepts throughout all grades; practical application is more important than any specific content.
		B	MS	• The test should allow similar assessment of PEs regardless of varying middle school course/curriculum options.
	2	C	MS	• The test should assess end-of-middle-school science PEs.
		D	MS	• The test should assess end-of-middle-school science PEs.
	3	E	MS	• The test should assess all middle school science PEs.
		F	MS	• The test should assess all middle school science PEs.

Grade Span Ten to Twelve, Inclusive

The subject of when to test at high school elicited the greatest discussion among participants. A number of factors were considered, including:

- Some schools do not mandate science courses until tenth grade.
- Only two years of science are required for graduation in California high schools.
- There are heavy testing burdens at eleventh grade for students taking the California High School Exit Examination, Advanced Placement (AP) tests, and SATs.
- Motivating twelfth graders to try their hardest on a test that will have no impact on them will be difficult.

Table 4.5 summarizes stakeholder recommendations about the grade level for the ESEA-mandated test for grade span ten to twelve, including the prevailing rationale for each group's recommendation.

Table 4.5 Summary of Stakeholder Recommendations for Grade Level, Grade Span Ten to Twelve

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	11	• An eleventh grade test provides the opportunity to assess three years of science, but it should emphasize Life Science performance expectations (PEs).
		B	10 or 11	• The test should be administered in the tenth grade to avoid testing overload at grade eleven. • The test should be administered in the eleventh grade to provide an opportunity for students to gain three years of high school science instruction prior to testing.
	2	C	10	• The test should be a culminating assessment of high school science, but avoid testing during the Smarter Balanced administration year.
		D	11	• The test should be administered in grade eleven because students are likely to be finished with science instruction.
	3	E	11	• The test should be an assessment of assess all high school science PEs.
		F	11	• A test in grade eleven allows for the varied start of high school science instruction (grade nine or grade ten), and emphasizes a push for three years of science instruction.
7/17–7/18	1	A	10	• The test should be administered in grade ten, because most students will complete the two years of science requirements by grade ten.
		B	12	• A test in grade twelve provides students with an opportunity to complete several years of high school science instruction and avoid additional testing at grade eleven.
	2	C	10 or 11	• All students will fit most LEA course designs, whether students complete required science courses by grade ten or grade eleven as designated by their LEA. • The group was split between testing at grade ten or grade eleven because of the variances that exist among LEAs.
		D	11	• A test should be administered in grade eleven, because students are likely to be finished with science instruction.

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/17–7/18 (cont.)	3	E	11	• A test should be administered in grade eleven, covering all PEs. This will give students the opportunity to complete three years of science.
		F	10 or 11	• A test in tenth grade will mirror a majority of current course designs. • A test in eleventh grade will allow for three years of science before assessment of all high school science PEs.

There was little agreement among stakeholders about what content should be tested. Some groups supported an emphasis on crosscutting concepts and practices, while others felt that course-specific content was more appropriate depending upon the grade assessed (e.g., Life Science at grade ten).

Table 4.6 summarizes stakeholder recommendations about the content area for the ESEA-mandated test for grade span ten to twelve, including the prevailing rationale for each group's recommendation.

Table 4.6 Summary of Stakeholder Recommendations for Content Area, Grade Span Ten to Twelve

Meeting Dates	Room	Group	Performance Expectations	Rationale
7/15–7/16	1	A	High School (HS)	• The test should emphasize crosscutting concepts throughout all grades; practical application is more important than any specific content.
		B	Life Science or HS	• The test should be focused on Life Science if given at grade ten because of a lack of opportunity to take three years of high school science before the assessment.
	2	C	End-of-Course (EOC)	• The test should contain content appropriate to year of instruction.
		D	HS	• The test should incorporate all high school science performance expectations (PEs).
	3	E	HS	• The test should incorporate all high school science PEs.
		F	HS	• The NGSS demand assessment of all science PEs.
7/17–7/18	1	A	EOC	• The test should emphasize crosscutting concepts throughout all grades; practical application is more important than any specific content.
		B	HS	• The test should assess all high school science PEs.
	2	C	HS	• The test should assess all high school science PEs.
		D	HS	• The test should assess all high school science PEs.
	3	E	EOC or HS	• The test should focus on course instruction; if there is emphasis on three years of science, then all high school science PEs could be assessed.
		F	HS	• The test should focus on practices and concepts rather than core ideas.

Note: The term "EOC" was used by participants in their recommendations. An EOC assessment is based on the idea that grade-level instruction in science may be course-specific and an assessment such as the CST for Chemistry may be administered near or at the completion of the course. These courses may span all or part of a school year.

4B. Type of Assessment

Participants were enthusiastic about online assessments, although there was disagreement about the viability of computer-adaptive testing (CAT). While many educators liked the idea of CAT for California students, others felt that given the costs it would be impractical to administer at all grades and they would prefer less expensive computer-based testing (CBT) if that meant they could test at more grades beyond those that are federally mandated. There was some support for non-computer performance-based tasks, described as hands-on tasks that could be scored by raters. These tasks were recommended to be standardized and materials provided through the assessment system.

Despite fruitful discussions about the value of formative versus summative assessments, all groups ultimately recommended summative assessments. These were seen as providing more useful feedback at the LEA level or above. They also recommended that a formative item bank aligned to the NGSS would be a valuable tool for teachers and students.

For All Grade Spans

Table 4.7 summarizes stakeholder recommendations at the group level on the assessment options for the ESEA-mandated tests.

Table 4.7 Summary of Stakeholder Recommendations for Assessment Options, ESEA-mandated Tests, All Grade Spans

Meeting Dates	Room	Group	Assessment Options
7/15–7/16	1	A	computer-based testing (CBT)
		B	computer-adaptive testing (CAT)
	2	C	multistage (MSg) CAT
		D	CAT
	3	E	CAT
		F	CBT, CAT, and PT
7/17–7/18	1	A	CAT
		B	CAT
	2	C	MSg CAT
		D	CAT
	3	E	CAT and PT
		F	CAT and PT

Table 4.8 summarizes stakeholder recommendations at the group level on the assessment type for the ESEA-mandated tests.

Table 4.8 Summary of Stakeholder Recommendations for Assessment Types, ESEA-mandated Tests

Meeting Dates	Room	Group	Assessment Options
7/15–7/16	1	A	Summative
		B	Summative
	2	C	Summative
		D	Summative
	3	E	Summative
		F	Summative
7/17–7/18	1	A	Summative
		B	Summative
	2	C	Summative
		D	Summative
	3	E	Summative
		F	Summative

Section 5: Discussion and Feedback Pursuant to *EC* Section 60640(c)

The subsections of Section 5 expand on the major discussion points and group notes from each session of the stakeholder meetings regarding additional assessments beyond those recommended for *EC* Section 60640(b). Notes addressed both common themes and minority opinions of meeting participants and/or groups.

Table 5.1 through Table 5.8 provide the meeting dates, room number, and designations of groups. Table 5.1 through Table 5.6 also provide recommendations and rationales for grade levels or content area (by performance expectations). Table 5.7 and Table 5.8 provide recommendations and rationales for the assessment options. The information in the tables is organized horizontally.

5A. Grade Levels

Grade Span Kindergarten to Grade Five

Many participants were concerned that science is not actively being taught in elementary school grades and felt that earlier assessments would remedy that. A variety of annual assessment types, developed either at the LEA level or as part of a statewide assessment system, were discussed as valuable tools for teachers to determine progress and identify areas for specific focus, including formative, summative, and interim assessments. A large portion of the groups also recommended a state-developed item bank that could be used by LEAs to generate benchmark or unit assessments at the classroom level.

Table 5.1 summarizes stakeholder recommendations about the grade level for the non–ESEA-mandated tests for grade span three to five, including the prevailing rationale for each group's recommendation.

Table 5.1 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Three to Five

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	3, 4	• LEAs should have end-of-course assessments in all content areas and grade levels.
		B	Kindergarten (K)–4	• Formative benchmarks for K–2 and all grades up through grade twelve should be available from a state-created test bank.
	2	C	4, 5	• There should be annual benchmark assessments that teachers can use as a formative tool.
		D	3, 5	• There should be annual assessments to show student progress.
	3	E	3, 4	• There should be annual assessments to show student progress.
		F	4, 5	• There should be annual benchmark assessments.

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/17–7/18	1	A	4	• There should be small summative assessments at non-federally required grades to evaluate student knowledge and performance.
		B	3, 4	• There should be item banks created by the state that teachers can draw from. This would allow regular and systematic ways to evaluate student learning each year.
	2	C	K–3 or 4	• A test covering K–4 would make K–3 teachers accountable to these standards; that being said, these contents do build on each other. K–2 is very basic, and grades three through four are similar. • A fourth grade test can determine whether a child is at/above proficient or not proficient.
		D	2, 3, 4	• There should be an item bank available for every grade level based on performance expectations.
	3	E	3, 4	• There should be annual census assessments to show student progress.
		F	2, 3, 4	• The test should emphasize the importance of primary science and effectiveness of instruction throughout the school experience.

Participants supported assessing grade-specific content because the performance expectations at the elementary school-level are organized by grade level rather than being spanned across several grades, such as the middle school and high school performance expectations.

Table 5.2 summarizes stakeholder recommendations about the content area for the non–ESEA-mandated tests for grade span three to five, including the prevailing rationale for each group's recommendation. Note that an EOY assessment is based on the idea that grade-level instruction in science may be more of an integration of science domains as described by the NGSS and, as such, an assessment may be administered near or at the end of the year.

Table 5.2 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Three to Five

Meeting Dates	Room	Group	Performance Expectations	Rationale
7/15–7/16	1	A	End-of-Year (EOY)	• A test should cover all performance expectations across the span of grade levels.
		B	EOY	• The assessment system should focus on the growth of the student between tests.
	2	C	EOY	• The assessment system should focus on learning progressions across the year through reflection of previous-year assessments.
		D	EOY	• There should be a summative test at the beginning of the third grade, as it gives you an opportunity in third, fourth, and fifth grades to remediate everything.

Meeting Dates	Room	Group	Performance Expectations	Rationale
7/15–7/16 (cont.)	3	E	EOY	• Tests should assess what is taught in a course through a specific grade level.
		F	EOY	• A test in the third and fourth grades should integrate with their specific grade level.
7/17–7/18	1	A	EOY	• The assessment system should emphasize practices and processes in science.
		B	EOY	• The assessment system should promote development of state-mandated end-of-course tests for high school and an optional item bank for kindergarten (K)–12.
	2	C	EOY	• There should be an item bank with test items ranging from very simple to very complex that integrates K–4 concepts, but with the focus on fourth grade PEs, including practice and crosscutting concepts.
		D	EOY	• The assessment system should evaluate student science knowledge each year.
	3	E	EOY	• Students should be evaluated on knowledge of specific content, practices, and concepts taught each year.
		F	EOY	• Assessments should address larger ideas in previous years, linking big ideas back to current learning.

Grade Span Six to Nine

Participants supported annual assessments, developed either at the LEA level or as part of a statewide assessment system. They felt that these benchmarks would be valuable tools for teachers to determine progress and identify areas for specific focus. Some groups recommended these assessments be summative in nature, allowing state-level comparisons of student knowledge and skill, while other groups recommended these be formative assessments and incorporated into the federal accountability reporting system. However, they were unsure if this would be practical given the current system of ESEA assessments.

Table 5.3 summarizes stakeholder recommendations about the grade level for the non–ESEA-mandated tests for grade span six to nine, including the prevailing rationale for each group's recommendation.

Table 5.3 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Six to Nine

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	6, 7, 9	• LEAs should have end-of-course assessments in all content areas and grade levels.
		B	6, 7	• Formative benchmarks for kindergarten (K)–2 and all grades up through grade twelve should be available from a state-created test bank.
	2	C	7, 8, 9	• There should be annual benchmark assessments that teachers can use as a formative tool.

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16 (cont.)	2 (cont.)	D	6, 7	• There should be annual assessments to show student progress.
	3	E	6, 7, 9	• There should be annual assessments to support understanding of student knowledge.
		F	7, 8	• There should be annual benchmark assessments.
7/17–7/18	1	A	6, 7, 9	• The assessment system should emphasize practices and processes in science.
		B	6, 7, 9	• There should be item banks created by the state that teachers can draw from. This would allow regular and systematic ways to evaluate student learning each year.
	2	C	6, 7, 9	• There should be formative benchmarks from an available test bank.
		D	6, 7, 9	• An item bank based on performance expectations should be available for every grade level.
	3	E	6, 7, 9	• There should be a census test of students each year.
		F	6, 7, 9	• The assessment system should emphasize the importance of primary science and effectiveness of instruction throughout the school experience.

Participants disagreed as to what content or concepts should be covered in non–ESEA-mandated tests in middle school. Some supported assessments that tracked growth across grades, while others preferred to focus on specific content, practices, and concepts taught each year. Table 5.4 summarizes stakeholder recommendations about the content area for the non–ESEA-mandated tests for grade span six to nine, including the prevailing rationale for each group's recommendation.

Table 5.4 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Six to Nine

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	End-of-Course (EOC)/End-of-Year (EOY)	• Tests should cover all performance expectations across a span of grade levels.
		B	EOC/EOY	• The assessment system should focus on the growth of the student from test-to-test assessments and should be set on student growth.
	2	C	EOC/EOY	• There should be a bank with test items ranging from very simple to very complex.
		D	EOC/EOY	• There should be a summative test at the beginning of the sixth grade, as it gives you an opportunity to mix up sixth, seventh, and eighth to remediate everything.
	3	E	EOC/EOY	• Tests should assess what is taught in a course during that year of instruction at specific grade level (six, seven, nine).

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16 (cont.)	3 (cont.)	F	EOC/EOY	• One group opinion held that sixth and seventh grade middle schools can be integrated or not, so you would have to test on the lowest common denominator if administering a single test. • Another group opinion wanted two separate tests for the separate paths.
7/17–7/18	1	A	EOC/EOY	• The assessment system should emphasize practices and processes in science.
		B	EOC/EOY	• The assessment system should promote development of state-mandated EOC tests for high school and an optional item bank for kindergarten through grade twelve.
	2	C	EOC/EOY	• The NGSS essentially build on previous knowledge; therefore tests should be integrated as well.
		D	EOC/EOY	• The assessment system should evaluate science knowledge each year.
	3	E	EOC/EOY	• Students should be evaluated on specific content, practices, and concepts taught each year.
		F	EOC/EOY	• Assessments should address larger ideas in previous years, linking big ideas back to current learning.

Note: The terms "EOC" and "EOY" were used by participants in their recommendations. An EOC assessment is based on the idea that grade-level instruction in science may be course-specific and an assessment such as the CST for Chemistry may be administered near or at the completion of the course. These courses may span all or part of a school year. An EOY assessment is based on the idea that grade-level instruction in science may be more of an integration of science domains as described by the NGSS and, as such, an assessment may be administered near or at the end of the year.

Grade Span Ten to Twelve

Participants supported annual assessments, developed either at the LEA level or as part of a statewide assessment system. They felt that these benchmarks would be valuable tools for teachers to determine progress and identify areas for specific focus.

Table 5.5 summarizes stakeholder recommendations about the grade level for the non–ESEA–mandated tests for grade span ten to twelve, including the prevailing rationale for each group's recommendation.

Table 5.5 Summary of Stakeholder Recommendations for Grade Level of Non–ESEA-mandated Tests, Grade Span Ten to Twelve

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	10	• LEAs should have end-of-course assessments in all content areas and grade levels.
		B	9, 10 or 11, 12	• Formative benchmarks for kindergarten (K)–2 and all grades up through grade twelve should be available from a state-created test bank.

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16 (cont.)	2	C	11, 12	• There should be annual benchmark assessments that teachers can use as a formative tool.
		D	10	• There should be annual assessments to show student progress.
	3	E	10	• There should be annual assessments to support understanding of student knowledge.
		F	10, 12	• There should be annual benchmark assessments.
7/17–7/18	1	A	10	• The assessment system should allow testing before the jump in student drop-out rates at upper grades. Also, students are still interested in science at this grade level (tenth).
		B	10, 11	• There should be item banks created by the state that teachers can draw from, which would allow regular and systematic ways to evaluate student learning each year.
	2	C	10 or 11, 12	• There should be formative benchmarks from an available test bank.
		D	10	• An assessment bank based on performance expectations should be available for every grade level.
	3	E	10	• There should be a census test of students each year.
		F	10 or 11, 12	• A test should be a measure of the effectiveness of instruction throughout the school experience.

As with middle school, there was little agreement about what should be assessed at high school. Some participants advocated assessments that draw from multiple courses and grades and focused on learning progressions, whereas others felt that course-specific content (similar to the current end-of-course model) was more appropriate.

Table 5.6 summarizes stakeholder recommendations about the content area for the non–ESEA-mandated tests for grade span ten to twelve, including the prevailing rationale for each group's recommendation.

Table 5.6 Summary of Stakeholder Recommendations for Content Area of Non–ESEA-mandated Tests, Grade Span Ten to Twelve

Meeting Dates	Room	Group	Recommended Grade(s)	Rationale
7/15–7/16	1	A	End-of-Course (EOC)	• Stakeholders want coverage of all performance expectations across span of grade levels.
		B	High School (HS)	• EOC tests would give flexibility and local control to LEAs , rather than be required by state mandate to administer certain tests.
	2	C	EOC	• The assessment system should focus on learning progressions across the year through reflection of previous-year assessments.

7/15–7/16 (cont.)	2 (cont.)	D	EOC	• The assessment system should emphasize culminating learning progressions throughout kindergarten (K)–12 science.
	3	E	EOC	• The assessment system should assess what is taught in a course through a specific grade level.
		F	HS	• A test should represent a benchmark that would assess everything students have been taught up to that point.
7/17–7/18	1	A	EOC	• Tests should allow assessment of performance expectations specifically targeted during a course.
		B	EOC	• An assessment system should promote development of state-mandated EOC tests for high school and an optional item bank for K–12.
	2	C	HS	• There should be an item bank with test items ranging from very simple to very complex.
		D	HS	• The assessment system should evaluate student knowledge for each year of science learning.
	3	E	EOC	• Students should be evaluated on specific content, practices, and concepts taught each year.
		F	HS	• Assessments should address larger ideas from previous years, linking big ideas back to current learning.

Note: The term "EOC" was used by participants in their recommendations. An EOC assessment is based on the idea that grade-level instruction in science may be course-specific and an assessment such as the CST for Chemistry may be administered near or at the completion of the course. These courses may span all or part of a school year.

5B. Type of Assessment

Similar to the recommendations for ESEA-mandated assessments, participants suggested that online assessments be the staple for non–ESEA-mandated assessments. There was also some support for non-computer performance-based tasks. In particular, many of the stakeholders referenced the previous California assessment system, the Golden State Examinations, which utilized hands-on performance-based tasks in assessments of student science skills and knowledge. These performance-based tasks were provided by the state as kits and scored locally.

There was more support for formative assessments at the non–ESEA-mandated grades than at the mandated grades, although participants were still split between the two. Some teachers felt that receiving feedback about student performance earlier in the school year would help identify struggling students and provide opportunities for remediation.

Table 5.7 summarizes stakeholder recommendations (at the group level) on the assessment options for non–ESEA-mandated tests.

Table 5.7 Summary of Stakeholder Recommendations for Assessment Options, Non–ESEA-mandated Tests

Meeting Dates	Room	Group	Assessment Options
7/15–7/16	1	A	computer-adaptive testing (CAT)
		B	computer-based testing (CBT)
	2	C	multistage CAT
		D	CAT and performance-based task (PT)
	3	E	CAT/PT
		F	CBT, CAT, and PT
7/17–7/18	1	A	CAT
		B	Item Bank
	2	C	Varied
		D	CBT and Varied
	3	E	CAT and PT
		F	CAT and PT

Table 5.8 summarizes stakeholder recommendations at the group level on the assessment type for non–ESEA-mandated tests.

Table 5.8 Summary of Stakeholder Recommendations for Assessment Types, Non–ESEA-mandated Tests

Meeting Dates	Room	Group	Assessment Options
7/15–7/16	1	A	Formative
		B	Formative
	2	C	Formative
		D	Summative
	3	E	Summative
		F	Summative
7/17–7/18	1	A	Summative
		B	Interim
	2	C	Formative and Summative
		D	Formative and Summative
	3	E	Formative, Summative, and Interim
		F	Summative

5C. NGSS Consortium–Developed Assessments

Stakeholders expressed interest in assessments developed by an NGSS consortium, citing benefits of a larger pool of NGSS-aligned items and tests that would reduce the costs and time needed to develop state-exclusive assessment materials. However, there was limited information to share since no national initiative for NGSS Consortium–developed assessments was underway at the time of this meeting.

