

Independent Evaluation Study of Certain Students Who Used Modifications and/or Accommodations on the California High School Exit Examination (CAHSEE)

Final Report

Submitted to:

Diane Hernandez, Contract Monitor
Statewide Assessment Division
California Department of Education
1430 "N" Street
Sacramento, CA 95814

Submitted by:

American Institutes for Research
2800 Campus Drive, Suite 200
San Mateo, CA 94403

April 1, 2010

Table of Contents

Executive Summary	1
Introduction.....	2
Task 2: Identification of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE	3
Purpose and Research Questions.....	3
Conceptual Framework	3
Population and Sample.....	5
Methods and Procedures	12
Data Analysis	15
Task 2 Conclusions and Discussion	24
Task 3: Identify and Analyze Possible Alternative Means.....	26
Purpose and Research Questions.....	26
Conceptual Framework	27
Population and Sample.....	28
Methods and Procedures	31
Data Analysis	33
Task 3 Conclusions and Discussion	35
Public Comment.....	36
Discussion and Recommendations	37
References.....	44
Appendix A: Examples of Assessment Probes in Mathematics	46
Appendix B: Examples of Assessment Probes in English Language Arts	47
Appendix C: Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (eight items).....	48
Appendix D: Handouts to AB 2040 Panel.....	54

Executive Summary

The purpose of this study was to investigate and report on a subgroup of high school students who had taken the California High School Exit Examination (CAHSEE) with modifications and/or accommodations specified in their respective individualized education programs (IEPs) or Section 504 plans, and who had not passed the CAHSEE, but who had satisfied, or would satisfy, all other requirements for graduating from high school.

The study had two major tasks. In the first of these we conducted in-depth individual assessments, based on released CAHSEE items and employing cognitive interviewing techniques, with a sample of students meeting the criteria of the study.

- Two of 42 students in the mathematics sample (about 5 percent) and 15 of 47 students in the English-Language Arts (ELA) sample (about 32 percent) had mastery of CAHSEE content. Projecting to the population, we estimate that the number of eligible students in the 2007–08 eleventh grade cohort who had established mastery of CAHSEE content by the end of their senior year was greater than zero but less than 300 in mathematics, and between 550 and 900 in ELA.

In the second major task we investigated alternative means that might better allow qualified students to demonstrate mastery. A pilot test of a written examination employing item and form adaptations (many of which were also used in the individual assessments) was carried out.

- Based on the evidence, we believe that in-depth individual assessments, combined with item and form adaptations that do not change the construct being measured and incorporate principles of universal design, offer the most promise for a fair and feasible alternative means for students to demonstrate content mastery.

We also reviewed the strategies employed by other states that provide alternative means for students to satisfy their high school exit examination requirement and receive a fully equivalent diploma. The most common strategy is to evaluate a collection of evidence.

- We believe that the California Department of Education should be cautious in adopting this approach because it is difficult to confirm that the work submitted actually reflects the student's own knowledge and skills. Unless this option is pursued carefully, it could have the unfortunate effect of enabling students who really do not know the content assessed by CAHSEE to inappropriately meet graduation requirements.

Finally, we note that many students assessed in this study do not have mastery of the content assessed by CAHSEE. Curriculum and instruction for these students need to be reviewed, and greater attention should be focused on content mastery when students are initially exposed to the CAHSEE content. In addition, students who fail to achieve passing scores on their first CAHSEE attempt should be considered for diagnostic testing and early interventions to address skill deficiencies.

Introduction

The California Department of Education (CDE) contracted with the American Institutes for Research (AIR), a non-profit education research organization, to conduct the Independent Evaluation Study of Certain Students Who Used Modifications and/or Accommodations on the California High School Exit Examination (CAHSEE). The purpose of the study was to investigate and report on a subgroup of high school students who had taken the CAHSEE with modifications and/or accommodations specified in their respective individualized education programs (IEPs) or Section 504 plans, and who had not passed the CAHSEE, but who had satisfied, or would satisfy, all other requirements for graduating from high school. This final report describes the process and findings of the study and concludes with recommendations to the CDE.

The study comprised two major research tasks¹:

- Task 2: Identify certain students who used modifications and/or accommodations on the CAHSEE
- Task 3: Identify and analyze possible alternative means

Task 3 was contingent on finding that there were students, among those who met the criteria for inclusion in task 2, who had mastered CAHSEE content in reading, writing, and mathematics.

This report includes a comprehensive discussion of our work on tasks 2 and 3, including our conceptual framework, population and sample, methods and procedures, data analysis, and findings for each task. We also summarize the comments received at the two required public meetings conducted under task 1. Finally, we conclude with a discussion and a set of recommendations on the types of alternative means that the State of California might adopt to better meet the needs of certain students with disabilities (SWDs). As part of this discussion, we present such evidence as is available from this study and from the experience of other states that have adopted alternative means as to the efficacy, feasibility, and validity of the various alternatives.

¹ The first study task (task 1) involved a variety of non-research tasks, including the conduct of two required public meetings to solicit comments on the study. Tasks 4 through 6 covered report writing, security, and consulting services.

Task 2: Identification of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Purpose and Research Questions

The central purpose of task 2 was to investigate the knowledge and skills of students with IEPs or Section 504 plans who were nearing the end of high school and were unable to perform at a passing level on the CAHSEE, despite the use of modifications and/or accommodations. Specifically, task 2 was designed to answer four research questions about a population of students who took CAHSEE as eleventh grade students sometime during the 2007–08 school year. Most of these students were in twelfth grade during the 2008–09 school year, when the study was conducted.

The four research questions were:

1. Among eleventh-grade students who took the CAHSEE with modifications and/or accommodations during the 2007–08 school year and failed to achieve a passing score on one or both parts of the examination, what proportion had completed course work that puts them on track for graduation in summer 2009?
2. Among the subset of these students who were on track for graduation, what proportion appear to have mastered CAHSEE content in reading, writing, and mathematics?
3. Among the contrasting subset of these students who do *not* appear to have mastered CAHSEE content, what proportion appear to have had reasonable opportunity to learn in that they have been exposed to a curriculum that is significantly aligned with CAHSEE content standards?
4. Finally, for those who *do* appear to have mastered CAHSEE content, what disability-related factors appear to have created barriers to passing the examination?

Conceptual Framework

California wishes to hold SWDs to the same standards as other students who earn diplomas in California high schools. However, the percentage of SWDs who pass the high school exit examination has remained close to 50 percent in recent years, even for students who have taken the examination on multiple occasions. Of particular interest are those SWDs who have failed to pass (or to obtain scores that are high enough to apply for a waiver) despite having availed themselves of the remedies available under the law—the use of modifications and/or accommodations specified by their IEP or 504 plan. Some of these students may in fact have mastered the content at a level equivalent to that required by CAHSEE, but this equivalence is difficult to verify without an agreed on and validated alternative measure. In the absence of such

a measure, we selected, for use in task 2, an in-depth individual assessment, based on a sample of released CAHSEE items and using cognitive interviewing techniques.

This approach was selected both because it was feasible given the time and resource constraints and because it eliminated many of the barriers that may prevent students from demonstrating their knowledge. A further advantage is that, since it used actual CAHSEE items, we did not need to demonstrate that the assessment tasks instantiated CAHSEE content standards.

One-on-one oral assessments are common adaptations for testing SWDs. To this basic structure we added probes built on cognitive interviewing principles and designed to provide further evidence regarding the students' knowledge and skills. Similar types of cognitive interviews have been used in many previous studies to probe the qualifications of either the test taker or the test instrument itself. For example, Jakwerth, Stancavage, and Reed (1999) used cognitive interviews in a National Assessment of Educational Progress (NAEP) Validity Panel study to investigate why students do not respond to questions (in other words, why students leave certain questions blank) on the NAEP. The analysis concluded that, in most cases, students who skipped questions did so because they did not know the content being tested. However, in the case of some constructed response questions, the students did know at least enough to earn a partial score on the item, but failed to respond because they lacked confidence in their knowledge.

Levine and his colleagues (Levine and Huberman, 2000; Reed, Levine, and Huberman, 2000) used similar techniques to evaluate the quality of test questions being prepared for the proposed Voluntary National Test (VNT). In particular, the cognitive interviews were used to examine whether or not the test items actually tapped the intended construct, as judged by the line of reasoning through which students produced correct or incorrect responses. Both SWDs and regular students were included in the interview process. Del-Rio and Levine (2000) also employed cognitive interviews to study the efficacy of accommodations for English learners.

We consider the application of cognitive interview techniques in this evaluation to be an instance of test validation research. Several citations from the literature reinforce the appropriateness of cognitive interview techniques for research, including research with high stakes tests. For example, Zucker, Sassman, and Case (2004) note that "carefully executed cognitive labs yield valuable qualitative information that complements quantitative data from empirical research and statistical analysis. These qualitative data have repeatedly been demonstrated as invaluable in refining and validating assessments at both the early and late stages of development. As a leading publisher of high-quality assessment systems, Pearson Inc. has used cognitive labs as a regular activity in the development of educational assessments, including the Stanford Suite of Assessment Products and assessments for several state education agencies."

In another instance, Measured Progress, Inc. and SRI International convened a group of researchers who had completed or were engaged in research involving SWDs and assessment of

reading achievement for an Invitational Research Symposium on Alternate Assessments Based on Modified Achievement Standards in Reading. The symposium, which was held in Arlington, Virginia on July 24, 2008, led to the conclusion that appropriately structured cognitive interviews were a valuable tool for use in studies on the design, development, and validation of alternate assessments based on modified academic achievement standards in reading (Almond, et al., 2009).

This study had substantial limitations on time and resources. The contract was awarded at the end of January 2009, and all instrument development, sampling, and field work for task 2 had to be completed four and one-half months later, before the sampled students dispersed for the summer. Moreover, the budget as well as the schedule limited the amount of travel that could be undertaken. Given these constraints, we adopted a design for task 2 that involved in-depth individual assessments with approximately 100 students, clustered in ten districts and one special school for SWDs.

In a later stage of the analysis, we extrapolated from the assessed sample to estimate the statewide percentages of students in the subgroup of interest who were likely to have mastered CAHSEE content. Where feasible, we used variables from the CAHSEE and California Special Education Management Information System (CASEMIS) data files to refine these estimates by making model-based predictions of the prevalence of students with the same attributes as those determined to have different outcomes in our sample. Variables that were entered into the model included such factors as students' scores on prior CAHSEE attempts and percent time in general education classes—a variable that Becker and Watters (2008) found to be a strong predictor of the likelihood that a student would pass CAHSEE using the allowed modifications and/or accommodations.

While we feel that this was the best design given the resources available to the project, we acknowledge that it also has limitations. These include the relatively small sample of students evaluated in depth, the fact that students did not necessarily participate in the study with the same level of motivation that they would bring to an actual CAHSEE administration, and the fact that the in-depth individual assessments, being relatively time consuming, included fewer than half as many items as the operational CAHSEE.²

Population and Sample

Most students take the CAHSEE for the first time in tenth grade. If they do not pass initially, they can continue to take the CAHSEE multiple times in subsequent years. The population of interest to us is the population that has still not achieved a passing score on one or both subtests of the CAHSEE by the time they would otherwise graduate, despite the use of modifications and/or accommodations. Although this would suggest the use of a twelfth grade sample, we

² On the other hand, the additional information generated through the cognitive probes allowed a more reliable judgment as to whether or not the student knew the underlying content for each item that was assessed.

decided to base our sampling frame on students who were in eleventh grade during the 2007–08 school year so that a sample of these students could be assessed (and thereby evaluated for mastery of CAHSEE content) during the course of the contract.

More specifically, the sampling frame that we chose for task 2 included all eleventh grade students who took CAHSEE at any time during the 2007–08 school year,³ used a modification and/or accommodation on the mathematics and/or English-language arts (ELA) subtests, and scored below the passing level (below 350) on at least one of the subtests for which they had used a modification and/or accommodation.⁴ The number of students meeting these criteria was 11,163.⁵

We then compiled all CAHSEE records for these 11,163 students through the most recently available test administration (which was November 2008 at the time of sample selection) in order to screen out students who no longer scored below the passing level on either subtest of the CAHSEE.⁶ There were 8,740 students remaining in this pool of study-eligible students, which formed the basis for our in-depth study sample.⁷ Among the 8,740 students, 7,025 had failed to achieve a passing score on the mathematics subtest, and 6,474 had failed to achieve a passing score on the ELA subtest. More than half (54 percent) had failed to achieve a passing score on either subtest.

We wanted to screen for “on track to graduate” and then assess a sample of students that would be as representative as possible of the 8,740 students who had still not passed CAHSEE as of November 2008. However, it was also necessary to constrain our sample geographically so that the individual assessments could be completed within the resources available to the project. We therefore began by selecting a sample of districts, then schools within districts, and finally students within schools.

District Selection

Our design called for assessing ten students in each of ten districts, for a total sample size of 100. Given that time and cost constraints only allowed us one data collection trip to each district, we

³ Our data file included records from the following CAHSEE administrations: July 2007, October 2007, November 2007, December 2007, February 2008, March 2008, and May 2008.

⁴ Students entered the sample if they had used a modification and/or accommodation (and failed to achieve a passing score) on any CAHSEE administration during 2007–08. It was not necessary that they had used the modification and/or accommodation on their most recent attempt.

⁵ We initially thought the pool was 11,173 students, but later discovered that ten of these students each had two records in the file. This arose either because of two statewide student identifiers (SSIDs) being assigned to the same student (eight students) or because of a single student appearing to have two CAHSEE records on the same date (two students). It is possible that other duplicates may have gone undetected.

⁶ Additional screens were applied later to remove students from the assessment and/or analysis samples who completed their CAHSEE requirements between December 2008 and May 2009.

⁷ The other criterion defining the population of interest—that the students be on track to graduate—could not be confirmed from the CAHSEE records. Instead, we asked participating districts to screen potential in-depth study subjects against this criterion. In most cases, districts had to collect the information from the students’ schools.

planned to qualify and issue invitations to 20 students per district so that we would be able to use our time on site efficiently. In order to secure 20 invitees per district, and because we were not able to screen students for being on track to graduate until after we had confirmed district participation, we decided that it would be impractical to sample any districts that had fewer than 50 students in our pool of 8,740 study-eligible students. Forty-five districts met this criterion, and these 45 districts represented 50.7 percent of all of the study-eligible students that we identified.

Based on the population distribution of the state, the 45 qualifying districts were categorized into southern California districts and all other districts.⁸ Five districts were sampled from each category, with probability proportional to size (where size equals number of study-eligible students). Two districts, one in southern California and one from the “all other” category, declined to participate. Replacement districts were sampled using the same protocol. In addition, we sampled one of the special schools for SWDs that are not housed within regular districts. The names of the ten participating districts and the participating special school are shown in table 1, along with the numbers of study-eligible students in each.

Table 1. Participating districts and special school

District name	Number of study-eligible students¹
Alhambra Unified School District	53
Antioch Unified School District	71
Desert Sands Unified School District	79
Fresno Unified School District	176
Long Beach Unified School District	132
Los Angeles Unified School District	667
Modesto City Schools District	172
Moreno Valley Unified School District	58
Oakland Unified School District	61
San Francisco Unified School District	97
California School for the Deaf, Riverside	10
TOTAL	1576

Source: CAHSEE administration records for the period July 2007 through November 2008.

¹ Study-eligible students are students who were in eleventh grade during the 2007–2008 school year and, as of November 2008, had failed to achieve a passing score on at least one subtest of the CAHSEE, despite the use of modifications and/or accommodations.

Selection of Schools and Students within District

Once we had secured district participation, we further delimited our in-depth study sample by selecting primary and backup schools within each district. To select the primary schools, we first identified all schools that had ten or more study-eligible students. If there were more than three

⁸ We define southern California as comprising the counties of San Bernardino, Los Angeles, San Luis Obispo, Ventura, San Diego, Riverside, Orange, Santa Barbara, Kern, and Imperial. About 60 percent of the California population resides within these nine counties.

such schools, we randomly selected three with probability proportional to size (where size equals the number of study-eligible students).⁹ After the primary schools were chosen, up to three backup schools per district were selected from the remaining schools, using the same method.