5D. Various Item Types

The stakeholders recommended use of performance-based tasks to assess the majority of NGSS PEs. Performance-based tasks were defined as context-based activities encompassing a variety of item types, including open-ended constructed-response as well as physical actions. There was a mixture of recommendations for "hands-on" and "virtual" tasks using both in-person manipulatives and CBT simulations. The stakeholders recommended using as many technology-enhanced (TE) item types as possible.

There were also recommendations to limit use of stand-alone multiple choice items, though these item types would be appropriate for embedding in performance-based tasks.

5E. Online Testing

The stakeholder group recommended computer-adaptive testing for both ESEA and non-ESEA assessments.

Section 6: Results from CAASPP Science Stakeholders Meetings

6A. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(b)

Subsection 6A summarize stakeholder recommendations and major rationales for grade-level assessments within each grade span—three through five, six through nine, and ten through twelve—as required by *EC* Section 60640(b). Notes addressed both common themes and minority opinions of meeting participants/groups.

Grade Span Three to Five, Inclusive

Six of the 12 groups recommended assessment at grade five, citing students' grade-level maturity and maximized time for exposure to elementary science standards for students as their main concern. Two of the 12 groups recommended assessment at grade three, citing the need for early emphasis on science in elementary school education as well as a desire to advocate for accountability for teachers instructing science in lower grades. One of the 12 groups recommended assessment at grade four, while another group recommended grade four or five based on a split group discussion. The remaining 2 groups of the 12 recommended assessments at both grade three and grade five to reinforce the same rationales as those expressed above. Stakeholders referred to cumulative content as including all science topics normally taught at the targeted grade level and those taught at lower grade levels when they recommended integrated content for this grade level.

Grade Span Six to Nine, Inclusive

Ten of the 12 groups recommended assessment at grade eight due to the logical progression of end-of-middle-school performance expectations. Two of the 12 groups recommended assessment at grade six, consistent with their recommendations for assessment at grade three, to reinforce early emphasis of science education. They also described the desire to use early grade-level assessments in the growth of student performance and knowledge. Stakeholders provided the rationale of class offerings being variable in content type and sequence during middle school, necessitating an assessment at grade eight. Grade eight is frequently the final grade level of most middle schools, which would mean that the maximum amount of content may be assessed.

Grade Span Ten to Twelve, Inclusive

Six of the 12 groups recommended assessment at grade eleven based on the desire to allow time for students to experience three years of science, covering all high school NGSS performance expectations. This grade level also would provide flexibility for all students to take the minimum two years of required science courses before taking the test. Two of the 12 groups recommended assessment at grade ten because of the desire to avoid additional testing at the eleventh and twelfth grade levels, where Smarter Balanced testing, AP testing, and other assessments for entering a career and college will occur. Three of the 12 groups recommended assessment at either grade ten or eleven based on a split group discussion, citing similar rationales as stated above. One of the groups recommended assessment at grade twelve to

emphasize several years of science instruction and allow students to demonstrate full knowledge and learning progressions in K–12 science education.

Summary of Recommendations

Table 6.1 summarizes the stakeholder recommendations for grade-level assessments within each grade span—three to five, six to nine, and ten to twelve—as required by *EC* Section 60640(b). Notes addressed both common themes and minority opinions of meeting participants/ groups. Full results, including rationales, can be found in Table 4.1 through Table 4.8.

Table 6.1 Summary of Stakeholder Recommendations for Grade-level Assessments

	***EC* Section 60640(b), ESEA**											
	Meeting 1						**Meeting 2**					
Room	1						1					
Group	A			B			A			B		
Grade Level	5	8	11	5	8	10 or 11	3, 5	8	10	3, 5	8	12
Content	G5	MS	HS	G3-5	MS	LS or HS	G3, G5	MS	EOC	G3, G5	MS	HS
Option	CBT	CBT	CBT	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT
Type	S	S	S	S	S	S	S	S	S	S	S	S
Room	2						2					
Group	C			D			C			D		
Grade Level	3	6	10	4	8	11	4 or 5	8	10 or 11	5	8	11
Content	G3	EOC/EOY	EOC	G3-4	MS	HS	G3-4 or G3-5	MS	HS	G3-5	MS	HS
Option	MSg CAT	MSg CAT	MSg CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT	CAT
Type	S	S	S	S	S	S	S	S	S	S	S	S
Room	3						3					
Group	E			F			E			F		
Grade Level	5	8	11	3	6	11	5	8	11	5	8	10 or 11
Content	G3-5	MS	HS	G3-5	ES/EOY	HS	G3-5	MS	EOC or HS	G3-5	MS	HS
Option	CAT	CAT	CAT	CBT/CAT/PT	CBT/CAT/PT	CBT/CAT/PT	CAT/PT	CAT/PT	CAT/PT	CAT/PT	CAT/PT	CAT/PT
Type	S	S	S	S	S	S	S	S	S	S	S	S

Assessment Key
Grade Levels For *EC* Section 60640(b) at least one assessment per grade span (GS) 3–5, 6–9, and 10–12; for *EC* Section 60640(c) additional K–12 assessments to those proposed in *EC* Section 60640(b)
Content: Subjects to be assessed, grade level (elementary school [ES], middle school [MS], high school [HS]) performance expectation (PE), end-of-course (EOC), end-of-year (EOY), Life Science (LS)
Options: Paper-pencil (P/P), computer-based testing (CBT), computer-adaptive testing (CAT) or multistage (MSg) CAT, item type (such as technology enhanced [TE]), locally scored performance-based task (PT), portfolio, Consortium-developed item bank (IB), varied (V)
Assessment Types: Formative (F), Interim (I), Summative (S)

6B. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of Non–ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(c)

Subsection 6B summarizes stakeholder recommendations and major rationales for additional assessments at grade level as described by *EC* Section 60640(c). Notes addressed both common themes and minority opinions of meeting participants/groups.

Grade Span Kindergarten to Twelve

All of the groups recommended additional assessments for each grade spanning three through eleven. The main rationales behind these recommendations describe the desire to provide annual assessments that teachers, parents/guardians, and students could use to evaluate knowledge and skills. Two of the 12 groups recommended science assessments begin in kindergarten and continue through grade twelve as described and supported by the NGSS structure. Two of the 12 groups recommended including grade two assessments as part of their complete assessment system package. They suggested grade two would provide the benchmark assessment for K–2 science education before the main science content focus begins at grade three. There was significant discussion about the need for elementary school teachers to be held accountable at every grade level so that there is less pressure on the grade level chosen for an ESEA high-stakes assessment and to encourage science curriculum to be taught in every grade. However, the group discussions did not address how the assessments would promote accountability, other than being given at each grade level.

Another key discussion point among stakeholders was the feeling that science should be assessed at every grade level to provide students, parents/guardians, and teachers with information on growth within science learning progressions in a timely manner that would allow additional learning opportunities to be implemented before the ESEA high-stakes assessment occurred.

Table 6.2 summarizes stakeholder recommendations for additional assessments at grade level as described by *EC* Section 60640(c). Notes addressed both common themes and minority opinions of meeting participants/groups. Full results, including rationales, can be found in Table 5.1 through Table 5.8.

Table 6.2 Summary of Stakeholder Recommendations for Additional Assessments

	***EC* Section 60640(c), non-ESEA**											
	Meeting 1						**Meeting 2**					
Room	1						1					
Group	A			B			A			B		
Grade Level	3, 4	6, 7, 9	10	K–4	6, 7	9, 10 or 11, 12	4	6, 7, 9	10	3, 4	6, 7, 9	10, 11
Content	EOY	EOC/EOY	EOC	EOY	EOC/EOY	HS	EOY	EOC/EOY	EOC	EOY	EOC/EOY	EOC
Option	CAT	CAT	CAT	CBT	CBT	CBT	CAT	CAT	CAT	IB	IB	IB
Type	F	F	F	F	F	F	S	S	S	I	I	I/S

	EC Section 60640(c), non-ESEA											
	Meeting 1						Meeting 2					
Room	2						2					
Group	C			D			C			D		
Grade Level	4, 5	7, 8, 9	11, 12	3, 5	6, 7	10	K-3 or 4	6, 7, 9	10 or 11, 12	2, 3, 4	6, 7, 9	10
Content	EOY	EOC/EOY	EOC	EOY	EOC/EOY	EOC	EOY	EOC/EOY	HS	EOY	EOC/EOY	HS
Option	MS CAT	MS CAT	MS CAT	CAT	CAT/PT	CAT/PT	V	V	V	CBT/V	CBT/V	CBT/V
Type	F	F	F	S	S	S	F/S	F/S	F/S	F/S	F/S	F/S
Room	3						3					
Group	E			F			E			F		
Grade Level	3, 4	6, 7, 9	10	4, 5	7, 8	10, 12	3, 4	6, 7, 9	10	2, 3, 4	6, 7, 9	10 or 11, 12
Content	EOY	EOC/EOY	EOC	EOY	EOC/EOY	HS	EOY	EOC/EOY	EOC	EOY	EOC/EOY	HS
Option	CAT	CAT/PT	CAT/PT	CBT/CAT/PT	CBT/CAT/PT	CAT/CBT	CAT/PT	CAT/PT	CAT/PT	CAT/PT	CAT/PT	CAT/PT
Type	S	S	S	S	S	S	F/S	F	I	S	S	S

Assessment Key
Grade Levels For *EC* Section 60640(b) at least one assessment per grade span (GS) 3–5, 6–9, and 10–12; for *EC* Section 60640(c) additional K–12 assessments to those proposed in *EC* Section 60640(b)
Content: Subjects to be assessed, grade level (elementary school [ES], middle school [MS], high school [HS]) performance expectation (PE), end-of-course (EOC), end-of-year (EOY), Life Science (LS)
Options: Paper-pencil (P/P), computer-based testing (CBT), computer-adaptive testing (CAT) or multistage (MSg) CAT, item type (such as technology enhanced [TE]), locally scored performance-based task (PT), portfolio, Consortium-developed item bank (IB), varied (V)
Assessment Types: Formative (F), Interim (I), Summative (S)

6C. Summary of the Qualitative and Quantitative Data Gathered from the Stakeholder Discussions of Alternate NGSS Assessments Implemented Beyond ESEA-mandated Grade Spans Pursuant to *EC* Section 60640(c)

The 12 groups were also asked to recommend alternate assessments to meet the specialized needs of the one to two percent of the student population with the most significant cognitive disabilities by providing greater access to an assessment that helps measure how well they are achieving science content standards.

The stakeholders recommended that these assessments occur only at the same grade levels as those chosen to meet ESEA requirements to prevent students within this population from being overburdened. The majority of stakeholders also recommended assessments similar in style to the CMA and CAPA with the content focused on the NGSS "because we believe [the] NGSS [are] for all students" (from the Transcript of Meeting 2, J Table Group stakeholders' conversation).

Section 7: Results from the Online Survey

7A. Survey Background

Survey Administration

To obtain further input from both stakeholders who participated in a meeting and stakeholders who were unable to attend a meeting, ETS administered an online survey. The survey was launched on August 8, 2014; Section 7 analyzes the 422 responses that were received by August 20, 2014. An announcement e-mail with a URL to the survey was distributed to the following groups:

- Stakeholder meeting participants,
- Stakeholder meeting applicants unable to attend,
- LEA CAASPP Coordinators, and
- Individuals from organizations that represented stakeholder groups outlined in AB 484 who were originally contacted to recruit stakeholder meeting participants.

Recipients were encouraged to share the survey among their colleagues, fellow organization members, and any other individuals in California who might be interested in providing input.

Survey Details

The survey, which is presented in Appendix I, was separated into four parts. Part 1 focused on assessments pertaining to federal ESEA requirements. Part 2 focused on assessments pertaining to non-ESEA requirements. Part 3 focused on measurement considerations for testing. Part 4 elicited feedback on the science assessment system as a whole. The survey also included an optional demographic data section. In addition, survey respondents who indicated that they had attended the Science Stakeholder Meetings were asked to complete a brief evaluation of the meetings at the end of the survey. A summary of these meeting evaluations from meeting attendees can be found in Appendix K.

The survey included a variety of item types. There were two types of selected-response questions. Depending on the information elicited by the question, some selected-response questions allowed respondents to select, at most, only one option, whereas others allowed respondents to select as many options as applicable. The survey also included opportunities for the respondents to provide their rationale for their selections in their own words.

Process for Summarizing Survey Results

Subsections 7C, 7D, 7E, and 7F provide quantitative summaries of the respondents' selections as well as brief qualitative summaries of some of their rationales. The quantitative summaries describe the numbers of respondents who selected available options. After responses for the open-ended rationales were read, codes were developed that described the frequent themes, and then rationales were categorized by the codes. See Appendix L for a list of the codes. In some cases, respondents' rationales included multiple themes; these were counted for all applicable themes. The reported codes (common themes) and corresponding counts are preliminary evidence of respondents' rationales that might need to be replicated.

7B. Characteristics of Survey Respondents

A total of 422 stakeholders responded to the online survey. Of the 422, 74 (18%) attended one of the Science Stakeholder Meetings and 348 (82%) did not attend any of the meetings. As shown in Table 7.1, respondents represented a variety of stakeholder roles. Table 7.1 provides the breakdown of the survey respondents by primary stakeholder role using the categories provided in the survey question. The four most-selected roles were high school teacher, middle school teacher, K–12 administrator, and K–5 teacher. These roles made up 78 percent of the respondents' selections. Thirty-seven of the respondents selected "Other" and wrote in one or multiple roles, such as "parent and electrical engineer," "retired science educator," "curriculum coordinator," "district administrator," and "teacher ed professor." About 6 percent of the respondents did not select any role (i.e., did not respond to this survey question).

Table 7.1 Breakdown of Primary Stakeholder Roles of Survey Respondents

Primary Role as a Stakeholder	Count	Percent
High school (grades 9–12) teacher	114	27%
Middle school (grades 6–8) teacher	107	25%
K–12 administrator	64	15%
K–5 teacher	48	11%
Higher education expert	8	2%
Expert in teaching students with disabilities	6	1%
Parent	4	1%
Scientist, researcher, and/or engineer	4	1%
Expert in teaching English learners	2	0%
Measurement expert	2	0%
Other	37	9%
Missing	26	6%
Total	**422**	**100%**

Note: The percentages do not sum to 100 percent due to rounding.

The survey respondents can be further characterized by their gender and ethnicity. Table 7.2 shows that across all respondents, about two-thirds are female, 27 percent are male, and 7 percent chose not to respond. For the four most prevalent stakeholder roles (listed in the first four rows of the table), the breakdown is 61 percent to 83 percent female. The 37 respondents who identified with "other" roles also showed a similar breakdown by gender, with 68 percent female and 30 percent male.

Table 7.3 provides the cross-tabulation of science stakeholder role by ethnic background of the survey respondents. Across all respondents, 66 percent are White, 9 percent are Hispanic, and 4 percent are Asian. Some chosen ethnicities had small counts of fewer than 10 respondents and, for simplicity, were combined with the counts of respondents who selected the "Other" ethnicity option in the survey. The combined "Other" ethnic background group includes specifications such as "Black or African American" (n=6, 1%), "Asian White" (n=4, 1%), "mixed" (<1%), and "Pacific Islander" (<1%), among others. For the four most frequently selected stakeholder roles—teachers in K–5, middle, and high school as well as K–12 administrators—the ethnicity compositions are generally similar.

Table 7.2 Primary Science Stakeholder Role of Survey Respondents by Gender

	Female		Male		Missing		Total	
Primary Role	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
High school (grades 9–12) teacher	69	61%	44	39%	1	1%	**114**	**27%**
Middle school (grades 6–8) teacher	78	73%	27	25%	2	2%	**107**	**25%**
K–12 administrator	46	72%	18	28%	0	0%	**64**	**15%**
K–5 teacher	40	83%	7	15%	1	2%	**48**	**11%**
Higher education expert	5	63%	3	38%	0	0%	**8**	**2%**
Expert in teaching students with disabilities	6	100%	0	0%	0	0%	**6**	**1%**
Parent	4	100%	0	0%	0	0%	**4**	**1%**
Scientist, researcher, and/or engineer	1	25%	3	75%	0	0%	**4**	**1%**
Expert in teaching English learners	2	100%	0	0%	0	0%	**2**	**0%**
Measurement expert	0	0%	2	100%	0	0%	**2**	**0%**
Other	25	68%	11	30%	1	3%	**37**	**9%**
Missing	1	4%	0	0%	25	96%	**26**	**6%**
Total	277	66%	115	27%	30	7%	**422**	**100%**

Note: The percentages do not sum to 100 percent due to rounding.

Table 7.3 Primary Science Stakeholder Role of Survey Respondents by Ethnic Background

	White		Hispanic or Latino		Asian		Other		Missing		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
High school (grades 9–12) teacher	86	75%	7	6%	2	2%	13	11%	6	5%	**114**	**27%**
Middle school (grades 6–8) teacher	74	69%	12	11%	6	6%	10	9%	5	5%	**107**	**25%**
K–12 administrator	39	61%	9	14%	3	5%	11	17%	2	3%	**64**	**15%**
K–5 teacher	34	71%	5	10%	3	6%	2	4%	4	8%	**48**	**11%**
Higher education expert	6	75%	1	13%	0	0%	0	0%	1	13%	**8**	**2%**
Expert in teaching students with disabilities	4	67%	2	33%	0	0%	0	0%	0	0%	**6**	**1%**
Parent	3	75%	0	0%	0	0%	1	25%	0	0%	**4**	**1%**
Scientist, researcher, and/or engineer	4	100%	0	0%	0	0%	0	0%	0	0%	**4**	**1%**
Expert in teaching English learners	2	100%	0	0%	0	0%	0	0%	0	0%	**2**	**0%**
Measurement expert	2	100%	0	0%	0	0%	0	0%	0	0%	**2**	**0%**
Other	24	65%	3	8%	2	5%	5	14%	3	8%	**37**	**9%**
Missing	0	0%	1	4%	0	0%	0	0%	25	96%	**26**	**6%**
Total	**278**	**66%**	**40**	**9%**	**16**	**4%**	**42**	**10%**	**46**	**11%**	**422**	**100%**

Note: "Other" includes the respondents who selected "Other," made several selections, and those who selected an ethnic background with small counts, including Black or African American (*n*=6) and Asian White (*n*=4).

The percentages do not sum to 100 percent due to rounding.

7C. Summary of Part 1 Responses on ESEA-mandated CAASPP Assessments

In Part 1 of the online survey, respondents were asked a series of questions related to preferences for ESEA-mandated CAASPP science assessments (see Appendix I for survey questions). Within each of the three ESEA-mandated grade spans (grades three to five, grades six to nine, and grades ten to twelve), respondents first selected their preferred grade level and could then provide a rationale for the selected grade. They then selected what content domain(s), assessment type(s), and item type(s) they wanted for their selected ESEA grade-level test and could provide a rationale for their selections. The summary of these responses is arranged by grade span.

Grade Span Three to Five

Selection of Grade Levels:

Figure 7.1 illustrates the survey respondents' selections for the ESEA-mandated CAASPP science test in grades three through five. Out of the 378 respondents who selected a grade level, 279 (74%) selected grade five, with a close to even split between preferences for grades three and four.

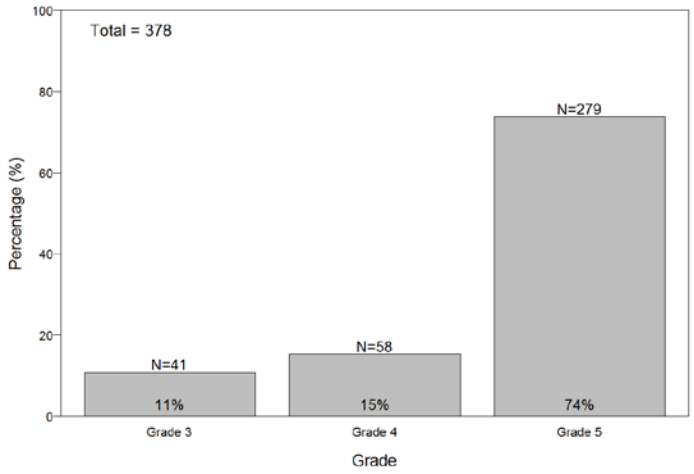

Figure 7.1 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Three to Five Span

A review of the rationales revealed a few common themes for each selection. Only 16 of the 41 respondents who selected grade three provided a rationale. The common themes in these rationales were:

- Early test will force science to be taught (*n*=7),
- Early test provides baseline (*n*=4), and
- Students are developmentally ready/have skills to take test at this grade (*n*=3).

The number of rationales exhibited for each of these themes is provided in parentheses next to the rationale in the itemized list above. Note that some respondents' rationales could have exhibited multiple themes, whereas others could have been unique and not fallen into any of these categories. Accordingly, the sum of the counts does not necessarily sum to the total number of provided rationales. This holds for all subsequent discussions of rationales as well.