We then asked the districts to screen all of the study-eligible students at the sampled schools to determine their educational status and suitability for inclusion in the individual assessment sample. That is, we wanted to know which of the study-eligible students a) were on track to graduate, b) had not passed CAHSEE since November 2008, and c) were still attending the sampled schools.¹⁰ The screening process was quite labor intensive and generally required the districts to obtain the information directly from the schools. For this reason districts were not able to screen all of the study-eligible students in the entire district for being on track to graduate, although this had been our original plan.

Two primary schools were replaced by backup schools at this stage of the sampling. In one case this was because a primary school closed temporarily because of an H1N1 flu outbreak. In the other case, the substitution was made because of the relative numbers of assessment eligible students at each school (that is, the relative numbers of students at each school who were on track to graduate and had not passed CAHSEE since November 2008).

In the final selection step, we randomly selected 20 assessment-eligible students from the primary schools (or their replacements) and invited them to participate in our study. If there were fewer than 20 assessment-eligible students, all were invited. Participating students were offered \$20 to acknowledge their efforts.¹¹ Parent permission was sought for all students younger than 18 and, at the district's request, for all Long Beach students, regardless of age. In all cases, informed consent was also obtained from students.

In total, 212 students were invited for individual assessments, and 106 individual assessments were completed. This includes 51 students who completed the assessment protocol for mathematics and 55 students who completed the assessment protocol for ELA. Tables 2 through 5 describe the demographics and disabilities of the students who were assessed. For comparison, tables 2 through 5 also display the demographic and disability data for the population of 8,740 study-eligible students from which we derived the assessment sample and, where available, for all twelfth grade students in California in 2008–09.

Both the assessment sample and the population of study-eligible students include high percentages of African American and Hispanic students and high percentages of English

⁹ School sampling was not used in Long Beach, which requested that we enroll all of their high schools in the study.

¹⁰ The study did not have sufficient time or resources to locate students who were no longer attending the sampled schools.

¹¹ In Long Beach, at the district's request, students were not given a financial incentive to participate.

learners. More than two-thirds of each group have a specific learning disability as a primary disability, and males somewhat outnumber females.¹²

Table 2. Distribution of assessed students, all study-eligible students, and all twelfth grade students in California, by race/ethnicity

Race/ethnicity	Assessed students		Study-eligible population ¹		12 th grade students in CA (2008–09)	
	N	%	N	%	N	%
American Indian or Alaskan Native	0	0.0	98	1.1	3,839	0.8
African American or Black	36	34.0	1,573	18.0	37,289	7.8
Asian	11	10.4	356	4.1	58,239	12.2
Hispanic or Latino	45	42.5	4,552	52.1	206,340	43.3
Pacific Islander	1	0.9	53	0.6	3,164	0.7
White (not of Hispanic origin)	12	11.3	1,990	22.8	155,898	32.7
Declined to state/Multiple Response	1	0.9	96	1.1	11,463	2.4
Unknown	0	0.0	22	0.3	-	-
TOTAL	106	100.0	8,740	100.0	476,232	100.0

Source: CAHSEE administration records for the period July 2007 through November 2008 and CA Department of Education DataQuest Database, 2008–09 Enrollment by Gender, Grade and Ethnic Designation, Statewide Report.

¹ The study-eligible population includes all students who were in eleventh grade during the 2007–2008 school year and, as of November 2008, had failed to achieve a passing score on at least one subtest of the CAHSEE, despite the use of modifications and/or accommodations.

¹² There are several reasons why the demographics of the assessment sample do not exactly match those of the study-eligible population. These include the fact that the study-eligible population includes students who were not on track to graduate at the time of the data collection, as well as the fact that all the assessed students were sampled from districts that each had at least 50 study-eligible students. As noted previously, only 50.7 percent of all of the study-eligible students attend school in such districts.

Table 3. Distribution of assessed students and all study-eligible students, by English proficiency level¹

English proficiency level	Assessed students		Study-eligible population ²	
	N	%	N	%
English only	60	56.6	5,257	60.2
Initially fluent English proficient	1	0.9	268	3.1
English learner	44	41.5	2,770	31.7
Reclassified fluent English proficient	1	0.9	268	3.1
Missing	0	0.0	22	0.3
Unknown	0	0.0	155	1.8
TOTAL	106	100.0	8,740	100.0

Source: CAHSEE administration records for the period July 2007 through November 2008.

¹Data on English proficiency level for all California students is not available by grade level.

² The study-eligible population includes all students who were in eleventh grade during the 2007–2008 school year and, as of November 2008, had failed to achieve a passing score on at least one subtest of the CAHSEE, despite the use of modifications and/or accommodations.

Table 4. Distribution of assessed students, all study-eligible students, and all twelfth grade students in California, by gender

Gender	Assessed students		Study-eligible population ¹		12 th grade students in CA (2008–09)	
	N	%	N	%	N	%
Male	61	57.5	5,628	64.4	242,414	50.9
Female	45	42.5	3,104	35.5	233,818	49.1
Unknown	0	0.0	8	0.1	0	0
TOTAL	106	100.0	8,740	100.0	476,232	100.0

Source: CAHSEE administration records for the period July 2007 through November 2008 and CA Department of Education DataQuest Database, 2008–09 Enrollment by Gender, Grade and Ethnic Designation, Statewide Report.

¹ The study-eligible population includes all students who were in eleventh grade during the 2007–2008 school year and, as of November 2008, had failed to achieve a passing score on at least one subtest of the CAHSEE, despite the use of modifications and/or accommodations.

Table 5. Distribution of assessed students, all study-eligible students, and all twelfth grade students in California, by primary disability

Primary disability	Assessed students ¹		Study-eligible population ^{1,2}		12 th grade students in CA (2008–09)	
	N	%	N	%	N	%
No Special Education Services	2	1.9	833	9.5	-	-
Mental Retardation	6	5.7	300	3.4	4,379	9.4
Hard of Hearing	1	0.9	78	0.9	543	1.2
Deafness	1	0.9	78	0.9	337	0.7
Speech or Language Impairment	7	6.6	328	3.8	1,526	3.3
Visual Impairment	0	0.0	49	0.6	349	0.7
Emotional Disturbance	2	1.9	579	6.6	3,832	8.2
Orthopedic Impairment	1	0.9	60	0.7	1,336	2.9
Other Health Impairment	5	4.7	431	4.9	3,536	7.6
Specific Learning Disability	76	71.7	5,831	66.7	28,142	60.1
Multiple Disabilities	0	0.0	18	0.2	419	0.9
Autism	4	3.8	126	1.4	2,155	4.6
Traumatic Brain Injury	1	0.9	29	0.3	250	0.5
TOTAL	106	100.0	8,740	100.0	46,804	100.0

Source: CAHSEE administration records for the period July 2007 through November 2008 and CA Department of Education DataQuest Database, 2008–09 Special Education Enrollment by Grade and Disability, Statewide Report.

¹ For students receiving special education services under an IEP, the CAHSEE file contains a record of the student's primary disability. Other students, including students classified under Section 504, are coded as "No special education services."

² The study-eligible population includes all students who were in eleventh grade during the 2007–2008 school year and, as of November 2008, had failed to achieve a passing score on at least one subtest of the CAHSEE, despite the use of modifications and/or accommodations.

Table 6 provides information on the disposition of the 106 students who were invited for individual assessments, but not assessed. Students in the top three rows of the table—those who no longer attended the school, were not currently in session, or had recently passed the CAHSEE—were not actually eligible for the in-depth study and would have been screened out during the final stage of sample selection if data provided by the district had been complete.

Students in the bottom row—additional students not needed—are indicative of the strategy, noted earlier, of intentionally issuing more invitations than required. This was done in order to make sure that the assessment administrators would be able to fully utilize their time on site once other reasons for non-participation, such as late determination of ineligibility, refusals, and absences, were accounted for. When there were more potential subjects than required, study staff sampled randomly from among the available students.