Thirteen of the 58 respondents who selected grade four as their preferred grades three through five ESEA-mandated grade-level test also gave rationales. These rationales tended to mention the following common themes:

- Students are developmentally ready/have skills to take a test at this grade (*n*=5),
- Results inform next year of instruction (*n*=3), and
- Assessment will hold elementary school teachers accountable (*n*=3).

Only 36 out of the 279 survey respondents who selected grade five also provided a rationale for their selection. The most frequent reasons for selecting grade five were as follows:

- A grade five test would serve as a summative/capstone assessment looking back on elementary grades (*n*=15),
- Students are more mature so the test will better mirror their understanding (*n*=8), and
- Late-bloomers and English learners have a chance to develop so that the test better measures their science proficiency (as opposed to their reading ability) (*n*=6).

For the chosen grade level, respondents were then asked to provide their preferences for various characteristics of the assessment, including the content domain(s), test type(s), and item type(s). Given the majority preference for grade five, only the selections for this chosen grade-level assessment are summarized here. See Appendix J for a summary of these assessment characteristic selections for those respondents who preferred grade three or four for the ESEA-mandated test in the grades three to five span.

Selection of Content Domains:

Table 7.4 summarizes the selections for which content domain(s) to assess for those respondents who selected to test in grade five. Out of the total 279 respondents who selected grade five (see Figure 7.1), 277 made selections for their preferred content domains to assess. Table 7.4 gives the counts of these 277 respondents who selected each content domain option. Respondents were allowed to select as many content domain options as they wanted assessed. Ninety-three respondents selected more than one content domain option. Thus, a particular respondent could appear in multiple counts, making the sum of counts equal more than the total number of respondents and the percentages sum to more than 100 percent.

As shown in Table 7.4, the most frequently chosen option was Integrated Science with 181 (65%) selecting this content domain. Of these 181 respondents, 138 respondents selected only Integrated Science, whereas the remaining 43 also selected at least one other content domain (with the majority, *n*=38, selecting all content domain options). In addition to the 181 (65%) survey respondents who preferred to assess Integrated Science in grade five, 31 (11%) of the 277 respondents selected all three core disciplinary ideas (Biological Science/Life Science, Earth and Space Science, and Physical Science). It is not clear how an assessment covering all three core disciplinary ideas would differ from a test assessing Integrated Science that draws on all three core disciplinary ideas. Accordingly, the preference for Integrated Science may be even more than is shown in Table 7.4.

Table 7.4 Preferences of Content Domains for Respondents Who Selected Grade Five

Content	Count	Percent
Biological Science/Life Science	100	36%
Earth and Space Science	101	36%
Physical Science	96	35%
Integrated Science	181	65%
Total Respondents	**277**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

The rationales for content domain selections were given in response to a single survey question asking respondents to provide rationales for their selections of content domain(s), test type(s), and item type(s). A total of 212 respondents wrote a rationale. These 212 rationales were coded for main themes for each of the assessment characteristic selections. In some cases, respondents' rationales gave an overall motivation for all their selections, whereas others focused on one or more specific selections for content domain(s), test type(s), and item type(s). Accordingly, not all 212 rationales were relevant for all of these assessment characteristics. Those that were related to each assessment characteristic were reviewed and coded for common themes. Those rationales that provided an overall response primarily indicated that their selections promoted critical thinking (*n*=16) or that their selections for content domain(s), test type(s), and item type(s) matched/corresponded with the NGSS (*n*=12).

As the majority of respondents included Integrated Science among their preferred content domain selections, most of the content domain–related rationales were for this choice. The frequently provided rationales for these respondents who preferred Integrated Science are:

- Wants students to know basics across all disciplines or believes these content areas are foundational (*n*=40),
- Matches/corresponds with the NGSS (*n*=13),
- Wants content to cover full grade span (not just selected grade level within the ESEA grade span) (*n*=10), and
- Content reflects real science (*n*=8).

Selection of Test Types:

Respondents also made selections for preferred test type, selecting from "Computer adaptive," "Computer-based," "Paper-pencil," and "Other," with the option to write in a suggestion. As with the content domain survey question, respondents could select as many options as applicable. Table 7.5 summarizes the counts of each item option across all selections for all respondents.

As shown in Table 7.5, computer-based and computer-adaptive tests were each selected by more than half of the respondents. Further, from an analysis of the unique combinations of selections that respondents made, the top three most frequently selected combinations were computer adaptive (*n*=73, 26%), computer-based (*n*=56, 20%), and both computer adaptive and computer-based (*n*=36, 13%). The fourth most selected response, at 13 percent (*n*=35), was all three test types: computer adaptive, computer-based, and paper-pencil. Out of all 278 responses, 236 (85%) selected computer-based and/or computer adaptive among their selections.

Overall, 102 respondents included paper-pencil among their preferred test types, but only 27, or 10 percent, selected paper-pencil exclusively. Respondents were allowed to write in "other" possible test types. Some of these write-in test types were "hands-on," "project-based," "task-based," "performance[-based] assessment," and "lab portion," suggesting some interest in some type of "hands-on" component of the test.

Table 7.5 Preferences of Test Types for Respondents Who Selected Grade Five

Test Mode	Count	Percent
Computer-based	151	54%
Computer adaptive	162	58%
Paper-pencil	102	37%
Other	15	5%
Total Respondents	**278**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

The majority of rationales related to the test-type selection were given by respondents who selected computer-based and/or computer-adaptive tests (*n*=78). The main rationales for these computer test-type selections were:

- Provides better measure of student ability (*n*=26), and
- Takes advantage of technology (*n*=14).

Another somewhat common rationale for selecting a particular test mode was flexibility (*n*=15), but this motivation sometimes referred to having flexibility in test type (i.e., for those respondents who selected more than one mode) and other times referred to the flexibility a particular test type affords. Familiarity or appropriateness for examinee age was also a rationale that was frequently mentioned. In order to understand the test type respondents thought students are familiar with, this common rationale is broken down by test type selections: 33 of these familiarity rationales are for respondents with computer-based or computer adaptive among their selections, while 13 are for respondents with paper-pencil among their selections of test types.

Selection of Item Types:

Survey respondents were also presented with four item types as well as an "Other" option for the types of items they would like on their selected grade five test. As with content domain and test type, respondents could select as many options as they thought applicable. About 75 percent of the 277 responses for this item involved multiple selections. Table 7.6 provides the total counts of respondents who selected each of the possible item types. Selected-response/multiple-choice items were the most frequently selected item type at 71 percent, with task-centered items a close second with 69 percent. Constructed-response and technology-enhanced items were also each selected by more than half of the respondents.

As respondents generally selected more than one item type, their rationales tended to support the combination of their choices. These included:

- Allows assessing specific skills (*n*=44),
- A variety of item types is beneficial (*n*=22),
- Should follow Smarter Balanced ELA and mathematics example (*n*=17),

- Emphasizes hands-on/de-emphasizes memorization (*n*=13),
- Allows for assessing multiple levels of cognition (*n*=9),
- Allows for access to all students (*n*=8),
- Matches/corresponds with the NGSS (*n*=8), and
- Reflects authentic/real science (*n*=6).

Among the 44 respondents who articulated that their item-type selections allowed assessing specific skills, some specified one particular item type that was particularly good for this purpose: 12 said only task-centered, 8 said technology-enhanced, 5 said constructed-response, and 4 said selected-response/multiple choice. The remaining 15 (of the 44) specified more than one item type. The other pattern of interest was that for all the "emphasizing hands-on/de-emphasizing memorization" rationales, respondents included task-centered and/or technology-enhanced in their rationale, suggesting that respondents thought these types of items were best suited for assessing higher-order, critical-thinking skills and/or allowing for "doing" science.

Table 7.6 Preferences of Item Types for Respondents Who Selected Grade Five

Item Type	Count	Percent
Selected-response/multiple-choice items	198	71%
Technology-enhanced items	151	55%
Constructed-response items	168	61%
Task-centered items	190	69%
Other	7	3%
Total Respondents	**277**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Grade Span Six to Nine

Selection of Grade Levels:

Figure 7.2 illustrates survey-respondent selections for the preferred grade level to test for the ESEA-mandated test in grades six through nine. It clearly indicates that the preferred choice is grade eight, with 72 percent of the 374 respondents who selected a grade level selecting it. About 10 percent of the respondents selected each of the other three grade levels in the grades six to nine span.

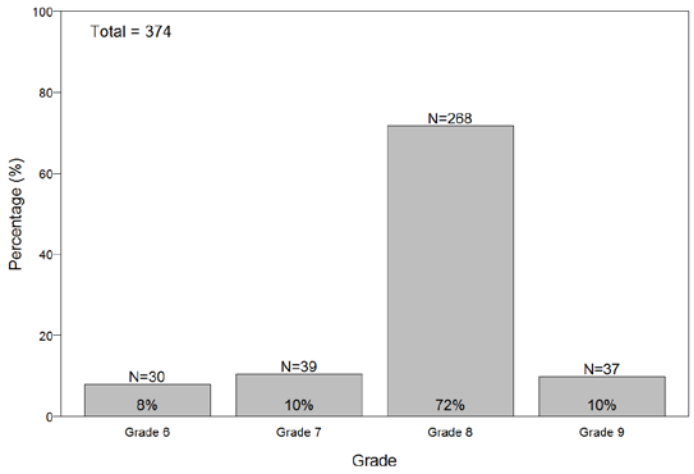

Figure 7.2 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Six to Nine Span

A review of the respondent-written rationales revealed several common themes. Out of the 30 respondents who selected grade six, 8 provided rationales. The common themes in these rationales were:

- A grade six test would serve as a summative/capstone assessment looking back on elementary grades (*n*=4), and
- Grade six is the first year of middle school, so a grade six test would give middle school teachers a platform to build on (*n*=3).

As seen also in the analysis of grade span three to five, some rationales include more than one common theme, whereas others are unique and do not fall into any common category, meaning the common rationales likely do not sum to the total number of rationales.

Thirty-three of the 39 respondents who selected grade seven provided rationales. The most frequently cited reasons for choosing this grade level were:

- Seventh grade testing allows for remediation/intervention in eighth grade (*n*=11),
- Seventh grade testing allows for eighth grade teachers to prepare students for high school standards (*n*=11), and
- Grade seven is the midpoint between middle school grades six to eight (*n*=3).

Of the 268 respondents who selected grade eight, 219 supported their choice. They typically included the following in their supporting statements:

- A grade eight test would serve as a capstone/summative assessment for middle schools (as most middle schools end at/have an eighth grade) (*n*=117),
- A grade eight test would inform high school instruction and/or placement of students (*n*=36),
- Testing at eighth grade ensures that all students who received instruction on either a domain-specific or integrated model would be prepared (*n*=15),

- By grade eight, students have exposure to all three disciplines (Biological Science/Life Science, Earth and Space Science, and Physical Science) (*n*=11), and
- Because grade nine is a high school grade level, choosing grade nine over grade eight means middle school would not be tested (*n*=7).

Thirty of the 39 respondents who selected grade nine for ESEA testing in the grades six to nine span also provided a rationale. These rationales in support of grade nine included the following common themes:

- Ninth grade students should know Earth Science (n=6),
- A ninth grade test allows for assessing middle school science learning (n=6),
- A ninth grade test serves as a benchmark to inform high school instruction (n=5), and
- Students are more mature (by grade nine) so the test will better mirror their understanding (n=3).

Given that the majority of survey respondents (who selected a grade level to test in the six-to-nine grade span) selected grade eight, only the preferences for the assessment characteristics for these respondents are given here. Preferences for assessment characteristics for respondents who selected grades six, seven, or nine are given in Appendix J.

Selection of Content Domains:

Table 7.7 summarizes the selections of content domain preferences for those who selected a grade eight ESEA test. Respondents were allowed to select as many content domains as they felt applicable; 171 selected only one content domain, while 93 selected more than one content domain. Of the 171 who selected only one content domain, 126 selected Integrated Science. An additional 48 respondents selected Integrated Science and at least one other content domain, resulting in a total of 174 (66%) of respondents recommending Integrated Science, as shown in Table 7.7. In addition, 28 respondents selected all three content domains that reflect the NGSS core disciplinary ideas of Biological Science/Life Science, Earth and Space Science, and Physical Science. It is not clear how these respondents believed an assessment covering all three core disciplinary ideas would differ from an assessment that covers Integrated Science, which draws on all three core disciplinary ideas. Biological Science/Life Science and Earth and Space Science were each selected by about 30 percent of the respondents, and Physical Science was selected by almost half of the respondents.

Table 7.7 Preferences of Content Domains for Respondents Who Selected Grade Eight

Content	Count	Percent
Biological Science/Life Science	84	32%
Earth and Space Science	75	28%
Physical Science	123	47%
Integrated Science	174	66%
Total Respondents	**264**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Of the 268 respondents who selected grade eight, 219 respondents also wrote a rationale supporting their assessment characteristic selections, including their content domain, test type,

and item-type choices. These rationales ranged from providing an overarching response supporting all of their selections generally to providing specific rationales for one or more of their chosen assessment characteristics. Accordingly, not all 219 rationales were relevant for each assessment characteristic. The rationales were reviewed for common themes related to each assessment characteristic they discussed. The rationales that gave general explanations for all selections across content domain, test type, and item type tended to indicate that either their selections promoted critical thinking (*n*=18) or they matched/corresponded with the NGSS (*n*=2).

As the majority of respondents included Integrated Science in their selections, most of the content domain–related rationales were for these respondents. The common themes for these respondents' supportive explanations for including Integrated Science in their selections were:

- Wants students to know basics across all disciplines or believes these content areas are foundational (*n*=68),
- Wants content to cover full grade span (not just selected grade level within the ESEA grade span) (*n*=27),
- Matches/corresponds with the NGSS (*n*=8), and
- Content reflects real science (*n*=8).

Several respondents who instead selected all three core disciplinary ideas (and excluded Integrated Science) also said they wanted students to know basics across all disciplines (*n*=10) or wanted the test content to cover the full grade span (*n*=6). For the respondents who selected Physical Science, the next most common content domain, their main rationale was that it constituted foundational knowledge (*n*=16).

Selection of Test Types:

Table 7.8 summarizes the preferences for test type for the grade eight ESEA test. As in grade five, the computer mode assessments were the most preferred; but for grade eight, more respondents selected computer adaptive (*n*=187) than computer-based (*n*=152). Among all combinations of selections, the most popular choice was for computer adaptive only (*n*=74, 28%), followed by both computer adaptive and computer-based (*n*=51, 19%), and only computer-based (*n*=42, 16%). Only 30% of respondents included paper-pencil among their selections. About half of the respondents selected only one test type, and the other half selected more than one test type. The 27 respondents who specified "Other" test type also wrote in specific suggestions, such as "hands on," "performance[-based] assessment," "with a physical lab portion," and "task based."

Most of the rationales related to supporting test type selections were for respondents who selected computer-based and/or computer adaptive among their selections. Their main rationales were:

- Provides better measure of student ability (*n*=22),
- Takes advantage of technology (*n*=6), and
- Mirrors/follows Smarter Balanced example (*n*=5).

As in grade five test-type rationales, some respondents (*n*=10) mentioned flexibility, with some indicating that the mode itself was flexible and others that specifying several test types allows for flexibility in testing. The familiarity or appropriateness of the test type for the

examinee age group was another common theme. Fifteen respondents who mentioned familiarity in their rationale had computer adaptive or computer-based among their selections, while six had paper-pencil among their selections.

Table 7.8 Preferences of Test Types for Respondents Who Selected Grade Eight

Test Mode	Count	Percent
Computer-based	152	58%
Computer adaptive	187	71%
Paper-pencil	78	30%
Other	27	10%
Total Respondents	**264**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Selection of Item Types:

In terms of item-type preferences, respondents selecting grade eight for the grades six to nine ESEA test tended to select more than one item type: 217 out of the 265 (82%) selected multiple item types. Table 7.9 shows that all item types were popular among these respondents. In fact, 37 percent (*n*=98) selected all four provided item-type choices. As shown in Table 7.9, the highest proportion of respondents selected task-centered items, but all the item types were selected by at least 65 percent of the respondents. Thirteen respondents also specified other item types, including "engineering practice," "task-centered should include inquiry/lab experience," "technology or live experiments," "integrate content and processes used," "NGSS practices-based," and "portfolios of student work." These suggestions are generally hands-on or performance-based activities.

Table 7.9 Preferences of Item Types for Respondents Who Selected Grade Eight

Item Type	Count	Percent
Selected-response/multiple-choice items	176	66%
Technology-enhanced items	179	68%
Constructed-response items	197	74%
Task-centered items	209	79%
Other	13	5%
Total Respondents	**265**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Respondents supported their multiple item-type selections with the following common rationales:

- Allows for assessing specific skills (*n*=24),
- Allows for assessing multiple levels of cognition (*n*=22),
- Ensures access to all students (*n*=14),
- Matches/corresponds with the NGSS (*n*= 9),
- Emphasizes hands-on/de-emphasizes memorization (*n*=9),

- Reflects authentic/real science (*n*=7), and
- Should follow Smarter Balanced ELA and mathematics example (*n*=3).

Grade Span Ten to Twelve

Selection of Grade Levels:

A total of 347 survey respondents selected one preferred grade level to assess in the ESEA-mandated grades ten to twelve span. As shown in Figure 7.3, the majority choice was for grade eleven, with 51 percent of the respondents selecting it. Grade ten was the second most-selected grade level with 27 of percent respondents, and grade twelve was a close third with 22 percent.

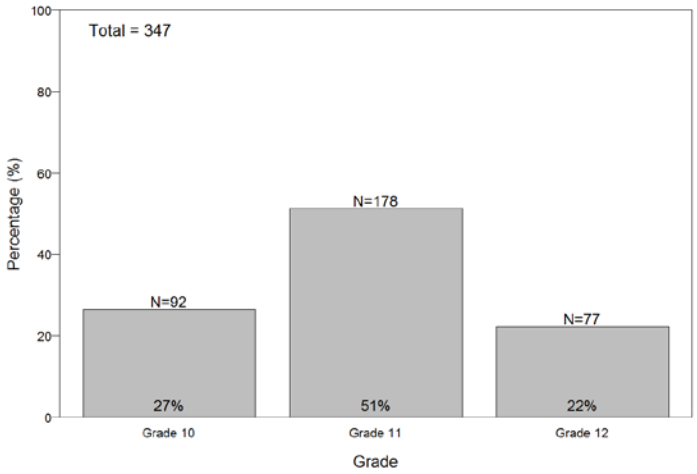

Figure 7.3 Barplot of Survey Respondent Selections for Grade Level to Test in the ESEA-mandated Grades Ten to Twelve Span

Of the 92 respondents who selected grade ten for ESEA testing, 63 respondents provided supporting rationales. A review of these rationales revealed the following common themes:

- Testing in grades eleven and twelve is undesirable (due to multiple testing in grade eleven and lack of student motivation in grade twelve or ability to use grade twelve results to inform instruction) (*n*=22),
- Testing in grade ten would correspond with the high school requirement for two years of science (so if students complete their science course requirements in grades nine and ten, the end of grade ten is appropriate for testing instead of waiting a year after they have no science instruction in grade eleven) (*n*=15),
- Testing in grade ten would be a continuation of current/past practices (*n*=5), and
- Most students would have Biology by grade ten so there will be common content to assess, which is often difficult to find in high school given the diversity of course trajectories (*n*=5).

For the 178 grade eleven supporters, 148 provided rationales. The common themes in these rationales were:

- Testing later (in high school) allows for more instruction in all domains of science and autonomy for students to choose their science courses (*n*=62),
- Testing in grade twelve is too late (due to lack of student motivation or ability to use grade twelve results to inform instruction) (*n*=41), and

- Allows for using test results in college admissions and provides students an additional year (grade twelve) to improve in areas in which they are deficient before attending college or pursuing a career (*n*=18).

Sixty-five of the 77 respondents who were in support of a grade twelve ESEA test supported their choice with a rationale. The common themes in these rationales were:

- Would serve as a capstone/summative assessment for all of K–12 science instruction and would assess the extent to which students can think scientifically before they move on to college or a career (*n*=28),
- Provides students with an incentive for taking four years of science instruction (*n*=18), and
- Grade twelve is not as heavily tested as grade eleven is (*n*=4).

Survey respondents were then asked to make assessment characteristics choices for content domain(s), test type(s), and item type(s). These selections are summarized for grade eleven as it was the most selected grade. The assessment characteristics selections for grades ten and twelve are summarized in Appendix J.

Selection of Content Domains:

Of the 178 respondents who selected grade eleven for ESEA testing, 173 selected which content domains they wanted assessed. Table 7.10 summarizes these selections. As with grades five and eight, the most selected content domain was Integrated Science, with 61 percent selecting it. However, for grade eleven, Biological Science/Life Science was also selected by about the same proportion of respondents.

It is of interest to note, however, that in general, respondents who selected Biological Science/Life Science also selected at least one other content domain; only 13 respondents selected this content domain exclusively compared to 61 respondents who selected Integrated Science exclusively. Similar to selections for grades five and eight, several respondents selected all four content domains (*n*=35), and several selected all but Integrated Science (*n*=28). Given that Integrated Science cuts across all three core disciplinary ideas, it is unclear how an assessment that assesses all three core disciplinary ideas differs from one that assesses Integrated Science, suggesting there may be further support for Integrated Science.

Table 7.10 Preferences of Content Domains for Respondents Who Selected Grade Eleven

Content	Count	Percent
Biological Science/Life Science	104	60%
Earth and Space Science	68	39%
Physical Science	89	51%
Integrated Science	105	61%
Total Respondents	**173**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Among all respondents who selected grade eleven and made selections for assessment characteristics, 137 responded to the survey prompt asking for a rationale for their assessment characteristic selections. These rationales varied in focus, with some giving a general, overall explanation for their selections and others providing specific explanations for one or more of

their assessment characteristic choices (for content domain, test type, and item type). Accordingly, not all 137 rationales were in support of selections for all three assessment characteristics. The rationales were reviewed for common themes for each assessment characteristic they referenced.

Some respondents provided an overall rationale for all selections. These included 17 respondents referring back to earlier rationales by saying "same as above," suggesting that some respondents felt that the same motivations for selecting assessment characteristics transcend the particular grade level. Six respondents articulated that their selections, in general, promoted critical thinking.

The respondents who included Integrated Science among their selections and provided a rationale (*n*=36) articulated the following common reasons:

- Wants students to know basics across all disciplines or believes these content areas are foundational (*n*=30),
- Wants content to cover the full grade span (not just a selected grade level within the ESEA grade span) (*n*=4), and
- Matches/corresponds with the NGSS (*n*=4).

Grade eleven supporters who included Biological Science/Life Science among their selections and provided a rationale (*n*=41) had a variety of rationales, with the most common themes being it (and any other content domains selected) represented foundational science content (*n*=16) and that Biology should be tested as it will be a common course that high school students would have taken by grade eleven (*n*=8).

Selection of Test Types:

Table 7.11 summarizes the test type selections for respondents who selected a grade eleven ESEA test. The most selected test type was computer-adaptive testing, with 77 percent of respondents selecting it, followed by computer-based testing, with 58 percent, and paper-pencil testing, with 28 percent. An analysis of the particular combinations of selections that respondents made reveals that 68 percent selected computer adaptive and/or computer-based exclusively, with 29 percent (*n*=50) selecting only computer adaptive, 26 percent (*n*=44) selecting both computer adaptive and computer-based, and 13 percent (*n*=22) selecting only computer-based. Some respondents (*n*=11) also wrote in "Other" test types, which generally indicated a type of hands-on or performance-based task; their write-ins included "task-oriented," "hands-on," "collaborative task," "hands-on test with scoring rubric," "practicum," "task," "performance[-based] with materials," and "student-designed and -executed experiment."

Table 7.11 Preferences of Test Types for Respondents Who Selected Grade Eleven

Test Mode	Count	Percent
Computer-based	100	58%
Computer adaptive	131	77%
Paper-pencil	48	28%
Other	11	6%
Total Respondents	**171**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Respondents supported their test type selections with the following common rationales:

- Provides better measure of student ability (*n*=16),
- Affords flexibility (*n*=13),
- Provides familiarity/appropriateness for age (*n*=4), and
- Takes advantage of technology (*n*=2).

All of these explanations corresponded to respondents who included computer-based and/or computer adaptive in their selections. The respondents who included flexibility in their rationale could mean flexibility in allowing examinees to choose from several test types and/or that a particular test type (e.g., computer adaptive) affords flexibility.

Selection of Item Types:

For item-type selections, 145 (84 %) of the 172 respondents who responded to this question selected more than one item type, with about 47 percent (*n*=80) selecting all four item types. Table 7.12 shows that about 85 percent of the respondents selected constructed-response items and task-centered items each, and almost three-fourths of the respondents selected technology-enhanced items. The least frequently selected item type was selected-response/multiple-choice, but even for this item type, 62 percent of respondents included it in their preferences. Respondents could also write in "Other" item types, but only four respondents did so, specifying "lab performance[-based] or simulation," "performance[-based] task," and "collaborative task."

Table 7.12 Preferences of Item Types for Respondents Who Selected Grade Eleven

Item Type	Count	Percent
Selected-response/multiple-choice items	107	62%
Technology-enhanced items	125	73%
Constructed-response items	144	84%
Task-centered items	147	85%
Other	4	2%
Total Respondents	**172**	

Note: Each percent is the count divided by the total number of respondents. The percents sum to more than 100 percent because survey respondents could select as many options as applicable.

Respondents generally selected several item types and provided the following rationales for these selections:

- Allows assessing specific skills (*n*=55),
- Emphasizes hands-on/de-emphasizes memorization (*n*=11),
- Reflects authentic/real science (*n*=11),
- Should follow Smarter Balanced ELA and mathematics example (*n*=5), and
- Matches/corresponds with the NGSS (*n*=2).

For the "allows assessing specific skills" rationale, some respondents specified a particular item type that was useful for this purpose: 15 (of the 55) respondents identified only constructed-response items, 7 only task-centered items, 7 only technology-enhanced items, and 5 only selected-response/multiple-choice items. The remaining 21 (of the 55) indicated that multiple item types were useful for assessing specific skills.

7D. Summary of Part 2 Responses on Additional CAASPP Assessments

Part 2 of the online survey generally asked respondents about additional CAASPP science assessments for non-ESEA uses. This section of the survey was further divided into two portions, with the first focusing on traditional/regular assessments and the second focusing on alternate assessments. Responses to each of these Part 2 survey sections are summarized separately here.

Additional CAASPP Assessments

For the first section of Part 2 of the online survey, respondents were asked to choose the grade levels in which they would like testing in addition to those grade levels they selected for ESEA-mandated tests in Part 1. For each selected grade level, respondents were then asked to select how content should be assessed, which content domain(s) should be assessed, and what assessment type(s) should be administered. Following these assessment characteristic selections for each grade level, respondents had the opportunity to write a rationale for all of their selections. Only a small sample (n=27 to 52) of the respondents provided rationales supporting selections for each grade level, and these rationales varied in their focus. Accordingly, detailed analysis of their common themes is not included in this report.

Table 7.13 provides the summary of selections for each grade level across the 312 respondents who responded to this survey question. Respondents were allowed to select as many grade levels as they were interested in having any testing, and most respondents (n=212, 68%) selected more than one grade level. Table 7.13 shows that about 30 to 40 percent of respondents indicated a preference for each of the grade levels, from grade three to grade eleven. Grade twelve has the lowest proportion of respondents, with only 24 percent selecting it. Table 7.13 indicates that there is at least some interest in testing each grade from grades three to twelve with no particular grade level receiving a majority vote.

Table 7.13 Summary of Selections of Non-ESEA Grade-level Tests

Grade	Count	Percent
3	124	40%
4	103	33%
5	120	38%
6	119	38%
7	119	38%
8	125	40%
9	126	40%
10	138	44%
11	107	34%
12	76	24%
Total Respondents	**312**	

Note: The percents do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents.

Table 7.14 summarizes respondents' selections for how content should be assessed in each selected non-ESEA test. Respondents were given the choices of "Integrated" and "Other" in all grades, and depending on the grade level, they were also given the choice of "End-of-year"

and/or "End-of-course." For instance, in grades three to five, the "End-of-course" option was not provided, and thus the table contains "NA" for "not applicable" in those cells. The survey included definitions of each of these ways of assessing content (see Appendix I for specific survey questions and provided definitions). Respondents were also allowed to select as many options as they felt appropriate. Specifically, this table reads, for example: Of the 123 respondents who selected grade three and provided a response to this follow-up question, 66 (54%) selected Integrated among their choices.

Table 7.14 shows that the choices for how content should be assessed tend to vary by grade level. However, similar proportions of preferences were made for grades within elementary school (grades three through five), middle school (grades six through eight), and high school (grades nine through twelve). In the elementary grades, about 50 to 58 percent of respondents selected Integrated and end-of-year, suggesting interest in having assessments that test content over multiple grades (Integrated). For the middle school grades, about 35 to 48 percent selected each of the provided options, Integrated, end-of-year, and end-of-course. Only in high school grades did respondents show a more clear preference for a single way of assessing content, with 73 to 82 percent selecting end-of-course among their choices.

Table 7.14 Summary of Selections for How Content Should Be Assessed in Non-ESEA Tests

	Integrated		End-of-year		End-of-course		Other		Total
Grade	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Respondents
3	66	54%	67	54%	NA	NA	7	6%	**123**
4	49	50%	57	58%	NA	NA	3	3%	**98**
5	62	55%	65	58%	NA	NA	6	5%	**113**
6	50	45%	48	43%	44	39%	3	3%	**112**
7	41	37%	49	44%	54	48%	4	4%	**112**
8	55	46%	43	36%	57	48%	7	6%	**120**
9	47	39%	NA	NA	89	74%	5	4%	**120**
10	44	34%	NA	NA	107	82%	5	4%	**130**
11	42	41%	NA	NA	75	73%	4	4%	**103**
12	31	43%	NA	NA	54	75%	6	8%	**72**

Note: The percentages across rows do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents in that row.

Respondents could also select which content domain(s) they believed should be assessed for each of their selected non-ESEA grade-level tests. The possible choices were the same as were provided for the ESEA-mandated tests—Biological Science/Life Science, Earth and Space Science, Physical Science, and Integrated Science—and respondents could select as many as applicable. Table 7.15 summarizes these selections by grade level.

For grades three through five (elementary school grades), the majority of respondents included Integrated Science among their selections, with 67 to 75 percent selecting it within each of these grade levels. For the middle school grades six through eight, the most preferential content domain differed by grade level. In grade six, the preferences were for Integrated Science followed by Earth and Space Science. In grade seven, both Integrated and Biological Science/ Life Science were picked by about 50 percent of the respondents (who chose grade seven and responded to this survey item). For grade eight, a similar pattern is seen, but for Integrated Science and Physical Science. The distributions also differ by high school grade level, although

for both grades nine and ten there is a majority preference for Biological Science/Life Science. In grade eleven, the highest preference is for Physical Science, and in grade twelve, all content domains have at least 55 percent of respondents selecting them, meaning that respondents generally selected more than one content domain and there is interest in assessing content across multiple core disciplinary ideas.

Table 7.15 Summary of Selections for What Content Domain(s) Should Be Assessed in Non-ESEA Tests

	Biological Science/ Life Science		Earth and Space Science		Physical Science		Integrated Science		Total
Grade	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Respondents
3	39	33%	33	28%	29	24%	90	75%	**120**
4	32	34%	40	43%	29	31%	63	68%	**93**
5	37	33%	41	37%	40	36%	75	67%	**112**
6	32	30%	54	50%	30	28%	64	60%	**107**
7	58	53%	26	24%	27	25%	56	51%	**109**
8	40	34%	40	34%	69	58%	64	54%	**118**
9	60	54%	54	48%	36	32%	50	45%	**112**
10	69	56%	33	27%	55	45%	46	37%	**123**
11	36	37%	36	37%	61	62%	47	48%	**98**
12	36	55%	37	56%	43	65%	39	59%	**66**

Note: The percentages across rows do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents in that row.

The options for type of assessment differed somewhat from those provided for the ESEA-mandated tests in Part 1 of the survey. Given that ESEA-mandated tests are summative, respondents did not have to specify whether they wanted summative, formative, and/or interim testing for their selected ESEA tests in Part 1, although it is useful to note that summative tests can sometimes be used for formative purposes.

However, for Part 2, respondents were asked for any additional science tests they would like, including specifying for what purpose(s) they will be used. The test-type questions in Part 2 also included "Computer-based," "Computer adaptive," "Paper-pencil," and "Other."

Table 7.16 summarizes the selections for test type for each selected non–ESEA-tested grade level. Among the computer-based, computer adaptive, and paper-pencil selections, respondents favored either of the computer mode test types over paper-pencil across all grade levels. Among the summative, formative, and interim selections, respondents generally favored summative and formative over interim. Given that the percents sum to a number greater than 100, there is a general interest in having assessments that serve multiple uses or distinct assessments for each use.

Table 7.16 Summary of Selections for Which Type(s) of Assessments Should Be Administered in Each Non-ESEA Grade-level Test

	Computer-based		Computer adaptive		Paper-pencil		Summative		Formative		Interim		Other		
Grade	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Total**
3	60	49%	67	54%	43	35%	53	43%	61	50%	21	17%	4	3%	123
4	45	45%	55	56%	34	34%	41	41%	45	45%	24	24%	3	3%	99
5	57	50%	64	56%	35	31%	60	53%	40	35%	21	18%	9	8%	114
6	54	48%	70	63%	31	28%	55	49%	46	41%	17	15%	4	4%	112
7	61	54%	73	65%	29	26%	57	50%	45	40%	21	19%	5	4%	113
8	66	55%	81	67%	37	31%	68	56%	49	40%	23	19%	7	6%	121
9	62	53%	76	65%	30	26%	70	60%	44	38%	25	21%	6	5%	117
10	70	54%	78	60%	33	25%	68	52%	47	36%	25	19%	9	7%	130
11	51	52%	68	69%	24	24%	55	56%	41	41%	13	13%	5	5%	99
12	46	66%	51	73%	23	33%	46	66%	25	36%	16	23%	6	9%	70

Note: The percentages across rows do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents in that row.

Alternate Assessments

The second section of Part 2 of the online survey asked respondents to select additional grade levels (from those selected for ESEA testing in Part 1) for which they would like testing, specifically for students with severe cognitive disabilities who are currently tested with the CAPA. After making the grade-level selections, they were then asked how content should be assessed and given space to write in a rationale for their selection. Only 12 to 25 respondents provided rationales for this selection at each grade level. Accordingly, analysis of these rationales is not given in this report.

As shown in Table 7.17, only 227 of the total 422 survey respondents selected any additional (non-ESEA) grade levels for assessing students with severe cognitive disabilities. Respondents were allowed to select as many grade levels as they felt applicable. Just less than half (*n*=109, 48%) selected only one grade level, and just over half (*n*=118, 52%) selected more than one grade level. There was no majority preference for any particular grade level, but grade eight had the highest proportion at 41 percent. As Table 7.17 shows, all other grade levels had between 13 and 33 percent, with grades three and four having the lowest respondent preference.

Table 7.17 Summary of Selections of Grade Levels to Test Students with Severe Cognitive Disabilities for Non-ESEA Purposes

Grade	Count	Percent
3	29	13%
4	30	13%
5	75	33%
6	48	21%
7	39	17%
8	92	41%
9	45	20%
10	49	22%
11	61	27%
12	42	19%
Total Respondents	**227**	

Note: The percents do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents.

As in the first section of Part 2 of the online survey, respondents were subsequently asked how content should be assessed for each selected grade level. In this case, respondents were always provided with the same three options: “Integrated,” “End-of-year,” and “Other.” For all grade levels, the “Other” option was rarely selected. As shown in Table 7.18, the majority preference for Integrated or end-of-year varies by grade level. Note that the sum of the Integrated and end-of-year selections exceeds 100 percent for each grade level as some respondents selected both of these ways of assessing content.

Table 7.18 Summary of Selections for How Content Should Be Assessed for Students with Severe Cognitive Disabilities on Additional, Non-ESEA Grade-level Tests

	Integrated		End-of-year		Other		Total
Grade	Count	Percent	Count	Percent	Count	Percent	Respondents
3	19	70%	11	41%	1	4%	27
4	20	71%	12	43%	0	0%	28
5	34	48%	45	63%	1	1%	71
6	27	61%	21	48%	0	0%	44
7	15	42%	28	78%	0	0%	36
8	42	49%	55	65%	2	2%	85
9	22	51%	25	58%	3	7%	43
10	23	51%	27	60%	1	2%	45
11	30	52%	35	60%	2	3%	58
12	14	36%	30	77%	0	0%	39

Note: The percents do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents in that row.

7E. Summary of Part 3 Responses on Measurement Considerations

In Part 3 of the online survey, respondents were asked to reflect on a few measurement considerations related to test administration sampling designs of test items and examinees, and what scoring procedure should be used for open-ended test questions. They were then asked to provide rationales for each of their selections. These selected and open-ended responses are summarized in the following subsections.

Test Administration Sampling Designs

Respondents were asked to consider two test administration sampling designs: matrix sampling, which involves assigning students different subsets of items that represent portions of the tested standards, and population sampling, which involves selecting a representative sample of students within a grade level to take the assessments each year. They were provided with definitions of each of these designs (see Appendix I for full definitions and specific survey questions) and asked whether each design should be used in administrating CAASPP science assessments. They could select “yes,” “no,” or “not sure” exclusively (they could not select more than one option) and were then asked to provide a rationale for their selection.

Figure 7.4 provides the breakdown of responses to the survey questions on using matrix sampling (Panel A) and population sampling (Panel B). Of all 422 survey respondents, 410 responded to the matrix-sampling question and 408 to the population-sampling question. For matrix sampling, the responses were almost evenly divided among the three options, whereas for population sampling, the majority of respondents (57%) were against population sampling.

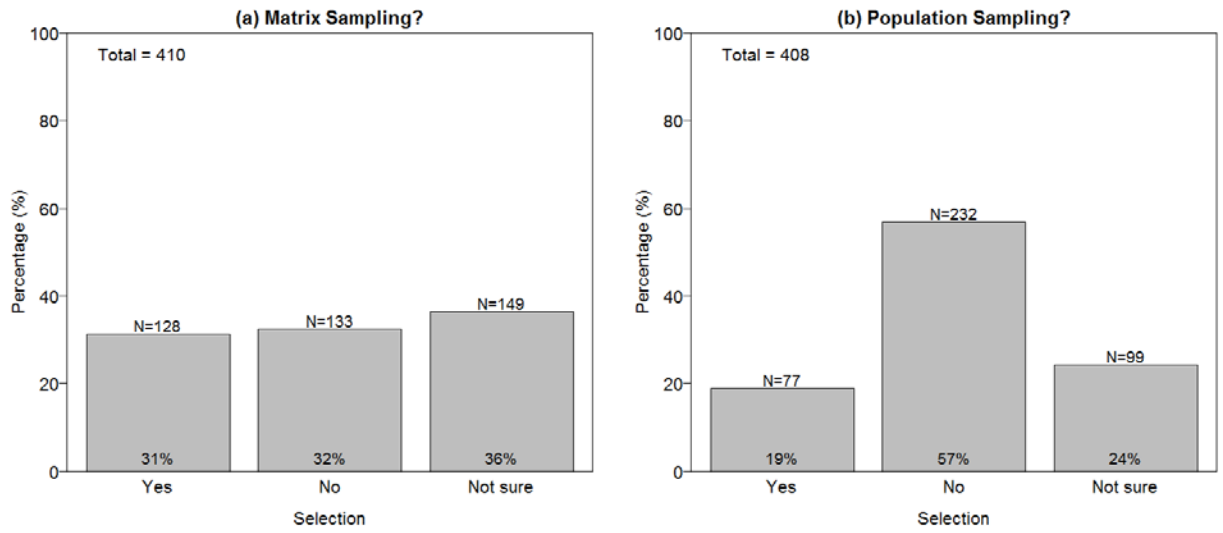

Figure 7.4 Barplots Showing Breakdown to Responses to Survey Questions on Using Matrix Sampling in CAASPP Science Assessments (Panel A) or Population Sampling (Panel B)

The rationales for the preferences for matrix sampling and population sampling were each analyzed given the response—all the rationales for "yes" to matrix sampling were analyzed together and then all for "no." Rationales for "not sure" tended to simply reiterate that respondents were not sure and/or did not have enough information or expertise to provide an informed selection. Common themes were then identified within each group of rationales, and rationales were coded by which common themes they included: some rationales specified several common themes and others were unique and did not fall within any of the common themes. The same procedure was then used for analyzing the rationales for population sampling.

Of the 128 respondents who indicated matrix sampling should be used, 87 provided rationales. The most common themes that appeared in these rationales were:

- Lowers the testing burden (*n*=28),
- Useful to use to inform aggregate decisions such as program evaluation (*n*=28),
- Allows for testing more standards and/or can better assess the NGSS (*n*=22),
- Provides more valid, accurate, or statistically sound results (*n*=8),
- Allows for including more complex tasks in the assessments (*n*=6), and
- Allows for depth over breadth (*n*=5).

Of the 133 respondents who indicated matrix sampling should not be used, 104 also gave rationales. The common themes that appeared for not using matrix sampling were:

- Values giving individuals scores, identifying individual strengths/weaknesses, and tracking student growth, but has concerns that matrix sampling would preclude such inferences (*n*=29),
- Has concerns with accuracy and fairness of sampling (e.g., that certain types of students would receive certain standards) (*n*=22),
- Values testing students on the *same* standards (with the same test) (*n*=15),
- Values testing *all* students on *all* standards (*n*=10),

- Values using test scores to inform instruction, but has concerns that matrix sampling would preclude such test use (*n*=10),
- Values fair comparisons among students and the belief that matrix sampling does not allow for comparability (*n*=10), and
- Has concerns that matrix sampling is not accurate for small samples (*n*=4).