Table 6. Disposition of students invited for individual assessments but not assessed

Disposition	N	%
No longer attends school	6	5.7
Year round school/C track not in session ¹	12	11.3
Passed CAHSEE	13	12.3
School declined to participate ²	4	3.8
Parent refusal/consent form not returned	17	16.0
Student refusal	16	15.1
Absent	14	13.2
Additional students not needed	24	22.6
TOTAL	106	100.0

¹Los Angeles schools are year round schools. The school year for C track students ends in March. Our assessments were conducted in May and June.

²Long Beach requested that all six of its high schools be included in the study, but one school declined to participate.

Methods and Procedures

Two types of data were collected for students in the in-depth study sample. First, the students themselves participated in in-depth individual assessments designed to evaluate their mastery of CAHSEE content and to provide clues about disability-related factors that might have impeded their performance when they were tested under operational conditions, despite the use of allowable modifications and/or accommodations. The second type of data came from interviews with special education coordinators or teachers who were familiar with these students and were able to comment on the curriculum and instruction to which the students had been exposed.

Individual assessments

The student assessment consisted of a one-on-one, highly supported administration of a sample of released CAHSEE items in either mathematics or ELA. In order to maximize the likelihood that students would be able to demonstrate their underlying knowledge, all material was read aloud to the students, the students responded orally, and a calculator was made available for the mathematics subtest. In addition, students were given as much time as they needed to complete the assessment protocol. These are all modifications that are consistent with operational CAHSEE practice and can lead to a waiver—at the discretion of the district’s governing board—if the student achieves an equivalent of a passing score. They are also the most common modifications in operational use. Any other test variation, accommodation, or modification required by the student’s IEP was also provided. For example, a sign-language translator was provided to students who required one.

In addition, students completed a graphic organizer before they wrote the ELA essay to help them organize their thoughts, and, after students gave their initial answer to each mathematics or ELA question, administrators asked a series of scripted probes—based on cognitive interviewing techniques—that provided additional scaffolding. For instance, the probes might include asking the student how they would start a mathematics problem, rephrasing a question using modified

vocabulary, or rereading a portion of the ELA stimulus. Example questions with accompanying probes are shown in appendices A and B. The students' responses to the scripted probes provided additional evidence about their mastery of CAHSEE content.

The mathematics and ELA protocols each consisted of 30 items, selected to have the same proportional representation of content strands as the full CAHSEE and to provide as much coverage of the individual content standards as possible, given the shorter length. Within this framework, items for the protocols were randomly selected from among all released CAHSEE items. Consequently, the distribution of item difficulties was representative of the full bank of released CAHSEE items. Figure 1 shows how the items in the assessment protocols were distributed across the strands.

Figure 1. Distribution of items by content strand — operational CAHSEE and study protocol

Strand	Number of items in CAHSEE	Number of items in study protocol
Mathematics		
Number Sense	14	5
Statistics, Data Analysis, and Probability	12	5
Algebra and Functions	17	6
Measurement and Geometry	17	6
Mathematical Reasoning	8	3
Algebra	12	5
TOTAL MATHEMATICS	80 multiple choice	30 multiple choice
English Language Arts		
Word Analysis ¹	7	1
Reading Comprehension	18	7
Literary Response and Analysis	20	9
Writing Strategies	12	6
Writing Conventions	15	6
Writing Applications	1	1
TOTAL ELA	72 multiple choice + 1 essay	29 multiple choice + 1 essay

Source: California High School Exit Examination (CAHSEE) Technical Report: July 2006 - May 2007 Administrations

¹The exact number of word analysis, reading comprehension, and literary response and analysis items was constrained by item distributions associated with the released reading passages.

The assessment protocols were developed by item writers experienced in adapting high-stakes test items for SWDs, and they were pilot tested with students in other districts who meet the study criteria before being finalized. The assessment administrators were experienced AIR researchers who received a day of training that included an introduction to working with SWDs from a certified special educator, modeling by expert cognitive interviewers, and practice administering the assessment protocols.

Administrators were not trained or expected to make judgments about students' mastery of content. Rather, they were trained to administer the assessment protocol in a standardized

fashion, to record their own probes and the students' oral responses verbatim, and to collect the students' written work. In addition, all assessments were videotaped or audiotaped unless the student refused permission. In the first phase of the analysis, all of these data were submitted to an expert panel that evaluated each student's mastery. The panel included certified special education experts and ELA and mathematics content experts.

Administrators were instructed to give students as much time as they needed to complete all of the items on the assessment protocols; however, class schedules, site visit logistics, and student attention levels prevented some students from completing all 30 items. For example, some students could only participate for one class period because of testing-related activities or class projects, and some students who started the assessment less than two hours before the end of school were not able to stay after school to finish the assessment. The average completion time for those students who completed all 30 items in the mathematics protocol was 98 minutes and the average completion time for those students who completed all 30 items in the ELA protocol was 90 minutes. These times do not include bathroom and rest breaks, which were offered to all students.

Seventy-five out of 106 students completed all items on the test they were assigned (including the essay on the ELA form); 22 additional students completed at least half of the items, leaving nine students who completed less than half. Three of those nine students completed too few items to be included in the analysis of mastery.

Interviews with Special Education Coordinators or Teachers

The interviews with special education coordinators or teachers focused on the students' mathematics and ELA classes in high school, with a particular emphasis on twelfth grade classes. If the student had a general education placement in one or both of these subject areas, we asked what special supports, such as individual assistance, use of audiotapes, or graphic organizers, were offered to help the student engage with the material. If the student had a special education placement, we asked about the credentials of the teacher(s) and the grade level of the material that the student was studying.

In addition, we asked for a description of any remedial assistance offered to the student on the basis of his or her unsuccessful attempts to pass either part of the CAHSEE, and we asked more generally about any strategies that the respondent had found to be effective in helping students like these to prepare for or take the CAHSEE.

While these interviews provided some general information about students' opportunities to learn, we would have liked to collect much more fine grained information if it had been feasible to do so. During the planning phase we quickly learned that there was no centralized source in each school that could provide details about the current educational experiences of the students in the study. Details about educational experiences from prior years, when most CAHSEE content is

first introduced, were nearly impossible to obtain. For this reason, the interview protocols were pitched at a fairly general level.¹³

In addition, many of the potential special education coordinator or teacher respondents were not available during our site visits or refused to be interviewed.¹⁴ By the conclusion of data collection, special education coordinator or teacher interviews were conducted with 17 respondents, each of whom reported on an average of 3.5 students. In total, we obtained opportunity to learn data on 59 of the students in our in-depth study sample. These students were distributed across 7 districts and the Riverside School for the Deaf.

Data Analysis

Scoring Student Assessments

The first step in the data analysis was to score the student assessments and establish each student's status as "passing" or "not passing." The criterion that we used for passing was whether the student achieved the same percentage of raw score points as was equivalent to a CAHSEE scale score of 350 at the time of the 2004 standard setting. These percentages were 55 percent for the mathematics subtest and 60 percent for the ELA subtest.¹⁵ All multiple choice items count as a single raw score point, whereas 20 percent of the total ELA raw score is based on the essay score.

Student assessments were scored item by item. First, we made a record of every item that the student answered correctly on their initial try. (That is, before any probes were employed by the administrator.) Second, we made a record of every item that the student answered incorrectly *and* for which the student did not provide any additional evidence that could be pertinent to mastery—either by responding to administrator probes or by providing written work in the student answer booklet. We then photocopied the student answer booklets and protocol records for all of the remaining items (items for which the student's evidence of mastery was "uncertain") and organized them by test question.

At this point we convened a small panel of experts to evaluate the "uncertain" items and make a judgment as to whether or not credit should be given for the student's response. Two special education experts and two subject matter experts from local school districts comprised the panel. The experts worked through the data test question by test question. For each test question, they first considered the content standard that the item was intended to address.¹⁶ Then they reviewed

¹³ Probing students about their own educational experiences would have taken time away from the completion of the assessment protocols required to evaluate mastery.