For the population sampling question, only 77 respondents indicated it should be used and only 48 provided a rationale for why they supported its use. The common themes in these rationales in favor of population sampling were:

- Provides information on key demographic groups and promotes equity (*n*=19),
- Is cost effective (*n*=4),
- Reduces the testing burden (*n*=4),
- Informs instructors and curriculum developers (*n*=4), and
- Informs aggregate-level decisions (*n*=4).

In addition, three respondents' rationales revealed that they mistakenly believed population sampling meant sampling all students (i.e., the full population) or census testing.

Of the 232 respondents who were against using population sampling, 164 explained their choice. The following common themes emerged from these rationales:

- Values testing *all* students (*n*=59),
- Has concerns on accuracy, fairness, and equity of sampling (e.g., belief that it is unfair to generalize performance of a group based on a selected subset of that group) (*n*=55),
- Values providing feedback to students, teachers, schools, or LEAs, but has concerns that population sampling would preclude this use of test score data (*n*=28),
- Values using test scores to inform instruction, but has concerns that population sampling would preclude such test use (*n*=11),
- Suggests that instead of using population sampling, data analysts/researchers can sample from test scores *after* testing all students (*n*=10),
- Has concerns that population sampling complicates test administration (e.g., what to do with non-test-takers during testing periods) (*n*=7),
- Has concerns that it places the testing burden on the selected subset (*n*=6),
- Has concerns that it is just politics or a political game (*n*=4),
- Has concerns that it de-motivates students to perform well on the test and/or in science class (*n*=4), and
- Has concerns on not getting information on the subset that was not tested (*n*=4).

Scoring Procedures

In addition to questions about test administration sampling designs, respondents were asked which scoring procedure they thought should be used for scoring open-ended items on the CAASPP science assessments. Respondents could select one option among five choices: automated scoring, centralized scoring, remote scoring, local scoring, and other. They were then asked to provide a rationale for their selection. As with the rationales for the test administration

sampling designs, the scoring procedure rationales were grouped based on response and then analyzed for common themes.

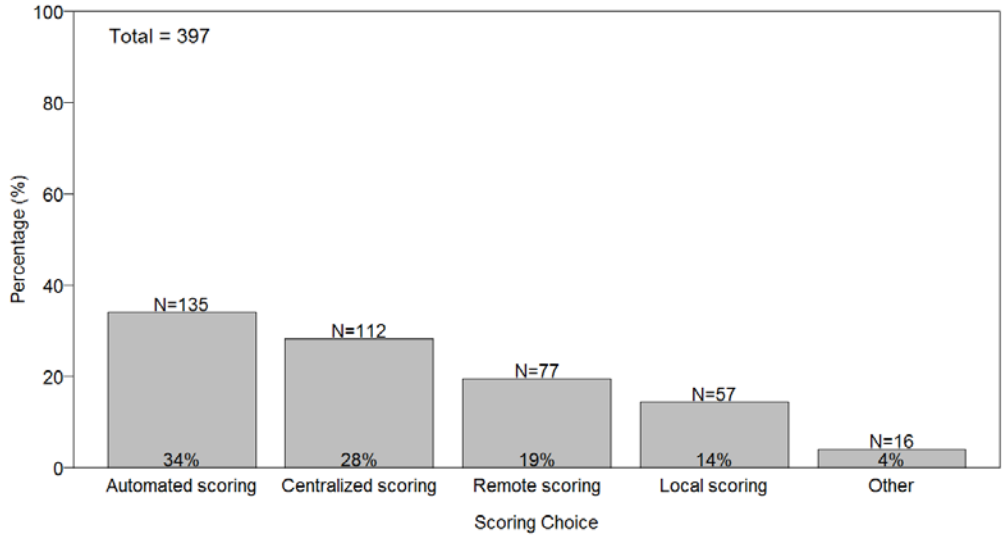

Figure 7.5 Barplot Showing Breakdown of Preferred Scoring Choice for Open-ended Test Items

Figure 7.5 gives the breakdown of respondent selections to this survey question on preferred scoring procedure. The scoring procedure options are ordered from most to least preferred in Figure 7.5. This figure shows that the most preferred scoring choice was automated scoring, with 135 (34%) out of the 397 who responded to this question selecting it. The next most preferred scoring choice was centralized scoring, with 28 percent, followed by remote scoring with 19 percent. Local scoring had the lowest support among the four provided scoring procedures. Sixteen (4%) of respondents selected "Other" and wrote in their preference. These preferences generally mentioned some combination of the four provided scoring procedures or had responses like "not sure," "depends on the questions," "depends on the reliability," "regional scoring," and "live scoring." Given the diversity of responses and that only 13 respondents provided rationales for their "Other" choice, further analysis of these rationales is not provided.

Of the 135 respondents who selected automated scoring, 76 provided rationales. The most common themes in these rationales in support of automated scoring were:

- Provides more fair/objective (or less biased) scoring (*n*=28),
- Provides faster/more timely feedback (*n*=28),
- Is cost effective (*n*=12),
- Is better than local scoring in that it can provide invalid/biased results and subjectivity in scoring (*n*=6),
- Is sophisticated enough now and reaching reliability levels of humans (*n*=5), and
- Alleviates burden on local teachers to score (*n*=4).

For centralized scoring, the next most popular scoring procedure, 77 out of the 112 who selected it also gave explanations. The main reasons for selecting centralized scoring were:

- Promotes training of raters and working together (*n*=19),
- Provides more fair/objective (or less biased) scoring (*n*=16),

- Expresses distrust in automated scoring (*n*=14),
- Is better than local scoring in that it can provide invalid/biased results and subjectivity in scoring (*n*=10),
- Believes that centralized scoring is used for scoring AP and/or the Golden State Exams (*n*=9),
- Provides faster/more timely feedback (*n*=4), and
- Is easier to monitor (*n*=4).

Remote scoring was preferred by 77 respondents, and 59 of them explained this preference. The common themes that arose in these rationales in support of remote scoring were:

- Minimizes bias/more consistent/less subjective (*n*=22),
- Values human raters and is wary of automated scoring (*n*=11),
- Is cost effective (especially in comparison with centralized scoring, as there are no travel or lodging expenses) (*n*=10),
- Is better than local scoring that can provide invalid/biased results and subjectivity in scoring (*n*=8),
- Allows more eligible raters to participate (as there are no geographical constraints) (*n*=8),
- Believes that it is used and works with College Board/AP scoring (*n*=5),
- Is the most flexible scoring option (*n*=5), and
- Provides faster/more timely feedback (*n*=4).

Although at least some respondents who were in favor of automated, centralized, or remote scoring expressed distrust for local scoring, 57 of the survey respondents selected it and 33 provided rationales supporting their choice. These rationales had the following common themes:

- Allows for geography and demographic composition to be taken into account (*n*=8),
- Provides feedback to teachers (*n*=8),
- Provides faster/more timely feedback (*n*=6),
- Is wary of/does not trust automated scoring (*n*=4),
- Involves training and oversight (*n*=4),
- Believes teachers know their students best (*n*=3),
- Does not trust centralized scoring (*n*=3), and
- Provides professional development to teachers (*n*=3).

7F. Summary of Part 4 Responses on Overall Feedback on the Future Science Assessment System

In Part 4 of the online survey, respondents were asked to provide overall feedback on the future CAASPP science assessments. This part of the survey included one selected-response item and two open-ended items asking respondents to express any other considerations they had on this future assessment system. The two open-ended questions were analyzed together as respondents tended to provide their considerations in one or the other of the provided text boxes.

The selected-response item asked respondents to select what their most important considerations were in the design of the California science assessments. Table 7.19 provides the options that respondents were given and the counts of respondents who selected each. For this item, respondents were allowed to select as many options as they felt were important to them. In general, respondents selected more than one important consideration: 70 percent selected multiple considerations. The only option that was selected exclusively was the most-selected consideration of "Including items that closely represent real-life science scenarios and thinking processes." Of the 361 of respondents (89%) who included this consideration among their selections, 121 selected it exclusively. The second most-selected response among respondent selections was "Reducing testing time for students," with 193 respondents selecting it. The "Assessing each tested student on the entire range of California NGSS for grade (grade span)" option was selected by 111 respondents, and the "Maximizing the number of grade levels that are assessed" option was selected by the fewest number of respondents, 86.

Table 7.19 Summary of Selections of Important Considerations for the Future CAASPP Science Assessments

Important Considerations	Count	Percent
Including items that closely represent real-life science scenarios and thinking processes	361	89%
Reducing testing time for students	193	47%
Assessing each tested student on the entire range of California NGSS for grade (grade span)	111	27%
Maximizing the number of grade levels that are assessed	86	21%
Total Respondents	**407**	

Note: The percents do not sum to 100 percent because survey respondents could select as many options as applicable. The percent is the count divided by the total number of respondents.

For the open-ended "other considerations" questions, 172 respondents provided responses. These responses had the following common themes:

- Emphasizes testing twenty-first century skills/real-life scenarios and skills (*n*=33),
- Emphasizes problem solving/critical thinking in assessments (*n*=18),
- Wants assessments like the Golden State Exams/performance-based/labs/practicum assessments (*n*=14),
- Wants supports for student learning/formative purposes (*n*=14),
- Emphasizes *not* taking time away from instruction/spend less time testing (*n*=14),
- Emphasizes attention to special groups such as English learners and accessibility such as keyboards skills/equity issues (*n*=11),

- Emphasizes *not* testing facts (*n*=9),
- Wants to test *all* students in *all* grades (*n*=7),
- Emphasizes college and career readiness skills (*n*=6),
- Wants timely turnaround of score reporting (*n*=5),
- Emphasizes *not* focusing on particular content domains (*n*=4),
- Emphasizes testing science earlier in elementary school so science would get taught (*n*=3), and
- Emphasizes providing useful information to schools and parents/guardians (*n*=3).

Section 8: Suggestions for Interpretation and Development of Recommendations

Through the CAASPP Science Stakeholder Meetings and online survey, stakeholders from across California had the opportunity to provide their input on various aspects of a new California science assessment system. The group discussions at the meetings and survey responses suggest that California science assessment stakeholders, including parents/guardians, educators, administrators, experts in assessing English learners or students with disabilities, and higher education experts are all invested in having a rich California science assessment system that is aligned to the California NGSS. Although stakeholders brought their own expertise and priorities to bear in the group discussions and survey responses, several common recommendations and rationales surfaced.

8A. Suggestions for Federally Mandated ESEA Testing

Suggested Grade Levels

For the federally mandated (ESEA) testing in science for the three grade spans—grades three to five, six to nine, and ten to twelve—the meeting stakeholder group and survey respondents considered which grades to assess, what content to assess, what type of test to administer, and which item types to include. Over the 12 meeting groups and 422 survey respondents, the most frequently recommended grade levels within each grade span were grades five, eight, and eleven, respectively. For both grades five and eight, an often-cited rationale across the discussion groups and survey respondents was that these tests would serve as capstone/culminating/ summative assessments of elementary and middle school science instruction. Supporters of ESEA testing in grade eleven often articulated that this would allow students to receive more of their required high school science instruction, or to have completed it altogether.

Suggested Content Domain

Both the groups at the meeting and individual survey respondents tended to favor integrated science assessments across grades and content domains. In the meeting discussions of what content to assess on the ESEA tests, groups tended to favor assessing California NGSS performance expectations over all the grades within a particular ESEA grade span as opposed to grade-specific performance expectations. Survey respondents were asked to select specific content domains to assess. They typically included Integrated Science in their selections or selected all three content domains that correspond with the NGSS core disciplinary ideas (Biological Science/Life Science, Earth and Space Science, and Physical Science) because they generally wanted students to have foundational knowledge across all core disciplinary ideas. Survey respondents supporting grade eleven testing also favored assessing Biological Science/Life Science, a common high school science course, in addition to Integrated Science.

Suggested Test Types

In general, stakeholders at the meetings and individual survey respondents both preferred computer-delivered assessments over paper-pencil tests. Specifically, the meeting groups showed a strong preference for computer-adaptive testing for providing potentially shorter tests and more precise scores. Similarly, survey respondents showed a strong preference for both computer-adaptive and computer-based testing. Some of the meeting groups and survey respondents also

expressed interest in having a paper-pencil option for testing (in addition to a computer-delivered test).

Suggested Item Types

To best assess the three dimensions of the NGSS, meeting groups generally favored performance-based "hands-on" and "virtual" tasks with limited use of discrete multiple-choice items. Survey respondents also expressed an interest in such performance-based tasks and de-emphasized including items that only require memorization of facts. They also showed strong preferences for including a variety of item types—constructed-response, selected-response, task-centered, and technology-enhanced items—to provide access to all students and best assess multiple levels of cognition.

8B. Suggestions for Non-ESEA Testing

Meeting stakeholder groups and survey respondents also provided feedback on additional, non-ESEA testing. These recommendations are more diverse in their grade-level preferences, specific content to assess, test types, and item types. Overall, there is interest in including summative, formative, and interim non-ESEA testing or tools (e.g., item banks) to inform instruction and provide information on students' science proficiency as they progress through their K–12 science instruction.

8C. Suggestions for Administering Alternate Assessments

Meeting stakeholder groups and survey respondents were also asked to provide feedback on administering alternate assessments to students with severe cognitive disabilities. The meeting groups generally recommended assessing this student group only at the same grade levels as those chosen to meet ESEA requirements to reduce the testing burden and to use tests similar to the current CMA and CAPA. Only about half of the survey respondents (227 out of 422) selected any grade level for additional, non-ESEA testing for this student group. No grade level was selected by a majority of these respondents: grades three and four were selected by the lowest proportion with 13 percent and grade eight with the highest proportion at 41 percent.

8D. Conclusion

Overall, California science stakeholder meeting groups and individual survey respondents often expressed similar preferences for a new California science assessment system. In addition, the meeting discussions and survey respondent rationales typically touched on several of the same underlying reasons for particular preferences. Given that only 18 percent (74 out of 422) of the survey respondents also attended one of the meetings, the common recommendations from these two events are not simply due to shared experiences, but rather, reflect the primary considerations and values of a large portion of the California science stakeholder community.

In summary, for ESEA testing, stakeholder meeting groups and survey respondents primarily recommend testing in grades five, eight, and eleven using a computer-delivered, integrated science assessment with a variety of item types that allow for students to demonstrate proficiency in science.

Appendix A: Organizations Contacted for Participant Recruitment

Organizations that were contacted to recruit meeting participants:

- American Association for the Advancement of Science (AAAS)
- American Association of Physics Teachers (AAPT)
- Association for Science Teacher Education (ASTE)
- Bechtel
- California Alliance of African American Educators (CAAAE)
- California Association of Bilingual Educators (CABE)
- California Association of Resource Specialists (CARS+)
- California Educational Research Association (CERA)
- California English Language Development Test (CELDT) District and Site Coordinators
- California Parent Teacher Association (PTA)
- California Science Teacher Association (CSTA)
- Chevron
- National Association for Research in Science Teaching (NARST)
- National Association of Biology Teachers (NABT)
- National Earth Science Teachers Association (NESTA)
- Project Lead The Way (PLTW)
- Regional Assessment Network (RAN)
- Science Expert Panel (SEP)
- Special Education Local Plan Area (SELPA)
- Technical Advisory Group (TAG)

Appendix B: Transcript of the Participant Application

2014 CAASPP Science Stakeholder Meeting Application

The California Department of Education (CDE), in collaboration with Educational Testing Service (ETS), is gathering input from stakeholders regarding science assessments aligned to the newly adopted science standards, called the Next Generation Science Standards (NGSS).

The input from stakeholders will be shared with State Superintendent of Public Instruction Tom Torlakson as he prepares recommendations for the California State Board of Education (SBE) for the new science assessments.

Two 2-day meetings will be held at the Hilton Arden West Hotel in Sacramento. The first meeting is scheduled to take place on **July 15 and 16, 2014,** and the second meeting is scheduled to take place on **July 17 and 18, 2014**. Each meeting day will be approximately eight hours long. Participants will be expected to attend both days of the two-day meeting. Travel and other expenses related to your participation will be provided.

If you are interested in participating in a meeting, please proceed with the application. If you have any questions, please contact the ETS CAASPP Program Coordinator, by e-mail or by phone.

Personal Information

Name

First Name Last Name Suffix

E-mail

Mailing Address

City State Zip Code

Phone

Which of the following best describes your role as a stakeholder? (Please check all that apply.)

- ☐ California K–12 teacher
- ☐ California K–12 administrator
- ☐ Higher education expert
- ☐ Expert in assessing English learners
- ☐ Expert in assessing students with disabilities
- ☐ Measurement expert
- ☐ Parent
- ☐ STEM professional
- ☐ Scientist, engineer and/or researcher
- ☐ Other:

Do you have any children currently enrolled in a public school in California?
[This question only appears if the "Parent" option is marked for "Which of the following best describes your role as a stakeholder?"]

● Yes ● No

Are you currently teaching or have you taught at a K–12 school in California?

● Yes ● No

Are you currently teaching or have you taught at a college/university level?

● Yes ● No

Organizational Affiliations: (Please check all that apply.)

- ☐ American Association for the Advancement of Science (AAAS)
- ☐ American Association of Physics Teachers (AAPT)
- ☐ Association for Science Teachers Education (ASTE)
- ☐ Bechtel
- ☐ California Alliance of African American Educators (CAAAE)
- ☐ California Association of Bilingual Educators (CABE)
- ☐ California Association of Resource Specialists (CARS+
- ☐ California Educational Research Association (CERA)
- ☐ California English Language Development Test (CELDT)
- ☐ California Parent Teacher Conference (PTA)
- ☐ California Science Teacher Association (CSTA)
- ☐ California Special Education Local Plan Areas (SELPA)
- ☐ California Department of Education (CDE) Technical Advisory Group (TAG)
- ☐ CDE Curriculum and Instruction Steering Committee (CISC)
- ☐ CDE Science Expert Panel (SEP)
- ☐ Chevron
- ☐ National Association for Research in Science Teaching (NARST)
- ☐ National Association of Biology Teachers (NABT)
- ☐ National Earth Science Teachers Association (NESTA)
- ☐ Project Lead the Way (PLTW)
- ☐ Regional Assessment Network (RAN)
- ☐ None of the Above
- ☐ Other:

Personal Education

Please list any undergraduate and postgraduate degrees obtained, most recent first.

Name of Institution **Degree Obtained** **Year Completed**

[Options for the “Degree Obtained” dropdown box are AA, BA, MA, EdD, and PhD]

Major

Name of Institution **Degree Obtained** **Year Completed**

[Options for the “Degree Obtained” dropdown box are AA, BA, MA, EdD, and PhD]

Major

Name of Institution **Degree Obtained** **Year Completed**

[Options for the “Degree Obtained” dropdown box are AA, BA, MA, EdD, and PhD]

Major

Name of Institution **Degree Obtained** **Year Completed**

[Options for the “Degree Obtained” dropdown box are AA, BA, MA, EdD, and PhD]

Major

Teaching Credentials

[Only those who have marked "Yes" option for "**Are you currently teaching or have you taught at a K–12 school in California?**" in "Personal Information" will see this page.]

NOTE: If you don't remember your teaching credential number, you may look it up at

---

Credential Type **Credential Number**

Subject(s)

[Dropdown with Elementary (Multiple Subject), Secondary (Single Subject) and Special Education as options]

Expiration Date

Month Year

Credential Type **Credential Number**

Subject(s)

[Dropdown with Elementary (Multiple Subject), Secondary (Single Subject) and Special Education as options]

Expiration Date

Month Year

Credential Type **Credential Number**

Subject(s)

[Dropdown with Elementary (Multiple Subject), Secondary (Single Subject) and Special Education as options]

Expiration Date

Month Year

Teaching Experience

[Only those who have marked "Yes" option for "**Are you currently teaching or have you taught at a K–12 school in California?**" or "Yes" for "**Are you currently teaching or have you taught at a college/university level?**" in "Personal Information" will see this page.]

List up to 3 experiences with the most recent first.

1) **Subject(s)**
[Drop down with Physical Science; Biological Science; Earth Science; Integrated Science; Chemistry; Physics; Biology; Mathematics; English–Language Arts; Earth, Planetary, or Environmental Science; Multiple Subjects (K–5); and Other as options]

- **If "Other" was selected, please specify subject(s) taught.**
- [Only appears if "Other" for previous dropdown was selected]
- **Total Number of Years Taught**

- **Population(s) Served**

☐ General Education
☐ Bilingual Education
☐ Special Education

- **Level**

☐ Kindergarten
☐ Grade 1
☐ Grade 2
☐ Grade 3
☐ Grade 4
☐ College/University level
☐ Grade 5
☐ Grade 6
☐ Grade 7
☐ Grade 8
☐ Grade 9
☐ Grade 10
☐ Grade 11
☐ Grade 12

- Check all that apply

2) **Subject(s)**
[Drop down with Physical Science; Biological Science; Earth Science; Integrated Science; Chemistry; Physics; Biology; Mathematics; English–Language Arts; Earth, Planetary, or Environmental Science; Multiple Subjects (K–5); and Other as options]

- **If "Other" was selected, please specify subject(s) taught.**
- [Only appears if "Other" for previous dropdown was selected]
- **Total Number of Years Taught**

- **Population(s) Served**

☐ General Education
☐ Bilingual Education
☐ Special Education

- **Level**

☐ Kindergarten
☐ Grade 1
☐ Grade 2
☐ Grade 3
☐ Grade 4
☐ College/University level
☐ Grade 5
☐ Grade 6
☐ Grade 7
☐ Grade 8
☐ Grade 9
☐ Grade 10
☐ Grade 11
☐ Grade 12

- Check all that apply

3) **Subject(s)**

[Drop down with Physical Science; Biological Science; Earth Science; Integrated Science; Chemistry; Physics; Biology; Mathematics; English–Language Arts; Earth, Planetary, or Environmental Science; Multiple Subjects (K–5); and Other as options]

- **If "Other" was selected, please specify subject(s) taught.**
- [Only appears if "Other" for previous dropdown was selected]

Total Number of Years Taught

- **Population(s) Served**

☐ General Education
☐ Bilingual Education
☐ Special Education

- **Level**

☐ Kindergarten
☐ Grade 1
☐ Grade 2
☐ Grade 3
☐ Grade 4
☐ College/University level
☐ Grade 5
☐ Grade 6
☐ Grade 7
☐ Grade 8
☐ Grade 9
☐ Grade 10
☐ Grade 11
☐ Grade 12

- Check all that apply

Have you ever taught students from disadvantaged and/or underrepresented backgrounds?

● Yes ● No

Years Taught

[If previous is marked 'yes', then will appear.]

Have you ever taught English learners?

● Yes ● No

Years Taught

[If previous is marked 'yes', then will appear.]

Have you ever taught students with severe cognitive disabilities?

● Yes ● No

Years Taught

[If previous is marked 'yes', then will appear.]

Employment

Current Position

Employer

Are you working for a school and/or local educational agency (LEA)?

● Yes ● No

School [Only appears if answer to previous is Yes]

Be sure to include the full name of your school. Please do not use initialisms.

LEA [Same]

Be sure to include the full name of your LEA. Please do not use initialisms.

Current LEA type [Same; drop down options: Urban, Suburban, and Rural]

Professional Experience

Please rate your familiarity with the Next Generation Science Standards.

- 1
- 2
- 3
- 4
- 5

1 = Not at all familiar; 5 = very familiar

Please provide any further information about your professional background that relates to the work of this meeting. (For example, coursework or training in science and/or assessments, programs implemented, etc.)

Please list any applicable local, state, and national professional organizations to which you belong that relate to the work of this meeting. (Please do not use initialisms.)

Demographic Data

Gender

● Male ● Female

Is Spanish your native language?

● Yes ● No

Ethnic Background

☐ Asian
☐ Black or African American
☐ Hispanic or Latino
☐ White
☐ Other:

Appendix C: Recommendations Outside the Scope of the Meetings

Many of the stakeholders participating in the meetings provided recommendations on how to handle various issues related to science assessment but were outside the purview of these meetings. These recommendations are as follows:

- Provide ready-to-use practice assessments to teachers.
- Focus on the Science and Engineering Practices domain of the NGSS at all grade levels.
- Build an item bank for use at the school level that teachers can access to assess where students are in the learning progression of a particular science topic.
- At the LEA/school level, mandate science journals for each grade level that students will be required to carry to the next grade level for use in content review.
- Provide boxes of lab material for use in a PT assessment to encourage a hands-on lab experience in all classrooms.
- Administer an early (fall) summative assessment to provide teachers with a benchmark of student progress.
- Shorten the length of the assessment to reduce loss of instruction time.
- Delay high-stakes assessments at least one year after the frameworks are developed and adopted to provide teachers with time to acquire professional development and implement new curriculum.
- Restructure assessments to feel more like a game to gain student buy-in; for example, include a visible score that can be seen during a game-type assessment.

Appendix D: General Session PowerPoint

2014 CAASPP
Science Stakeholders

Appendix E: Group Session PowerPoint and Handouts


2014 CAASPP Science Stakeholders


Handout1_CBT


Handout2_CR Items


Handout3_NGSS Reiser

Appendix F: Group Discussion Questions

Group Recommendations

1. **What will a California NGSS Assessment look like, measure, and require?** ***For each question, please provide a detailed rationale, citing both benefits and limitations of choice, based on evidence-based experience and best assessment practices.***

   1a. At which grade level, within each grade span (three through five, six through nine, ten through twelve), as referenced in 60640(b), should an NGSS assessment be administered?
   1b. What science content domains (Life Science, Physical Science, Earth and Space Science, Integrated Domains) should be targeted for assessment at each of the grade levels proposed in 1a?
   1c. At what grade levels, in addition to those proposed in 1a, should a science assessment be administered?
   1d. What science content domains (Chemistry, Physics, Biology, Earth Science, Environmental Science, Engineering, etc.) should be targeted for assessments at each of the grade levels proposed in 1c?
   1e. What type of assessments (computer-based, computer-adaptive, paper-pencil, etc.) should be implemented for the subjects proposed in 1b and 1d?
   1f. How will the challenges of developing questions to assess the 3 dimensions of NGSS (performance expectations) be addressed through the recommended assessment system?
   1g. NGSS storylines summarize major themes in NGSS science while emphasizing the practices and cross cutting concepts within and along the continuum of learning progressions. How should major NGSS storylines within grade bands inform assessment development?

2. **What assessment options should be considered for the California NGSS?** ***For each question, please provide a detailed rationale, citing both benefits and limitations of choice, based on evidence-based experience and best assessment practices.***

   2a. What item types (selected-response, technology-enhanced, constructed-response, task-centered, etc.) should be administered on each assessment?
   2b. If needed, what alternate California NGSS assessments should be implemented beyond the ESEA mandated grade spans (three to five, six through nine, ten through twelve)?
   2c. What sampling plan possibilities are recommended? What are the benefits and limitations of this plan?

Group Consensus:

Appendix G: NGSS Architecture

How to Read NGSS

A Look at NGSS

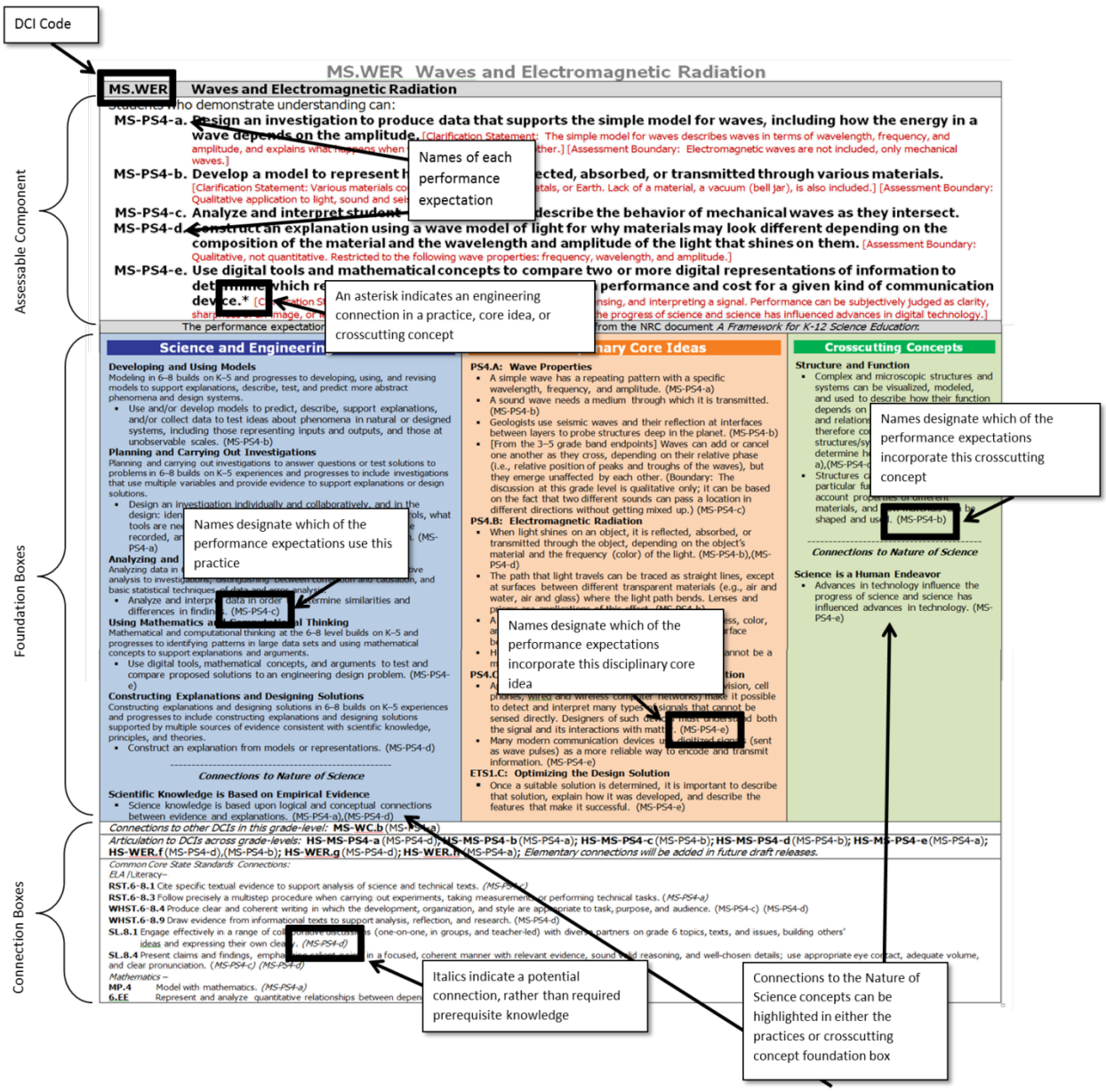

Figure G.1 Example of How to Analyze an NGSS Box

Appendix H: Acronyms, Initialisms, and Definitions

Acronyms and Initialisms

- **AAAS:** American Association for the Advancement of Science
- **AADE:** American Association of Diabetes Educators
- **AAPT:** American Association of Physics Teachers
- **AB:** Assembly Bill
- **ACS:** American Chemical Society
- **ACSA:** American School Counselor Association
- **AND:** Academy of Nutrition and Dietetics, formerly American Dietetic Association (ADA)
- **AP:** Advanced Placement
- **APHA:** American Public Health Association
- **ASCD:** Association for Supervision and Curriculum Development
- **ASTE:** Association for Science Teacher Education
- **CA:** California
- **CAAAE:** California Alliance of African American Educators
- **CAASPP:** California Assessment of Student Performance and Progress
- **CABE:** California Association of Bilingual Educators
- **CAG:** California Association for the Gifted
- **CAPA:** California Alternate Performance Assessment
- **CARS+:** California Association of Resource Specialists
- **CAT:** Computer-adaptive testing
- **CBT:** Computer-based testing
- **CCSSO:** Council of Chief State School Officers
- **CDE:** California Department of Education
- **CELDT:** California English Language Development Test
- **CERA:** California Educational Research Association
- **CISC:** Curriculum and Instruction Steering Committee
- **CMA:** California Modified Assessment
- **CSP:** California Science Project
- **CST:** California Standards Test
- **CSTA:** California Science Teacher Association
- **CUE:** Computer-Using Educators
- **DCI:** Disciplinary Core Idea
- ***EC:** Education Code*
- **EL:** English Learner
- **EOC:** End-of-Course

- **EOY:** End-of-Year
- **ES:** Elementary School
- **ESEA:** Elementary and Secondary Education Act
- **ETS:** Educational Testing Service
- **F:** Formative
- **GL:** Grade Level
- **GLOBE:** Global Legislators Organization for a Balanced Environment
- **HOT:** Hands-on Task
- **HS:** High School
- **I:** Interim
- **IB:** Item Bank
- **INT:** Integrated
- **K:** Kindergarten
- **LP:** Learning Progression
- **LS:** Life Science
- **MS:** Middle School
- **MSg:** Multistage
- **NABT:** National Association of Biology Teachers
- **NAGT:** National Association of Geoscience Teachers
- **NARST:** National Association for Research in Science Teaching
- **NBCT:** National Board Certified Teachers
- **NCEO:** National Center on Educational Outcomes
- **NCHEC:** National Commission for Health Education Credentialing
- **NESTA:** National Earth Science Teachers Association
- **NGA:** National Governors Association
- **NGSS:** Next Generation Science Standards
- **NMLSTA:** National Middle Level Science Teachers Association
- **NRC:** National Research Council
- **NSTA:** National Science Teachers Association
- **PARCC:** Partnership for Assessment of Readiness for College and Careers
- **PDF:** Portable Document Format
- **PE:** Performance Expectation
- **PL:** Performance Level
- **PLD:** Performance Level Descriptor
- **PLTW:** Project Lead the Way
- **P/P:** Paper/Pencil Test *or* Paper-pencil Test

- **PPT:** PowerPoint Presentation
- **PT:** Performance-based Task
- **PTA:** Parent Teacher Association
- **RAN:** Regional Assessment Network
- **S:** Summative
- **SBE:** State Board of Education
- **SCAS^2:** Southern California Association of Science Specialists
- **SDSA:** San Diego Science Alliance
- **SDSEA:** San Diego Science Educators Association
- **SELPA:** Special Education Local Plan Area
- **SEP:** Science Expert Panel
- **SIM:** Simulation
- **SLT:** Strategic Leadership Team
- **SPPI:** State Superintendent of Public Instruction
- **SRT:** State Review Team
- **STAR:** Standardized Testing and Reporting
- **TAG:** Technical Advisory Group
- **TE:** Technology Enhanced
- **TTSC:** Technology and Telecommunications Steering Committee

Definitions

Alternate Assessment: An assessment "used to evaluate the performance of students who are unable to participate in general state assessments even with accommodations; provides a mechanism for students with the most significant cognitive disabilities, and for other students with disabilities who may need alternate ways to access assessments, to be included in an educational accountability system." ("National Center on Educational Outcomes: *Alternates Assessments for Students with Disabilities*," 2013, para. 1)

Benchmark Assessment: Typically a short assessment that is often given several times during the school year to provide feedback on where students are in an LP; may be used to focus science DCIs on the educational needs of individual students; see also Interim Assessment.

Census Administration: An administration of items that cover an entire domain; given to all students within a tested grade level across a state; see Figure H.1 (CDE and ETS, 2014) for a brief overview of a census administration.

CENSUS ADMINISTRATION				
STUDENTS	ITEMS			
STUDENT 1	XXXXXX	XXXXXX	XXXXXX	XXXXXX
STUDENT 2	XXXXXX	XXXXXX	XXXXXX	XXXXXX
STUDENT 3	XXXXXX	XXXXXX	XXXXXX	XXXXXX
STUDENT 4	XXXXXX	XXXXXX	XXXXXX	XXXXXX

Figure H.1 Example of a Census Administration

Computer-adaptive Test (CAT): A type of CBT where the content being measured and the measurement process are altered as the student interacts with the computer in order to configure the assessment to the student by using answers to earlier questions to determine which questions are asked next, causing the assessment to change over time as the student's performance level is assessed.

Computer-based Test (CBT): An assessment delivered via the platform of a computer or tablet.

Consortium-developed Assessment: An assessment developed through a group partnership, such as a group of states or educators (e.g., Partnership for Assessment of Readiness for College and Careers [PARCC] and the Smarter Balanced Assessment Consortium [SBAC]) that formed to accomplish a common goal.

Content Framework: A foundation for the NGSS that is based on evidence by incorporating current scientific research and effective methods for how students learn science; may be modified to meet the particular needs of a state's student population; identifies the science content that K–12 students should know.

Crosscutting Concept: A concept that links different science domains and is applicable across all science domains (e.g., patterns, similarity, and diversity; cause and effect; scale, proportion, and quantity; systems and system models; energy and matter: flows, cycles, and conservation; structure and function; stability and change) by providing "an organizational schema for interrelating knowledge from various science fields into a coherent and scientifically-based view of the world." ("Next Generation Science Standards: *Three Dimensions*," 2014, para. 4)

Dimension: An aspect of the NGSS (e.g., Practices, Crosscutting Concepts, and disciplinary core ideas [DCIs]).

Disciplinary Core Idea (DCI): An idea that provides a focus for aspects of science in K–12 curriculum, instruction, and assessments; is an important, broad theme across multiple domains or is an organizational concept for a single domain; provides a tool for understanding or investigating complex ideas and solving problems; relates to the interests held by or life experiences of students, or connects to society or personal concerns requiring scientific or technological knowledge; is teachable and learnable over multiple grades while having an LP.

Discrete Item: Any item that is not part of a group of items tied to a text passage or graphic; has content that is independently answerable from all other items on the assessment; see also Standalone Item.

Domain: A group of disciplinary ideas (e.g., the physical sciences; the life sciences; the earth and space sciences; and engineering, technology, and applications of science).

End-of-Course (EOC) Assessment: An assessment for courses that are content-specific and cover explicit content objectives, such as Biology, Chemistry, and Physics, and usually given at the middle and high school levels.

End-of-Year (EOY) Assessment: An assessment for courses that have grade-specific content yet may be within a single domain, such as Physical Science, or include multiple domains of science, and is usually given at the elementary and middle school levels; however, an EOY assessment may also be given at the high school level through classes that are integrated.

Formative Assessment (F): An assessment developed for learning, administered during instructional units to improve instruction and identify students' strengths/weaknesses in order to evaluate where students are at in a learning progression.

Hands-on Task (HOT): An activity that requires students to use equipment and materials in a laboratory setting to conduct a science experiment in order for the students to demonstrate investigative, problem solving, and reasoning skills by applying scientific knowledge in a complex, real-world context.

High-stakes Assessment: Any assessment that is used to make decisions about the following: students, parents/guardians, educators, administrators, schools, LEAs, states, and/or nations for the purposes of accountability (i.e., to help determine the effectiveness of an education program in preparing students for college or careers); may be used to either reward or take disciplinary action against a person or entity; often administered at a statewide or national level.

Item Bank: A collection of items to be, being, or have been used on an assessment that can be accessed by the assessment developer and owner; may include practice items that are accessible to students, parents/guardians, and educators.

Interim Assessment (I): Typically a short assessment that is often given several times during the school year to provide feedback on where students are in an LP; may be used to focus science DCIs on the educational needs of individual students; see also Benchmark Assessment.

Learning Progression (LP): An "empirically grounded and testable hypothesis about how students' understanding of, and ability to use, core scientific concepts, explanations, and related scientific practices grow and become more sophisticated over time, with appropriate instruction." (Corcoran, Mosher, & Rogat, 2009, p. 20)

Locally Scored Assessment: Any assessment that is developed and scored at the classroom, school, or LEA level, rather than at the statewide or national level.

Manipulative: Any tangible object, tool, model, or mechanism that can be used by a student to demonstrate PL or location within an LP while completing a PT focused on science or engineering DCIs.

Matrix Sampling: An administration of a sample of items that cover a subset of a domain; different students may receive different items within a tested grade level across a state; see Figure H.2 (CDE & ETS, 2014) for a brief overview of matrix sampling.

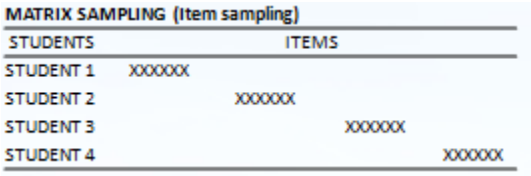

Figure H.2 Example of Matrix Sampling

Multi-stage (MSg) Computer Adaptive Test (CAT): A type of assessment with multiple stages where stage difficulty level (e.g., Easy, Medium, or Hard) is determined via CAT; a routing test (first stage) is given to a student and upon student completion the student's score determines which difficulty level of a second stage should be administered next to the student; well-performing students are assigned a second stage composed of items more difficult than those in the first stage, while struggling students are assigned a second stage composed of items easier than those in the first stage; after student completion of the second stage the assessment can either end with a final score compiled from performance across both the routing and second stages or more stages can be administered; see Figure H.3 (adapted from Davey, 2011) for a brief overview of a two-stage CAT.