¹⁴ The study was not funded to compensate special education coordinators or teachers for their time while participating in the interviews. This may have affected participation rates for this component of the study.

¹⁵ Due to equating, the raw score equivalent of a 350 scale score varies slightly from one CAHSEE form to the next. We therefore decided to use the score equivalents from the initial standard setting administration.

¹⁶ The CDE website provides the content standard classifications for all released items. As noted earlier, our assessment instrument was constructed from released CAHSEE items. Content standard classifications and released CAHSEE items can be found at this website: <http://www.cde.ca.gov/ta/tg/hs/resources.asp>

the probes in the assessment protocol and discussed the types of evidence that they would accept as indicative of mastery of CAHSEE content. After this discussion, each expert separately reviewed and scored all of the “uncertain” student responses to that test question. A response was judged as passing (received a score of 1) if it showed mastery of content vis a vis the content standard under which the test question was classified. It was also possible for the experts to assign a partial score (a score of 0.5) if the student response indicated partial mastery. Non-qualifying responses received a score of 0.

The individual scoring was followed by a reconciliation process in which the experts discussed any student responses for which they had not assigned the same scores. In many, but not all, cases, the experts agreed on a single score after their discussion. When they did not agree, their individual ratings were preserved, and the scores were subsequently adjudicated by the project manager and principal investigator, both of whom were present during the deliberations of the expert panel members. The first few test questions were discussed and scored by all four panel members. Panel members worked in pairs to score the remaining items.

The above procedure applied to the scoring of the multiple choice items. The student essays were scored by experienced contractor staff members using the four-point score rubric and anchor papers published by CDE. Two staff members scored each essay response. All discrepant scores were adjudicated by the table leader.¹⁷

All score data were entered into a spreadsheet, and score points for all correct or partially correct items were summed to establish an overall score for each student. As described in the methods and procedures section, some students skipped over items or ran out of time before they completed the assessment protocol. Skipped items were counted as incorrect, while not-reached items were counted as “not attempted.” Each student’s percent correct was calculated based on the percentage of *attempted* items that were answered correctly. (That is, students were not penalized for running out of time.¹⁸) However, an overall analysis of response patterns was conducted to determine whether a student who had achieved a “passing” percent correct based on fewer than 30 items should be counted as exhibiting mastery.¹⁹

¹⁷ Adjudication was done for all non-identical score pairs even though operational test programs typically only adjudicate non-adjacent scores.

¹⁸ However, there were three students in the sample who answered too few questions to be included in the analysis of mastery

¹⁹ In conducting the analysis of response patterns, it was necessary to consider which form of the interview protocol the student had completed. There was only a single form of the mathematics protocol in which items were ordered from least difficult to most difficult based on the item p-values provided by CDE. For ELA, however, there were three forms of the protocol in which the order of the items was varied. In form 1, students started with the reading passages (word analysis, reading comprehension, and literary response and analysis content strands). They then answered the multiple choice questions on writing strategies and writing conventions, and they ended by completing the essay for the writing applications strand. In form 2, students began with the multiple choice questions on writing strategies and writing conventions, then completed the essay, and finally addressed the reading passages. In form three, students began with the essay, then addressed the reading passages, and ended with the multiple choice items on writing strategies and writing conventions.

For example, scores were calculated separately for the first half and the second half of the mathematics protocol. Because the items were ordered by difficulty,²⁰ it was not surprising to see that nearly all students performed better on the first half of the test than they did on the second half. For this reason, a student who only completed half of the mathematics protocol would have had to have performed substantially above passing on the items he or she did complete in order to be classified as having mastery. (In fact, all of the students who were judged to have mastery in mathematics had completed the entire 30-item protocol.)

On the ELA protocol, we did not observe systematic differences in performance between the reading items and the multiple choice writing items. However, since an essay score of “1” was modal for this group of students, a student who failed to answer the essay could only be classified as having mastery if his or her score on the multiple choice items was high enough to offset a score of “1” on the essay.

Removing Students who Passed CAHSEE between December 2008 and May 2009

Before proceeding to estimate the numbers and percentages of students in the target population who were either on track to graduate and/or who exhibited mastery of CAHSEE content, we first had to remove from our file all students who had passed CAHSEE after the date on which we had compiled our initial sample—that is, after November 2008. We therefore obtained a file of CAHSEE data for the period from December 2008 through May 2009 and screened the students in our file against these new data. Many students had, in fact, passed CAHSEE during this time period, including some students in our in-depth study sample.

Among the 11,163 students with whom we started our analyses, there were 4,372 who had passed both subtests of CAHSEE by May 2009, leaving 6,791 who had not. The 6,791 included 1,222 students who had no CAHSEE attempts during the 2008–09 school year and 35 who were listed as being in tenth grade on their most recent CAHSEE record, suggesting that they should not have been included in our original pool of students. In order to be conservative in our estimates of numbers of students, statewide, who were on track to graduate or who exhibited mastery of CAHSEE content, we excluded the 1,222 students with no CAHSEE attempts and the 35 students with a grade ten designation.²¹ Twenty-seven other students were excluded because of problems with their data, leaving a pool of 5,507 potential students.

²⁰ Items were ordered by difficulty because we observed during the pilot test that many students became disheartened and stopped trying once they had confronted two or three items in a row for which they did not know the answers.

²¹ In the absence of other information, students with no records of CAHSEE attempts during the 2008–09 school year were excluded because of the likelihood that they had dropped out of school and/or had failed to accumulate sufficient credits to graduate at the end of the 2008–09 school year. There were, however, a few students in our in-depth study sample for whom we did not find any 2008–09 CAHSEE records despite the fact that they were still attending school and the school considered them to be on track to graduate at the end of the spring semester or by the end of the summer. None of the interviewed students who were evaluated as having mastery fell into this category.

Projecting to the Population—On Track to Graduate

We explored multiple approaches to answering our first research question: Among eleventh-grade students who took the CAHSEE with modifications and/or accommodations during the 2007–08 school year and failed to achieve a passing score on one or both parts of the examination, what proportion have completed course work that puts them on track for graduation in summer 2009?

The first approach was to examine the pool of 446 still-failing students for whom the districts provided information about their on-track status,²² calculate the percent on track, and apply this percentage to the population. The second approach was to estimate a model, based on available CAHSEE and CASEMIS variables, to predict on-track status for the 446 still-failing students for whom the outcome was known. We then used this model and a multiple imputation method to estimate the proportion of on-track students in our pool of 5,507 potential students.

Among the 446 known cases, 252 students, or nearly 57 percent, were reported to be on track to graduate (s.e. 2.35).²³ Calculating 57 percent of our total pool of 5,507 potential students, we obtained an estimate of 3,139 students who took CAHSEE with modifications and/or accommodations as eleventh grade students in 2007–08, had still not passed one or both subtests of CAHSEE as of May 2009, and were otherwise on track to graduate.

In the model-based approach, we used the 446 known cases to predict the odds of being on track to graduate. Predictors in the model included various known information on the students such as whether they had passed one of the two CAHSEE subtests, how recently they had attempted CAHSEE, percentage of the school day spent in general education, and demographic characteristics. We then used this model to impute multiple possible “on track” values (on track versus not on track) for each of the 5,507 students in the population. The results of the multiple imputations, which statistically summarize across the imputed values, suggested that 52.4 percent of the 5,507 students were on track to graduate (s.e. 1.8). This is equivalent to 2,885 students, but the procedure provides the percentage estimate only and does not identify actual students. In addition, there is, of course, uncertainty around the estimate. A 95 percent confidence interval suggests a true percentage anywhere between 48.8 percent (2,686 students) and 56.0 percent (3,083 students).

Projecting to the Population—Mastery of CAHSEE Skills

Our second research question asks: Among the subset of these students who are on track for graduation, what proportion appear to have mastered CAHSEE content in reading, writing, and mathematics? As with the analysis of on track to graduate, we can address this research question either by extrapolating from the percentage of our in-depth study sample that earned a passing

²² Students whom the districts reported as having left the sampled school were excluded from this analysis. Although many of these students had probably dropped out of school and would therefore not be on track to graduate, others may simply have transferred to another school or district.