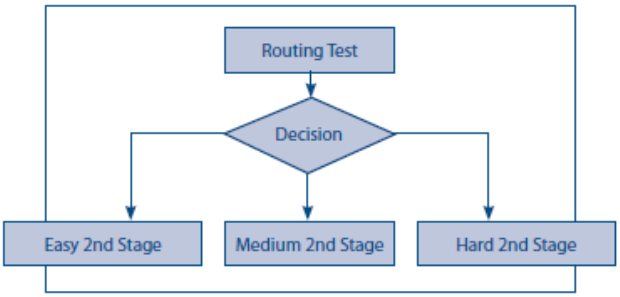

Figure H.3 Example of a Two-stage CAT

Non-ESEA Assessment: Any assessment that is outside the legal scope of the federal requirements of ESEA.

Partial Matrix Sampling: An administration of a set of items that all students are assessed with in common and a sample of items that cover a subset of a domain; different students may receive a different sample of items within a tested grade level across a state; see Figure H.4 (CDE & ETS, 2014) for a brief overview of partial matrix sampling.

PARTIAL MATRIX SAMPLING					
STUDENTS	ITEMS				
	Common				
STUDENT 1	XXX	XXX			
STUDENT 2	XXX		XXX		
STUDENT 3	XXX			XXX	
STUDENT 4	XXX				XXX

Figure H.4 Example of Partial Matrix Sampling

Performance Level (PL): An indicator of a student's level of proficiency in science content and practices (e.g., basic, proficient, and advanced).

Performance Level Descriptor (PLD): A description that identifies what students should know and be able to accomplish for each level of proficiency.

Performance-based Task (PT): A task that provides an opportunity for a student to demonstrate PL in the three dimensions of the NGSS, with evidence of PL based on observations of the student who is engaged in scientific or engineering practices related to DCIs; requires the student to construct an answer, produce a product, or perform an activity; often carried out in a classroom setting due to difficulty of monitoring this type of assessment at a large scale (e.g., statewide or national).

Population Sampling: An administration of items that cover an entire domain to a representative sample of students across a state; see Figure H.5 (CDE & ETS, 2014) for a brief overview of population sampling.

POPULATION SAMPLING (Student sampling)				
STUDENTS	ITEMS			
STUDENT 1	XXXXXX	XXXXXX	XXXXXX	XXXXX
STUDENT 2				
STUDENT 3	XXXXXX	XXXXXX	XXXXXX	XXXXX
STUDENT 4				

Figure H.5 Example of Population Sampling

Portfolio: A collection of a student's work gathered over the course of a unit or school year, which may include both artifacts of instruction (e.g., teaching materials) along with the student's assessment results.

Qualitative Data: Any descriptive data that comes from conceptual observations and narratives, such as interviews and subjective opinions or feelings.

Quantitative Data: Any numerical data that results from systematic measurements, such as the metric length of an object; is often more easily analyzed mathematically or statistically.

Reporting Plan: A process by which students' assessment scores will be distributed to the following: students, parents/guardians, educators, administrators, schools, LEAs, states, and/or nations.

Science Practice: A set of behaviors used by scientists while investigating the natural world or by engineers while designing then building models and systems.

Simulation Task (SIM): An activity that is unable to be easily recreated in a classroom setting so is delivered via a computer or tablet platform; may allow students to manipulate real-world data in a virtual environment.

Standalone Item: Any item that is not part of a group of items tied to a text passage or graphic; has content that is independently answerable from all other items on the assessment; see also Discrete Item.

Storyline: An overview of a major idea within a grade level's standards, includes emphasis on the practices and crosscutting concepts within and along the continuum of learning progressions.

Summative Assessment (S): An assessment of learning, administered at the end of instructional units (or at the conclusion of some defined period of instruction) in order to provide evidence of mastery of a particular content and aid in decision-making (e.g., assigning grades, promotion/retention, student classification by performance level).

Test Blueprint: A guide, usually in chart format, to the number of each DCI or PE that should be assessed in a given assessment year; helps determine the number of items needed in an item bank.

Virtual Environment: A computer-generated, often three-dimensional, representation of a scientific setting, such as a task requiring a student to redesign an electric car, a SIM allowing a student to conduct an acid-base reaction in a chemistry laboratory, or an item asking a student to measure the movement of an object over time within the solar system, in which a student perceives herself or himself to be in control of and can interact with the variables found in the setting.

Appendix I: Transcript of Online Survey

https://www.formstack.com/forms/?1770382-VDr42bxzl3

2014 CAASPP Science Stakeholders Online Survey

As stipulated in *Education Code* (*EC*) Section 60640, the California Department of Education, in collaboration with Educational Testing Service, is gathering input from stakeholders regarding science assessments aligned to the newly adopted Next Generation Science Standards (NGSS). The input from stakeholders will be shared with State Superintendent of Public Instruction Tom Torlakson as he prepares recommendations for the State Board of Education for the new K–12 science assessments.

To provide your input, please complete the following four-part online survey. **Part one** focuses on assessments pertaining to federal Elementary and Secondary Education Act (ESEA) requirements. **Part two** focuses on assessments pertaining to non-ESEA requirements. **Part three** focuses on measurement considerations for testing. **Part four** elicits feedback on the science assessment system as a whole.

In preparing your responses to the survey questions, please view the "Overview of NGSS and Assessments" Webcast prior to filling out the survey at http://californiatac.org/training/webcast/ngss.html

Did you participate in the CAASPP Science Stakeholders Meeting held in July 2014?

- ○ Yes
- ○ No

Part 1: ESEA-mandated CAASPP Assessments

Pursuant to *EC* 60640(b)

Please refer to the following definitions for the questions in this section.

Item Types

- **Selected-response/multiple-choice item:** A type of item that requires pupils to select one or more responses from a set of options.
- **Technology-enhanced item:** A type of item that uses technology to collect evidence through a non-traditional response type.
- **Constructed-response item:** A type of item that prompts students to produce a text or numerical response in order to collect evidence about their knowledge or understanding of a given core idea.
- **Task-centered item:** A type of item that assesses a set of core ideas as opposed to a narrow focus on just one or two ideas, as is typically the case with selected-response and constructed-response items. *Note:* Sub-items can be of different item types; i.e., selected-response, constructed-response, or technology-enhanced.

Sample item types posted by the Smarter Balanced Assessment Consortium can be viewed in the following video. Please note these are general item types and this clip does not contain NGSS items specifically: Video

Assessment Types

- **Computer-based assessments:** A test administered using an electronic computing device.
- **Computer adaptive assessments:** A computer-based test that uses a computer program to adjust the difficulty of test items throughout a testing session based on a test taker's responses to previous test items during that testing session.
- **Paper-pencil assessments:** A test administered using paper-based materials.

1) At which grade level, within each ESEA-mandated grade span, should a California NGSS assessment be administered for ESEA purposes? *Please select one grade per grade span and provide a rationale supported by evidence-based experience and assessment best practices.*

1a) Grades 3 through 5:

- ○ **Grade 3**
- ○ **Grade 4**
- ○ **Grade 5**

➔ ***These questions will appear only if "Grade 3" is selected above.***

Please provide your rationale for choosing Grade 3:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 3?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 3?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other:

Please select all that apply.

Which item type(s) should be administered in grade 3?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other:

Please select all that apply.

➔ ***These questions will appear only if "Grade 4" is selected above.***

Please provide your rationale for choosing Grade 4:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 4?
- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 4?
- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other: [ ]

Please select all that apply.

Which item type(s) should be administered in grade 4?
- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other: [ ]

Please select all that apply.

➔ ***These questions will appear only if "Grade 5" is selected above.***

Please provide your rationale for choosing Grade 5:

[ ]

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 5?
- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 5?
- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other: [ ]

Please select all that apply.

➔ ***This question will appear if any of the grades for grades 3 to 5 is selected above.***

Please provide a rationale for your selection of content domain, type of assessment, and item type for the grade you selected:

1,000 characters maximum

1b) Grades 6 through 9:

- ○ Grade 6
- ○ Grade 7
- ○ Grade 8
- ○ Grade 9

➔ ***These questions will appear only if "Grade 6" is selected above.***

Please provide your rationale for choosing Grade 6:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 6?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 6?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other:

Please select all that apply.

Which item type(s) should be administered in grade 6?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other:

Please select all that apply.

➔ ***These questions will appear only if "Grade 7" is selected above.***

Please provide your rationale for choosing Grade 7:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 7?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 7?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other:

Please select all that apply.

Which item type(s) should be administered in grade 7?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other:

Please select all that apply.

➔ ***These questions will appear only if "Grade 8" is selected above.***

Please provide your rationale for choosing Grade 8:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 8?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 8?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other:

Please select all that apply.

Which item type(s) should be administered in grade 8?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other:

Please select all that apply.

➔ ***These questions will appear only if "Grade 9" is selected above.***

Please provide your rationale for choosing Grade 9:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 9?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 9?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other:

Please select all that apply

Which item type(s) should be administered in grade 9?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other:

Please select all that apply.

➔ ***This question will appear if any of the grades for grades 6 to 9 is selected above.***

Please provide a rationale for your selection of content domain, type of assessment, and item type for the grade you selected:

1,000 characters maximum

1c) Grades 10 through12:

- ○ Grade 10
- ○ Grade 11
- ○ Grade 12

➔ ***These questions will only appear if "Grade 10" is selected above.***

Please provide your rationale for choosing Grade 10:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 10?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 10?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other: [ ]

Please select all that apply.

Which item type(s) should be administered in grade 10?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items
- ☐ Task-centered items
- ☐ Other: [ ]

Please select all that apply.

➔ ***These questions will appear only if "Grade 11" is selected above.***

Please provide your rationale for choosing Grade 11:

[ ]

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 11?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 11?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Other: [ ]

Please select all that apply.

Which item type(s) should be administered in grade 11?

- ☐ Selected-response/multiple-choice items
- ☐ Technology-enhanced items
- ☐ Constructed-response items

☐ Task-centered items
☐ Other:
Please select all that apply

➔ ***These questions will appear only if "Grade 12" is selected above.***

Please provide your rationale for choosing Grade 12:

1,000 characters maximum

Which science content domain(s) should be targeted for assessment in grade 12?
☐ Biological Science/Life Science
☐ Earth and Space Science
☐ Physical Science
☐ Integrated Science
Please select all that apply

Which type(s) of assessment should be available for administration in grade 12?
☐ Computer-based
☐ Computer adaptive
☐ Paper-pencil
☐ Other:
Please select all that apply

Which item type(s) should be administered in a grade 12?
☐ Selected-response/multiple-choice items
☐ Technology-enhanced items
☐ Constructed-response items
☐ Task-centered items
☐ Other:
Please select all that apply.

➔ ***This question will appear if any of the grades for grades 9 to 12 is selected above.***

Please provide a rationale for your selection of content domain, type of assessment, and item type for the grade you selected:

Part 2: Additional CAASPP Assessments

Pursuant to *EC* 60640(c)

Please refer to the following definitions for the questions in this section.

Ways content can be assessed

- **Integrated:** Assessing content from multiple grades or disciplines.
- **End-of-year (EOY):** Assessing content from a specific grade.
- **End-of-course (EOC):** Assessing content from a non-grade specific course or discipline. *Note:* EOC is offered only as an option for middle and high school grade levels (grades 6 through 12) as courses in elementary school typically span the full academic year.

Assessment types

- **Summative:** Summative assessments are assessments of learning. They usually are administered at the end of instructional units and assess mastery of all instructed content. They usually are used for providing evidence of mastery of a particular content or to aid in decision making (such as assigning grades, promotion/retention, student classification by performance level).
- **Formative:** Formative assessments are assessments for learning. They usually are administered during instructional units for providing immediate feedback to improve instruction and identify individual student strengths and weaknesses.
- **Interim:** Interim assessments are assessments of learning, as are summative assessments, but instead of being administered at the very end of instruction, they are administered at specified points in instruction to assess material covered within those periods. They sometimes are referred to as benchmark assessments, as they can be used to assess student mastery of specific content standards or benchmarks immediately after instruction of those standards.
- **Computer-based assessments:** Tests administered using an electronic computing device.
- **Computer adaptive assessments:** Computer-based tests that use a computer program to adjust the difficulty of test items throughout a testing session based on a test taker's responses to previous test items during that testing session.
- **Paper-pencil assessments:** Tests administered using paper-based materials.

2) At which grade level(s), in addition to those you indicated previously in this survey, should a science assessment be administered? *Please select all that apply and provide a detailed rationale supported by evidence-based experience and assessment best practices.*

- ☐ **Grade 3**
- ☐ **Grade 4**
- ☐ **Grade 5**
- ☐ **Grade 6**
- ☐ **Grade 7**
- ☐ **Grade 8**
- ☐ **Grade 9**
- ☐ **Grade 10**
- ☐ **Grade 11**
- ☐ **Grade 12**

➔ ***These questions will appear only if "Grade 3" is selected above.***

How should content be assessed in grade 3?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 3?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 3?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: ☐

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 3:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 4" is selected above.***

How should content be assessed in grade 4?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 4?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science

☐ Integrated Science
Please select all that apply.

Which type(s) of assessment should be available for administration in grade 4?
☐ Computer-based
☐ Computer adaptive
☐ Paper-pencil
☐ Summative
☐ Formative
☐ Interim
☐ Other:
Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 4:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 5" is selected above.***
How should content be assessed in grade 5?
☐ Integrated
☐ End-of-year (EOY)
☐ Other:
Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 5?
☐ Biological Science/Life Science
☐ Earth and Space Science
☐ Physical Science
☐ Integrated Science
Please select all that apply.

Which type(s) of assessment should be available for administration in grade 5?
☐ Computer-based
☐ Computer adaptive
☐ Paper-pencil
☐ Summative
☐ Formative
☐ Interim
☐ Other:

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 5:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 6" is selected above.***

How should content be assessed in grade 6?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ End-of-course (EOC)
- ☐ Other:

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 6?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 6?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other:

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 6:

1,000 characters maximum

➔ ***These questions will only appear if "Grade 7" is selected above.***

How should content be assessed in grade 7?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ End-of-course (EOC)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 7?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 7?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: ☐

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 7:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 8" is selected above.***

How should content be assessed in grade 8?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ End-of-course (EOC)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 8?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 8?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: [ ]

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 8:

[ ]

1,000 characters maximum

➔ ***These questions will appear only if "Grade 9" is selected above.***

How should content be assessed in grade 9?

- ☐ Integrated
- ☐ End-of-course (EOC)
- ☐ Other: [ ]

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 9?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 9?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: [ ]

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 9:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 10" is selected above.***

How should content be assessed in grade 10?

- ☐ Integrated
- ☐ End-of-course (EOC)
- ☐ Other:

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 10?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 10?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other:

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 10:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 11" is selected above.***

How should content be assessed in grade 11?

- ☐ Integrated
- ☐ End-of-course (EOC)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 11?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply.

Which type(s) of assessment should be available for administration in grade 11?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: ☐

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 11:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 12" is selected above.***

How should content be assessed in grade 12?

- ☐ Integrated
- ☐ End-of-course (EOC)
- ☐ Other: ☐

Please select all that apply.

Which science content domain(s) should be targeted for assessment in grade 12?

- ☐ Biological Science/Life Science
- ☐ Earth and Space Science
- ☐ Physical Science
- ☐ Integrated Science

Please select all that apply

Which type(s) of assessment should be available for administration in grade 12?

- ☐ Computer-based
- ☐ Computer adaptive
- ☐ Paper-pencil
- ☐ Summative
- ☐ Formative
- ☐ Interim
- ☐ Other: [ ]

Please select all that apply.

Please provide a rationale for your selections above of how to assess content, what content domain, and which type(s) of assessment for grade 12:

[ ]

1,000 characters maximum

3) Federal legislation mandates science assessments for students with severe cognitive disabilities who are currently tested with the California Alternate Performance Assessment (CAPA). Other than your recommendations for the ESEA-mandated *tests* listed in question 1, at which grade level should additional test(s) be administered to this student group? *Please select all that apply and provide a detailed rationale supported by evidence-based experience and assessment best practices.*

- ☐ **Grade 3**
- ☐ **Grade 4**
- ☐ **Grade 5**
- ☐ **Grade 6**
- ☐ **Grade 7**
- ☐ **Grade 8**
- ☐ **Grade 9**
- ☐ **Grade 10**
- ☐ **Grade 11**
- ☐ **Grade 12**

➔ ***These questions will appear only if "Grade 3" is selected above.***

How should content be assessed in grade 3 for this student group?

- ☐ Integrated
- ☐ End-of-year (EOY)
- ☐ Other: [ ]

Please select all that apply.

Please provide a rationale for your selection of how this content should be assessed in grade 3 for this student group:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 7" is selected above.***

How should content be assessed in grade 7 for this student group?

☐ Integrated
☐ End-of-year (EOY)
☐ Other:

Please select all that apply.

Please provide a rationale for your selection of how this content should be assessed in grade 7 for this student group:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 8" is selected above.***

How should content be assessed in grade 8 for this student group?

☐ Integrated
☐ End-of-year (EOY)
☐ Other:

Please select all that apply.

Please provide a rationale for your selection of how this content should be assessed in grade 8 for this student group:

1,000 characters maximum

➔ ***These questions will appear only if "Grade 9" is selected above.***

How should content be assessed in grade 9 for this student group?

☐ Integrated
☐ End-of-year (EOY)
☐ Other:

Please select all that apply.

Please provide a rationale for your selection of how this content should be assessed in grade 9 for this student group:

Part 3: Measurement Considerations

Please refer to the following definition for **Question 4**:

Matrix sampling involves assigning students *different subsets* of items that represent portions of the tested standards. For this type of test administration, no individual student receives items covering all standards, but all standards are assessed over all the students, such as class/school/district/state.

4) Should matrix sampling be used for the California NGSS assessments?

- ○ Yes
- ○ No
- ○ Not sure

Please provide a rationale for why or why not matrix sampling should be used:

1,000 characters maximum

Please refer to the following definition for **Question 5:**

Population sampling involves selecting a representative sample (by race/ethnicity, gender, urban/rural, etc.) of students within a grade level to take the assessments each year.

5) **Should** population sampling be used in administering the California NGSS assessments?

- ○ Yes
- ○ No
- ○ Not sure

Please provide a rationale for why or why not population sampling should be used:

1,000 characters maximum

Please refer to the following definitions for **Question 6:**

Automated scoring: Scoring that uses complex scoring rules or artificial intelligence algorithms implemented in a computer program to assign scores to constructed-response items.

Local scoring: The scoring of constructed-response items by local teachers/test administrators for students in their geographic area.

Centralized scoring: The scoring of constructed-response items in a central location by a group of raters who receiving training, certification, and score monitoring at a specific site during a specific time period.

Remote scoring: The scoring process that allows trained raters to score the assessment and view rubrics on how to score the items even when the scorers are not centrally located. It also delivers those scores back to the Test Delivery and Data Warehouse components to be stored with the student responses.

6) For open-ended items, such as constructed-response and performance tasks, which scoring method should be used? ***Please provide a detailed rationale supported by evidence-based experience and assessment best practices.***

- ○ Automated scoring
- ○ Local scoring
- ○ Centralized scoring
- ○ Remote scoring
- ○ Other: [ ]

Please provide a rationale for your selection of scoring methods for open-ended items:

[ ]

1,000 characters maximum

Part 4: Overall Feedback about the Future Science Assessment System

This part of the survey asks for your feedback regarding the design of a science assessment system for California, including your vision of how to integrate local, statewide, and national (e.g., National Assessment of Educational Progress [NAEP]) assessments in order to provide information about student performance in science.

7) What are your most important considerations in the design of California science assessments? ***Please select all that apply.***

- ☐ Including items that closely represent real-life science scenarios and thinking processes
- ☐ Assessing each tested student on the entire range of California NGSS for grade (grade-span)
- ☐ Maximizing the number of grade levels that are assessed
- ☐ Reducing testing time for students

Other considerations:

1,000 characters maximum

8) Please provide any other considerations for assessment of science. ***Please provide a detailed rationale supported by evidence-based experience and assessment best practices.***

1,000 characters maximum

Demographic Data (optional)

Gender

- ○ **Male**
- ○ **Female**

Primary role as a stakeholder

- ○ **K–12 administrator**
- ○ **K–5 teacher**
- ○ **Middle school (grades 6–8) teacher**
- ○ **High school (grades 9–12) teacher**
- ○ **Expert in teaching English learners**
- ○ **Expert in teaching students with disabilities**
- ○ **Higher education expert**
- ○ **Measurement expert**
- ○ **Scientist, researcher, and/or engineer**
- ○ **Parent**
- ○ **Other:** _____

Ethnic background

- ☐ **Asian**
- ☐ **Black or African American**
- ☐ **Hispanic or Latino**
- ☐ **White**
- ☐ **Other:** _____

➔ This section will appear only if the respondent selects **"Yes"** to "**Did you participate in the 2014 CAASPP Science Stakeholders Meeting?"**

Science Stakeholders Meeting Evaluation

1) Please rate the following on a scale of 1–5:

	5 – Far above average	4 – Above average	3 - Average	2 - Below average	1 - Far below average
Meeting overall	O	O	O	O	O
Facilitator's style	O	O	O	O	O
Materials	O	O	O	O	O
Slides	O	O	O	O	O
Meeting location	O	O	O	O	O

2) Please provide your feedback on the meetings:

	Strongly agree	Agree	Disagree	Strongly disagree
The opening session was clearly presented and helpful for the rest of the meeting.	O	O	O	O
I understood the purpose of the meeting.	O	O	O	O
The meeting was well organized.	O	O	O	O
Sufficient time was devoted to the tasks.	O	O	O	O

3) Additional comments:

1,000 character maximum

Appendix J: Summary of Responses to Part 1 (ESEA Assessments) of Online Survey for All Grade Levels

The summary of responses to Part 1 of the online survey in Section 7 focused on the selections for assessment content, assessment mode, and item types for the *most* selected grade levels within each ESEA-mandated grade span. These were grade five in the grades three to five span, grade eight in the grades six to nine span, and grade eleven in the grades ten to twelve span. This appendix summarizes survey responses for all grade levels, not just the majority-selected grade levels.

Table J.1 Summary of Which Science Content Domain(s) Should Be Targeted for Assessment in the Selected ESEA Grade Test

ESEA Grade Span*	Grade	Biological Science/ Life Science		Earth and Space Science		Physical Science		Integrated Science		Total Respondents
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	
3–5	3	16	41%	19	49%	15	38%	30	77%	**39**
	4	18	33%	19	35%	17	31%	37	67%	**55**
	5	100	36%	101	36%	96	35%	181	65%	**277**
6–9	6	12	40%	17	57%	10	33%	19	63%	**30**
	7	23	61%	15	39%	13	34%	21	55%	**38**
	8	84	32%	75	28%	123	47%	174	66%	**264**
	9	16	43%	18	49%	13	35%	17	46%	**37**
10–12	10	54	64%	21	25%	24	29%	33	39%	**84**
	11	104	60%	68	39%	89	51%	105	61%	**173**
	12	35	47%	34	45%	36	48%	58	77%	**75**

* **Note:** Survey respondents were asked to select ONE grade level from each grade span and then to select which science content domains they think should be tested in their selected grade level. Respondents were allowed to select all options that applied so the sum of the counts in each row does not equal the total number of respondents.

Table J.2 Summary of Which Type(s) of Assessments Should Be Available for Administration in the Selected ESEA Grade Test

ESEA Grade Span*	Grade	Computer-based		Computer adaptive		Paper-pencil		Other		Total Respondents
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	
3–5	3	23	59%	27	69%	14	36%	4	10%	**39**
	4	25	45%	33	59%	19	34%	9	16%	**56**
	5	151	54%	162	58%	102	37%	15	5%	**278**
6–9	6	17	59%	23	79%	9	31%	2	7%	**29**
	7	19	50%	25	66%	7	18%	1	3%	**38**
	8	152	58%	187	71%	78	30%	27	10%	**264**
	9	23	62%	26	70%	15	41%	3	8%	**37**
10–12	10	54	63%	62	72%	27	31%	6	7%	**86**
	11	100	58%	131	77%	48	28%	11	6%	**171**
	12	51	68%	56	75%	25	33%	10	13%	**75**

* **Note:** Survey respondents were asked to select ONE grade level from each grade span for the ESEA-mandated test and then to select which type(s) of assessments should be available for their selected grade level. Respondents were allowed to select all options that applied so the sum of the counts in each row does not equal the total number of respondents. The percent is the count divided by the total number of respondents in that row.

Table J.3 Summary of Which Item Type(s) Should Be Administered in the Selected ESEA Grade Test

ESEA Grade Span*	Grade	Selected-response/ multiple-choice items		Technology-enhanced items		Constructed-response items		Task-centered items		Other		Total Respondents
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	
3–5	3	20	51%	23	59%	20	51%	29	74%	2	5%	**39**
	4	23	40%	25	44%	31	54%	38	67%	1	2%	**57**
	5	198	71%	151	55%	168	61%	190	69%	7	3%	**277**
6–9	6	17	57%	19	63%	19	63%	23	77%	3	10%	**30**
	7	18	47%	23	61%	25	66%	27	71%	1	3%	**38**
	8	176	66%	179	68%	197	74%	209	79%	13	5%	**265**
	9	27	73%	22	59%	27	73%	26	70%	1	3%	**37**
10–12	10	53	61%	55	63%	64	74%	68	78%	6	7%	**87**
	11	107	62%	125	73%	144	84%	147	85%	4	2%	**172**
	12	57	74%	58	75%	65	84%	67	87%	4	5%	**77**

* **Note:** Survey respondents were asked to select ONE grade level from each grade span for the ESEA-mandated test and then to select which item type(s) should be administered for their selected grade level. Respondents were allowed to select all options that applied so the sum of the counts in each row does not equal the total number of respondents. The percent is the count divided by the total number of respondents in that row.

Appendix K: Summary of Science Stakeholder Meeting Evaluations

Survey respondents who attended one of the Science Stakeholder Meetings were also presented with several additional questions asking them to evaluate their experience. Seventy-four of the 422 respondents attended one of the meetings and were presented with two sets of meeting evaluation selected-response questions and an opportunity to contribute additional comments. Of these 74 respondents, 69 provided a response to one of the evaluation questions. Summaries of the selected-response questions and the one open-ended question are presented here.

Selected-response Feedback Questions

For the first set of meeting evaluation questions, respondents were asked for ratings from 1 (far below average) to 5 (far above average) on five aspects of the meetings, including the meeting overall, facilitator's style (for their group discussions), materials, presentation slides, and meeting location. Table K.1 summarizes the respondents' ratings on each of these meeting aspects. The ratings were generally on the mid to high end of the scale with the average ratings over the 69 respondents ranging from 3.59 to 3.99.

Table K.1 Summary of Ratings for Aspects of the Science Stakeholder Meetings

	5 – Far above average		4 – Above average		3 - Average		2 - Below average		1 - Far below average		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Respondents	Average
Meeting overall	18	26%	35	51%	13	19%	3	4%	0	0%	69	3.99
Facilitator's style	11	16%	35	51%	18	26%	4	6%	1	1%	69	3.74
Materials	12	17%	28	41%	24	35%	5	7%	0	0%	69	3.68
Slides	7	10%	29	42%	31	45%	2	3%	0	0%	69	3.59
Meeting location	16	23%	24	35%	28	41%	1	1%	0	0%	69	3.80

The second set of Science Stakeholder Meeting evaluation questions involved presenting the respondents with a feedback statement and asking them to select the extent to which they agreed with the statement from "Strongly Disagree" to "Strongly Agree." These statements are given in Table K.2 with the corresponding counts of respondents who selected each statement of agreement. For all statements, respondents mostly selected "Agree" or "Strongly Agree," indicating that these respondents generally found that the opening session was helpful, understood the purpose of the meeting, felt the meeting was organized, and had sufficient time for the tasks.

Table K.2 Summary of Feedback Evaluations of the Science Stakeholder Meetings

	Strongly agree		Agree		Disagree		Strongly disagree		Total
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Respondents
The opening session was clearly presented and helpful for the rest of the meeting.	22	33%	32	48%	13	19%	0	0%	**67**
I understood the purpose of the meeting.	38	56%	27	40%	1	1%	2	3%	**68**
The meeting was well organized.	34	50%	30	44%	4	6%	0	0%	**68**
Sufficient time was devoted to the tasks.	22	32%	29	43%	15	22%	2	3%	**68**

Open-ended Additional Comments Question

The last Science Stakeholder Meeting evaluation survey item provided respondents the opportunity to write in any additional comments they had on their experience at the meetings. Of the 74 respondents who attended one of the meetings, 43 provided additional comments. These comments were reviewed, and the following four common themes appeared in their responses:

- Expressed appreciation/thankfulness for meeting (*n*=19),
- Expressed concerns with length of the opening session or repetitive information presented from the Webcast (*n*=13),
- Expressed some concern with their facilitators (*n*=7), and
- Expressed a desire for more discussion time in their groups (*n*=7).

Overall, this small set of meeting attendees were thankful for the opportunity to be part of these meetings and voiced interest in having even more time to discuss the various aspects of the future CAASPP science assessment system with their groups.

Appendix L: Codes for Describing Online Survey Responses

This appendix provides all the codes used for classifying open-ended rationale responses in the Online Survey. Codes with at least two responses are listed. The most common codes with their *n* counts are provided in Section 7.

Survey, Part 1: ESEA-mandated CAASPP Assessments

Rationale	Codes
For Grade Level Selection	
Grade 3	• Early test will force science to be taught. • A grade three test provides a baseline. • Grade three students have skills to take the test.
Grade 4	• A grade four test will hold elementary school teachers accountable. • Grade four test results inform the next year of instruction. • Grade four students have the skills to take the test.
Grade 5	• Grade five students are more mature. • A grade five test allows for development of late-bloomers/English learners. • A grade five test serves as a summative, capstone test looking back on elementary grades.
Grade 6	• A grade six test serves as a summative, capstone test looking back on elementary grades. • Grade six is the first year of middle school so a grade six test gives middle school teachers a platform to build on.
Grade 7	• A grade seven test allows for remediation/intervention at eighth grade. • A grade seven test allows for eighth grade teachers to prepare students for high-school standards. • Grade seven is the midpoint for grades six to eight middle schools.
Grade 8	• A grade eight test serves as capstone/summative test for middle school; most middle schools end at/have an eighth grade. • Grade eight test results inform high school instruction/placement of students. • Grade nine is high-school level; choosing grade nine over grade eight means middle schools would not be tested. • Testing at eighth grade ensures that all students who received instruction on either the domain-specific or integrated model offered by NGSS would be prepared. • By eighth grade, students should have exposure to all three disciplines (Earth Science, Life Science and Physical Science).
Grade 9	• A grade nine test would be a benchmark to inform high school instruction. • Grade nine students are more mature. • Ninth grade students should know Earth Science. • A grade nine test allows for assessing middle school science learning.

Rationale	Codes
Grade 10	• A grade ten test is a continuation of past/current practice. • Testing in eleventh and twelfth grades is undesirable. • Testing in grade ten would be at the end of two required years of high school science. • Biology provides a common testing area.
Grade 11	• At grade eleven test is a later test (in high school) so allows more instruction. • A twelfth grade test is too late. • A grade eleven test allows for using test results for college admissions.
Grade 12	• A grade twelve test serves as a capstone/final assessment before leaving the K–12 system. • Twelfth grade is not as tested as eleventh grade. • A grade twelve test provides incentive for four years of high school science.
For Selection of Content Domain, Type of Assessment, and Item Type for the Grade Selected	
Overall responses	• All choices promote critical thinking. • All choices match with the NGSS.
Content Domain	• The test should match/correspond with the NGSS. • The content reflects real science. • Students should know basics across all disciplines. • The selected content domain is a foundational content domain. • The content should cover a full grade span (not just a selected grade level within the ESEA grade span). • Biology/life science is a common course that most will have taken.
Type of Assessment	• The assessment type provides a better understanding/measure of examinee ability. • The assessment type affords flexibility. • The assessment type takes advantage of technology advances. • The assessment type affords familiarity/appropriateness for the student's age. • Assessment types should mirror Smarter Balanced format.
Item Type	• The item type(s) allows the assessment of specific skills. • A variety of item types is beneficial. • Item types should follow the Smarter Balanced English–language arts/literacy and mathematics examples. • The item type(s) emphasizes hands-on/de-emphasizes memorization. • The item type(s) allows for assessing multiple levels of cognition. • The item type(s) allows for access to all students. • The item type(s) matches/corresponds with the NGSS. • The item type(s) reflects authentic/real science.

Survey, Part 3: Measurement Considerations

Rationale	Codes
For Matrix Sampling	
Yes	• Matrix sampling lowers the testing burden. • Matrix sampling is useful to use to inform aggregate decisions such as program evaluation. • Matrix sampling allows for testing more standards and/or can better assess the NGSS. • Matrix sampling provides more valid, accurate, or statistically sound results. • Matrix sampling allows for including more complex tasks in the assessments. • Matrix sampling allows for depth over breadth. • Matrix sampling de-emphasizes accountability for a single teacher. • Matrix sampling encourages best instructional practices. • Matrix sampling represents best practice or is more fair. • Matrix sampling ensures students are prepared for all standards. • Matrix sampling prevents teaching to the test. • The respondent wants partial matrix sampling (common set of standards assessed across students). • Matrix sampling allows for focus on thought processes applicable to all science. • Matrix sampling provides an opportunity to learn. • Matrix sampling is cost-effective. • Matrix sampling does not judge performance.
No	• The respondent values giving individuals scores, identifying individual strengths/weaknesses, and tracking student growth, but has concerns that matrix sampling would preclude such inferences. • The respondent has concerns with the accuracy and fairness of sampling (e.g., that certain types of students would receive certain standards). • The respondent values testing students on the same standards (with the same test). • The respondent values testing *all* students on *all* standards. • The respondent values using test scores to inform instruction, but has concerns that matrix sampling would preclude such test use. • The respondent values fair comparisons among students and believes that matrix sampling does not allow for comparability. • The respondent has concerns that matrix sampling is not accurate for small samples. • The respondent has concerns on not getting adequate information for teachers, schools, or districts. • The respondent has concerns that matrix sampling is not in line with the NGSS. • The respondent has concerns that matrix sampling encourages teachers not to teach all students. • The respondent has concerns that matrix sampling would be problematic/ introduce complications in evaluating teacher performance. • The respondent believes that students should know which standards they will be tested on to prepare.

Rationale	Codes
For Population Sampling	
Yes	• Population sampling provides information on key demographic groups and promotes equity. • Population sampling is cost effective. • Population sampling reduces the testing burden. • Population sampling informs instructors and curriculum developers. • Population sampling informs aggregate-level decisions. • The respondent mistakenly thinks that population sampling means testing all students. • Population sampling provides useful data.
No	• The respondent values testing all students. • The respondent has concerns on accuracy, fairness, and equity of sampling (e.g., belief that it is unfair to generalize performance of a group based on a selected subset of that group). • The respondent values providing feedback to students, teachers, schools, or LEAs, but has concerns that population sampling would preclude this use of test score data. • The respondent values using test scores to inform instruction, but has concerns that population sampling would preclude such test use. • The respondent suggests that instead of using population sampling, data analysts/researchers can sample from test scores after testing all students. • The respondent has concerns that population sampling complicates test administration (e.g., what to do with non-test-takers during testing periods). • The respondent has concerns that population sampling places the testing burden on the selected subset. • The respondent has concerns that population sampling is just politics or a political game. • The respondent has concerns that population sampling de-motivates students to perform well on the test and/or in science class. • The respondent has concerns on not getting information on the subset that was not tested. • The respondent has concerns that population sampling complicates testing when only a subset of students is tested.

Rationale	Codes
For Selected Scoring Method	
Automated Scoring	• Automated scoring provides more fair/objective (or less biased) scoring. • Automated scoring provides faster/more timely feedback. • Automated scoring is cost-effective. • Automated scoring is better than local scoring in that it can provide invalid/biased results and subjectivity in scoring. • Automated scoring is sophisticated enough now and is reaching reliability levels of humans. • Automated scoring alleviates the burden on local teachers to score. • The respondent wants centralized scoring as well. • The respondent wants local scoring as well. • Automated scoring uses advances in technology (does not need to rely on humans). • Automated scoring is more efficient. • Automated scoring is the simplest method.
Centralized Scoring	• Centralized scoring promotes training of raters and working together. • Centralized scoring provides more fair/objective (or less biased) scoring. • The respondent expresses distrust in automated scoring. • Centralized scoring is better than local scoring in that it can provide invalid/biased results and subjectivity in scoring. • The respondent believes that centralized scoring is used for scoring AP and/or the Golden State Exams and wants to follow their example. • Centralized scoring provides faster/more timely feedback. • Centralized scoring is easier to monitor. • Centralized scoring provides accurate/reliable results. • Centralized scoring is an effective/good use of resources. • Centralized scoring provides professional development opportunity for teachers.
Remote Scoring	• Remote scoring minimizes bias, is more consistent and/or is less subjective. • The respondent values human raters and is wary of automated scoring. • Remote scoring is cost effective (especially in comparison with centralized scoring, as there are no travel or lodging expenses). • Remote scoring is better than local scoring in that it can provide invalid/biased results and subjectivity in scoring. • Remote scoring allows more eligible raters to participate (as there are no geographical constraints). • The respondent believes that remote scoring is used and works with the College Board/AP scoring. • Remote scoring is the most flexible scoring option. • Remote scoring provides faster/more timely feedback. • Remote scoring provides professional development opportunity for teachers. • Remote scoring uses trained professionals.

Rationale	Codes
Local Scoring	• Local scoring allows for geography and demographic composition to be taken into account. • Local scoring provides feedback to teachers. • Local scoring provides faster and/or more timely feedback. • The respondent is wary of and/or does not trust automated scoring. • Local scoring involves training and oversight. • The respondent believes teachers know their students best. • The respondent does not trust centralized scoring. • Local scoring provides professional development to teachers. • Local scoring demonstrates respect for teachers. • Local scoring provides more professional scoring.

Survey, Part 4: Overall Feedback about the Future Science Assessment System

Rationale	Codes
For any other considerations	• The respondent emphasizes testing twenty-first century skills/real life scenarios and skills. • The respondent emphasizes problem solving/critical thinking in assessments. • The respondent wants assessments like the Golden State Exams/performance/ labs/practicum assessments. • The respondent wants supports for student learning/formative purposes. • The respondent emphasizes not taking time away from instruction/spend less time testing. • The respondent emphasizes attention to special groups such as English learners and accessibility such as keyboards skills/equity issues. • The respondent emphasizes not testing facts. • The respondent wants to test all students in all grades. • The respondent emphasizes college and career readiness skills. • The respondent wants timely turnaround of score reporting. • The respondent emphasizes not focusing on particular content domains. • The respondent emphasizes testing science earlier in elementary school so science would get taught. • The respondent emphasizes providing useful information to schools and parents/guardians.

References

California Department of Education. (n.d.). *Pre-publication version of the standards.* Retrieved from http://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp

California Department of Education and Educational Testing Service. (2014). *California Assessment of Student Performance and Progress (CAASPP) Science Stakeholder Meetings*. [Presentation slides]. Slides 99–101.

Corcoran, T., Mosher, F., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform* (p. 20). Philadelphia, PA: Consortium for Policy Research in Education.

Davey, T. (2011). *Practical considerations in computer-based testing*. Princeton, NJ: ETS Research & Development Division.

Livingston, S. (2009). Constructed-response test questions: why we use them; how we score them. *R & D Connections,* No. 11. ETS Research & Development Division.

National Center for Education Statistics. (2012). *The nation's report card: science in action: hands-on and interactive computer tasks from the 2009 science assessment* (NCES 2012–468) (p. 2). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center on Educational Outcomes. (2013). *Alternate assessments for students with disabilities*. (para. 1). Retrieved August 27, 2014, from http://www.cehd.umn.edu/nceo/topicareas/alternateassessments/altassesstopic.htm

National Research Council. (2014). *Developing assessments for the Next Generation Science Standards* (p. 33). Committee on Developing Assessments of Science Proficiency in K–12. (Board on Testing and Assessment and Board on Science Education, J.W. Pellegrino, M.R. Wilson, J.A. Koening, & A.S. Beatty, Eds.). Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

Next Generation Science Standards. (2014). *Development overview.* (para. 2–4). Retrieved August 1, 2014, from http://www.nextgenscience.org/development-overview

Next Generation Science Standards. (2014). *Framework.* (para. 2). Retrieved August 27, 2014, from http://www.nextgenscience.org/development-overview

Next Generation Science Standards. (2014). *How to Read the Next Generation Science Standards.* (para. 1–2). Retrieved August 27, 2014, from http://www.nextgenscience.org/how-to-read-the-standards

Next Generation Science Standards. (2014). *Three Dimensions*. (para. 4). Retrieved August 26, 2014, from http://www.nextgenscience.org/three-dimensions

Pruitt, S. (2013). *Next Generation Standards.* [Presentation slides]. Retrieved September 22, 2014, from http://www.cde.ca.gov/pd/ca/sc/ngssintrod.asp

Reiser, B. (2013). *What professional development strategies are needed for successful implementation of the Next Generation Science Standards?* Princeton, NJ: K–12 Center at ETS.

The Glossary of Education Reform. (2014). High-stakes test. (para. 1). Retrieved August 27, 2014, from http://edglossary.org/high-stakes-testing/

Willard, T. (2013). *A look at the Next Generation Science Standards*. National Science Teachers Association. Downloaded August 27, 2014, from http://nstahosted.org/pdfs/ngss/InsideTheNGSSBox.pdf