²³ Recall that students who had left the sampled schools were not included in the denominator for this estimate.

score or by developing a model that predicts passing among the assessed students and then applying the model to the larger sample.

The model-based analysis is only possible if there are sufficient numbers of students in the in-depth study sample with each type of outcome (mastery and lack of mastery) to allow us to estimate a model. For this reason, as will be evident below, the model-based analysis could only be attempted in ELA. Furthermore, the model-based results for mastery are more tenuous than the results for the on-track-to-graduate analysis because of the much smaller size of the sample on which the model can be estimated (47 students in the ELA in-depth study pool versus 446 students in the pool of students with known status on the on-track variable).

Mathematics

After excluding students who subsequently passed the CAHSEE mathematics subtest, there were 42 students with analyzable data from the mathematics assessment.²⁴ Two of these students, or almost 5 percent, were evaluated as having mastery of CAHSEE content in mathematics. All of the students in the in-depth study pool were on track to graduate, so this can be extrapolated to 5 percent of on-track students.

Among the pool of 5,507 potential students (who were failing one or both subtests of the CAHSEE), 4,109 had still not passed the mathematics subtest as of May 2009. If we assume that 57 percent were on track to graduate, and 5 percent of the on-track students had mastery of CAHSEE content in mathematics, then we estimate that there are 117 students who took CAHSEE with modifications and/or accommodations as eleventh-grade students in 2007–08, had still not passed the mathematics subtest of CAHSEE as of May 2009, were otherwise on track to graduate, and had mastery of CAHSEE content in mathematics.

Although the number of assessed students found to have mathematics mastery was too small to allow for the development of a model-based population estimate of mastery, we can still make use of the model-based estimate of on-track status. Of the 43 students assessed for mathematics, three of them had no CAHSEE attempts in 2008-2009 and thus were not in the larger population of 4,109 students who still had not passed the mathematics subtest as of May 2009. These students were omitted from this further analysis, leaving 40 students, two of whom (5 percent) were evaluated as having mathematics mastery (s.e. 3.49).

Examining the imputed on-track values for the 4,109 students, we determined that 52.8 percent of them were estimated to be on track to graduate (s.e. 1.8); this corresponds to 2,171 students (95 percent confidence interval of 2,026 to 2,316). If we apply the observed mathematics mastery percentage of 5 percent, and its associated standard error, to the 2,171 figure, we get an estimate of 109 students (95 percent confidence interval of 0 to 257). However, taking into account that there is also some uncertainty around the estimate of 2,171 students on track, a more

²⁴ One other student was assessed, but did not complete enough of the protocol to be evaluated.

conservative estimate of the upper end of the range would be 274. The lower end of the range, meanwhile, can be adjusted upward to two students (that is, the two observed).

English Language Arts

The results in ELA are somewhat different. After excluding students who subsequently passed the CAHSEE ELA subtest, there were 47 students with analyzable data from the ELA assessment.²⁵ Fifteen of these students, or almost 32 percent, were evaluated as having mastery of CAHSEE content in ELA.

Among the pool of 5,507 potential students (who were failing one or both subtests of the CAHSEE), 3,919 had still not passed the ELA subtest as of May 2009. If we assume that 57 were on track to graduate, and 32 percent of the on-track students had mastery of CAHSEE content in ELA, then we estimate that there are 715 students who took CAHSEE with modifications and/or accommodations as eleventh grade students in 2007–08, had still not passed the ELA subtest of CAHSEE as of May 2009, were otherwise on track to graduate, and had mastery of CAHSEE content in ELA.

Of the 47 students remaining in the in-depth study sample, one of them had no CAHSEE attempts in 2008-2009 and thus was not in the larger population of 3,919 students who still had not passed the ELA subtest as of May 2009. Thus, this student was omitted from the model-based analysis, leaving 46 students, 15 of whom were evaluated as having ELA mastery. We used the 46 cases to predict the odds of having such mastery. Predictors in the model included various known information on the students such as demographic characteristics and how close they were to passing the ELA subtest of CAHSEE on their most recent operational attempt. We then used this model to impute multiple possible “mastery” values (mastery versus non-mastery) for each of the 3,919 students in the population.

However, the simple model-based estimate of how many of the 3,919 students had ELA mastery would not take into account how many of those students were also on track to graduate. We approached this problem in two different ways. First, we matched students’ imputed ELA mastery values to their imputed on-track values (first set of on-track values paired with first set of ELA mastery values, second set paired with second set, and so on), and then estimated the percentage of students with ELA mastery *only within the subset of students predicted to be on track to graduate*. The number of the 3,919 students predicted to be on track to graduate was 1,981, and the percentage of these students predicted to have ELA mastery was estimated at 35.9 percent (s.e. 4.03). This is equivalent to 711 students, with a 95 percent confidence interval of 555 to 868.

This method, however, does not take into account the uncertainty around the percentage of students estimated to be on track to graduate. To address this, we used an alternate method in which we looked at the imputed on-track values for the 3,919 students and determined that 51.1

²⁵ Two other students were assessed, but did not complete enough of the protocol to be evaluated.

percent of these students were estimated to be on track (s.e. 0.49); this corresponds to 2004 students, with a 95 percent confidence interval of 1,967 to 2,042. Then, we applied the estimated mastery percentage of 35.9 percent, and its associated standard error, to each end of this confidence interval, to get *two* confidence intervals of mastery (one applying to the estimate of 1,967 students on track and the other applying to the estimate of 2,042 students on track). The low end of the first interval is 551 students, while the high end of the second interval is 894 students.

Combining across these various methods, it seems that a safe (conservative) estimate of the number of students in our population of interest who were *both* on track to graduate *and* had mastery of ELA CAHSEE content would be 550 to 900 students.

Analysis of Special Education Coordinator and Teacher Interviews

Available interview data from special education coordinators and teachers were tabulated for the students who remained in our in-depth study sample after excluding those who were found to have passed the actual CAHSEE between December 2008 and May 2009. Among the 92 students remaining in our sample, 49 had coordinator/teacher interview data. Table 7 shows how these interview data are distributed across mathematics and ELA and across students who were judged to have exhibited mastery and those who were not.

Table 7. Availability of coordinator/teacher interview data for students participating in individual assessments, by students' subject area and evaluated mastery

Evaluated mastery	ELA	Math	Total
Students failing evaluation of mastery	15	23	38
Students passing evaluation of mastery	9	2	11
TOTAL	24	25	49

The coordinators/teachers were asked about the educational setting in which the students were receiving instruction for the subject area in which we were evaluating mastery. Although we asked about educational setting for all four years of high school, the best information was available for grade twelve, where we had non-missing data for 45 students. As shown in table 8, the majority of these 45 students were reported to be in special education or CAHSEE preparation classes only (33 in special education/CAHSEE prep versus ten in general education and two not currently taking classes).

Table 8. Primary twelfth grade educational setting, for students participating in individual assessments, by students' subject area and evaluated mastery

Evaluated mastery	ELA			Math			Total
	Gen ¹	S.E. ²	None	Gen ¹	S.E. ²	None	
Students failing evaluation of mastery	2	13	0	5	13	2	35
Students passing evaluation of mastery	3	5	0	0	2	0	10
TOTAL	5	18	0	5	15	2	45

¹General education classes

²S.E. includes special education classes and CAHSEE preparation classes as primary educational setting. Some students were enrolled in a CAHSEE preparation class in addition to their primary regular or special education class. Note: Within the set of 49 students with coordinator/teacher data, data on educational setting were missing for one student passing mastery in ELA and three students failing mastery in mathematics.

We also asked about the grade level of instruction and the teacher credentials for the special education/CAHSEE preparation classes. Respondents reported that students in these classes varied widely in ability, so teachers had to be able to differentiate their teaching strategies to try to reach each student at his or her level, while at the same time teaching the required grade-level standards. This variation also made it difficult to determine the exact standards and level of material to which each student was exposed.

With regard to teacher credentials, data were available on 27 of the 33 students reported to have special education or CAHSEE preparation as their primary educational setting (table 9). The majority of these teachers had special education credentials, although single subject credentials were relatively more frequent among the teachers of students who were evaluated as having mastery of CAHSEE content.

Table 9. Credentials of special education/CAHSEE preparation teachers, for students participating in individual assessments, by students' subject area and evaluated mastery

Evaluated mastery	ELA			Math			Total
	S.E. Cred ¹	S.S. Cred ²	None	S.E. Cred ¹	S.S. Cred ²	None	
Students failing evaluation of mastery	9	2	0	7	1	1	20
Students passing evaluation of mastery	3	2	0	1	1	0	7
TOTAL	12	4	0	8	2	1	27

¹Special education credential

²Single subject credential

Note: Within the set of 33 students reported to have special education or CAHSEE preparation classes as a primary education setting, data on teacher credentials were missing for two students failing mastery in ELA and four students failing mastery in mathematics.

Respondents reported that some of the supports given to SWDs in general education classes included:

- CAHSEE-specific curriculum or supports, including texts or computer programs such as *Measuring Up*, *Connecting Math Concepts*, *CAHSEE Revolution*, and *Study Island*
- Extra help from adults (for example, teaching assistants or parents)
- Audio tapes, spell check, and grammar check

- Graphic organizers
- Extended time on tasks
- Read-aloud on tests
- Calculators

In both general and special education classes, respondents reported providing the following CAHSEE-specific supports:

- Providing CAHSEE study time and materials such as *CAHSEE Revolution*, *Measuring Up*, Kaplan study guides, and *Study Island* in study halls/free periods
- Focusing on CAHSEE-covered standards in regular mathematics or ELA classes
- Offering after-school, summer, or Saturday small group CAHSEE tutoring sessions
- Enrolling students not passing CAHSEE in CAHSEE-specific or remedial classes
- Making students aware of and teaching them how to use calculators and dictionaries before the test
- Teaching test-taking skills
- Aligning resource and special day class curricula to state standards
- Allowing the regular classroom teacher to administer the test

Respondents also reported that the following accommodations and modifications seemed particularly helpful for students:

- Extended time, including four-day testing and frequent breaks
- Audio CDs (read-aloud) in both ELA and mathematics
- Individual test-taking or small group testing
- Writing the ELA essay on a computer
- Speech-to-text converter or dictation for the essay

Respondents suggested enrolling students in CAHSEE prep classes during the day instead of or as well as after school, on Saturdays, or over the summer to increase student attendance. Two respondents commented that they felt students in special day classes should not be required to take the CAHSEE, but that instead California should create a differentiated diploma for these students that acknowledged that the students were unable to pass the exam but that would still allow the students to apply for financial aid or join the military. One teacher commented that schools need to hire more subject-credentialed special education teachers. Finally, respondents were split on whether they thought having the CAHSEE requirement for SWDs helped motivate schools and teachers to hold these students to higher standards or discouraged students who have difficulty with the test from attempting to get a diploma.

Task 2 Conclusions and Discussion

The primary question of task 2 was whether there were students who met the criteria of the study—had failed to achieve a passing score on one or both subtests of the CAHSEE despite the use of allowable modifications or accommodations but were otherwise on track for graduation—who had mastery of CAHSEE content. In addition, if such students were found, we wished to estimate their prevalence in the population.

Students were evaluated for mastery using in-depth individual assessments that were constructed out of questions drawn from the pool of released CAHSEE items and enhanced with probes based on cognitive interview techniques. Based on the students' responses, we determined that there were, in fact, eligible students with mastery of CAHSEE content. Two of 43 students in the final mathematics sample (about 5 percent) and 15 of 47 students in the final ELA sample (about 32 percent) were judged to have mastery.

Estimating prevalence in the population was somewhat more difficult for several reasons, including the following:

- We did not have firm population figures for the numbers of students who met the qualifying condition of being otherwise on track to graduate.
- Neither our assessed sample nor our potential sample (which was used to estimate numbers on track to graduate) was fully representative of all California students since, for logistic reasons, we only sampled from districts with 50 or more study-eligible students.²⁶
- The numbers of assessed students who met the criteria for mastery were too small—especially in mathematics—to allow confident model-based estimates of prevalence.

Given these caveats, we nevertheless offer the following estimates of the numbers of eligible students in the 2007–08 eleventh grade cohort who had established mastery of CAHSEE content by the end of their senior year:

- In mathematics, we estimate that the number of eligible students with mastery is greater than zero but less than 300.
- In ELA, we estimate that the number is between 550 and 900 students.

There were two additional research questions that we also attempted to answer in task 2: these concerned the opportunities to learn CAHSEE content that were afforded students who were

²⁶ As a reminder, study-eligible refers to students who meet the first criterion for inclusion in task 2, namely, failing to achieve a passing score despite the use of allowable modifications or accommodations.

found to be failing in mastery and, among those who were judged to *have* mastery, the types of disability-related factors that appear to have created barriers to passing the regular CAHSEE examination.

With regard to the opportunity to learn question, we attempted to gather information on the students' educational context from their teachers and special education coordinators. We were only successful in obtaining this information for approximately half the sample, and the information that we did obtain was less complete than we would have wished. Nevertheless, it was apparent that the majority of students in our sample were receiving their primary mathematics or ELA instruction in special education or CAHSEE preparation classes rather than general education classes, that the grade level of material covered in these classes was ambiguous because of the wide range of students served in the same class, and that the majority of the teachers had special education rather than single subject credentials. Combined with comments made by students in the cognitive interviews to the effect that "I was never taught to do this," we conclude that many of the students who we evaluated as failing had not, in fact, had meaningful opportunities to learn CAHSEE content. To what extent their lack of exposure was a result—rather than a cause—of their generally low level of performance is beyond the scope of this study to address.

Finally, disability-related barriers to passing the regular CAHSEE are difficult to sort out since the students who passed under the cognitive interview conditions were provided with a package of adaptations designed to compensate for a range of disability-related conditions such as low reading skills, attention deficits, and others. Our findings from both tasks 2 and 3, however, suggest to us that low reading skills were a major contributing factor to the problems these students experienced with both subtests of the regular CAHSEE.

In support of our findings, a pilot study of item modifications for the California Modified Assessment conducted by Educational Testing Services (ETS) found that, "There is strong evidence that listening to the passage and stem of a question made the item more accessible than when these components were read by the test taker. There is evidence that listening to the stem of a question influenced the accessibility of math and science items in addition to ELA items," (ETS, 2007). This appears reasonable in light of the fact that the majority of students in our sample had a specific learning disability as their primary disability, and dyslexia is very common in this population. We did not, however, do any analyses to untangle whether or not the students with low reading skills who were able to display mastery in the cognitive interview setting had, in fact, taken advantage of the read-aloud accommodation/modification in their regular CAHSEE attempts. Knowing this would be more instructive as to the contribution of low reading skills versus other disability-related factors.

Task 3: Identify and Analyze Possible Alternative Means

Purpose and Research Questions

Having concluded, on the basis of task 2, that some students have mastered CAHSEE content in ELA and/or mathematics but are unable to demonstrate that knowledge through the CAHSEE despite their use of permissible modifications and/or accommodations, we proceeded to task 3. In task 3 we considered whether this subgroup of qualified students could demonstrate their knowledge of the CAHSEE standards through alternative means, and we conducted a small scale pilot test of one alternative, described in our proposal, that appeared to hold promise based on other research being carried out by AIR.

More generally, we attempted to answer the following research questions concerning alternative means:

1. What alternative means can be developed and used by the CDE that would allow qualified students to demonstrate their knowledge of the material assessed on the CAHSEE?
2. How would these alternative means allow for the setting of a passing standard equivalent to that required of students taking the CAHSEE?
3. Which alternative means do we recommend for use by the CDE?

In this section we discuss the conceptual framework, population and sample, methods and procedures, analyses, and conclusions associated with our development and pilot testing of one specific alternative based on a group administered version of the CAHSEE with item and form modifications intended to overcome barriers caused by students' disabilities. Because the alternative is closely based on the standard CAHSEE, it is easily amenable to setting a passing standard equivalent to that required of other students. However, evidence is needed to demonstrate that the adaptations can make a significant impact on the performance of qualified SWDs, and this is why we proposed to carry out a pilot test as part of the independent evaluation study.

A subsequent section of this report—Discussion and Recommendations—considers a broader range of alternative means that the State of California might adopt to better serve qualified SWDs. As part of this discussion, we present evidence from this study (tasks 2 and 3), and evidence from the experience of other states that have adopted alternative means, in order to address the efficacy, feasibility, and validity of the various alternatives.

Conceptual Framework

Our general approach to developing an alternative means of assessment incorporates item, form, and administrative adaptations designed to maintain alignment with the content standards for the CAHSEE and remove the challenges created by deficits in specific cognitive traits or abilities. The approach builds on results from a study funded by the U.S. Office of Special Education Programs (OSEP) that is being conducted by AIR with the states of Ohio, Minnesota, and Oregon to examine alternatives to assessments, including alternatives to secondary school assessments that are used by those states to meet requirements of the Elementary and Secondary Education Act.

Three basic premises underlie this work, which has progressed to the point of large-scale pilot testing in Ohio, Minnesota, and Oregon:

1. Deficits in specific cognitive traits or abilities have an impact on the ability of a student to access and perform assessment tasks in specific and identifiable ways.
2. Ideal test adaptations remove disability-imposed barriers to performance rather than simply make the test easier.
3. An appropriately adapted assessment can accurately and fairly measure proficiency on grade-level content among SWDs who have persistently performed poorly on an unadapted assessment.

Cognitive and neurological research has identified distinct, measurable traits and abilities (e.g., Royall et al., 2002). Deficits in some of these traits and abilities are often associated with, or used in the diagnosis of, various cognitive disabilities (e.g., Godefroy and Rousseaux, 1996; Seidman, et al., 1997; Williams, Goldstein, and Minshew, 2006). These deficits can result in specific difficulties for students attempting to perform tasks routinely required by conventional tests.

For example, students may have specific deficits in working memory, planning, or focused attention. Each of these can be ameliorated by adaptations such as:

- Providing examinees with graphic or other organizing devices to annotate as they work through problems (addresses deficits in working memory)
- Priming the student's memory in order to help him or her more readily see the solution to some problems (addresses deficits in planning)

- Visually highlighting important words in stimuli and questions to draw the attention of the students (addresses deficits in focused attention)

In addition, our approach incorporates the general principles of universal design for learning (UDL), including reducing the language load to the minimum necessary to communicate the ideas in the item, avoiding unnecessarily unfamiliar contexts, and making graphics clear and uncluttered.

Population and Sample

Given the resources available to the study, our goal in task 3 was to conduct a small scale pilot to provide preliminary evidence on the efficacy of one promising form of alternative means. As such, we did not require a representative sample that would allow us to develop population-based parameters, but we did require a sample of students who met the eligibility criteria stipulated by the independent evaluation, criteria which also applied to task 2. That is, the students needed to have taken the CAHSEE with allowable modifications and/or accommodations but not yet passed it, and they needed to have met or be expected to meet all other graduation requirements. Our approach was to recruit a sample of eligible students in eleventh and twelfth grades, as identified by personnel at local districts and schools that agreed to participate.

Students from 10 high schools in 2 districts in northern California participated in task 3. District- and school-level staff members were asked to identify 10 to 20 students in each school that met the study criteria. Parent and student permission was solicited for all participants. We then tested all consenting students except where class schedules or student absences made this impractical.

In total, we tested 103 students, including 46 in ELA and 57 in mathematics. The demographic and disability distributions of the students are shown in tables 10 through 14. Sixty-two percent of the students were in eleventh grade, with the remainder in twelfth grade. Sixty-seven percent were Hispanic or Latino, 46 percent were English learners, and 67 percent were male. As in the task 2 sample, the most common disability category was “specific learning disability.”

Table 10: Pilot test of alternative means: distribution of tested students by grade

Grade	Tested students	
	N	Percent
Eleventh	64	62.1
Twelfth	39	37.9
TOTAL	103	100.0

Source: Local education agency (LEA) and school records.

Table 11. Pilot test of alternative means: distribution of tested students by race/ethnicity

Race/ethnicity	Tested students	
	N	Percent
American Indian or Alaskan Native	0	0.0
African American or Black	8	7.8
Asian	13	12.6
Hispanic or Latino	69	67.0
Pacific Islander	1	1.0
White (not of Hispanic origin)	12	11.7
Declined to state/Multiple response	0	0.0
TOTAL	103	100.0

Source: Local education agency (LEA) and school records.

Table 12. Pilot test of alternative means: distribution of tested students by English proficiency level

English proficiency level	Tested students	
	N	Percent
English only	51	49.5
Initially fluent English proficient	3	2.9
English learner	46	44.7
Reclassified fluent English proficient	3	2.9
TOTAL	103	100.0

Source: Local education agency (LEA) and school records.

Table 13. Pilot test of alternative means: distribution of tested students by gender

Gender	Tested students	
	N	Percent
Male	69	67.0
Female	34	33.0
TOTAL	103	100.0

Source: Local education agency (LEA) and school records.

Table 14. Pilot test of alternative means: distribution of tested students by primary disability

Primary disability ¹	Tested students	
	N	Percent
No Special Education Services	0	0.0
Mental Retardation	1	1.0
Hard of Hearing	0	0.0
Deafness	1	1.0
Speech or Language Impairment	13	12.6
Visual Impairment	0	0.0
Emotional Disturbance	3	2.9
Orthopedic Impairment	1	1.0
Other Health Impairment	2	1.9
Specific Learning Disability	80	77.7
Multiple Disabilities	1	1.0
Autism	1	1.0
Traumatic Brain Injury	0	0.0
TOTAL	103	100.0

Source: Local education agency (LEA) and school records.

Although both the task 2 and task 3 students met the eligibility criteria specified by the independent evaluation, there are reasons to think that the two populations are not strictly comparable. The primary reason to suspect non-comparability is that the students in task 3 were tested earlier in their high school careers than the students in task 2. Not only did we open the task 3 sample to students in eleventh grade, but the task 3 testing also occurred nearly six months earlier in the school year—early December for task 3 versus May or June for task 2—so even the twelfth grade students in task 3 were not so far along in their schooling.

Because of this difference in timing, there are competing reasons to think that the task 3 sample may be more or less able than the task 2 sample. Arguing for a more able sample is the fact that students participating in task 3 had had fewer opportunities to attempt the CAHSEE at the point at which we tested them. Based on the patterns observed in the 2007–08 eleventh grade cohort that formed the basis for our sampling in task 2, we would expect significant numbers of these students to eventually pass the CAHSEE.²⁷ On the other hand, the earlier test dates also mean that students in the task 3 sample had had less opportunity for CAHSEE-specific remediation, and this may have had a lowering effect on their ability to master CAHSEE content.

Another reason to think that the samples may be non-comparable is that, for task 2, we were able to verify students' CAHSEE scores and their use of modifications and/or accommodations by referencing CAHSEE administrative records. For task 3, we relied on district and school personnel to screen students against these criteria. (For both samples, we relied on district and

²⁷ That is, this would be the expectation if one were to set aside potential influences arising from the CAHSEE waiver that is currently in effect for SWDs.

school personnel to confirm “on track to graduate” since there was no centralized source for this information.)

Methods and Procedures

To determine whether item and form adaptations could have an effect on student performance on the CAHSEE, adapted versions of CAHSEE items were constructed using the principles described above under the conceptual framework. The adapted items and the source items were then assembled into forms such that each student received equal numbers of adapted and unadapted items in the same subject area. This design is illustrated in figure 2.

Figure 2. Design for task 3

	Mathematics	ELA
Group 1	Item set A (20 items) – adapted Item set B (20 items) – unadapted	
Group 2	Item set A (20 items) – unadapted Item set B (20 items) – adapted	
Group 3		Item set C (19 items) – adapted Item set D (19 items) – unadapted 1 essay, either adapted or unadapted
Group 4		Item set C (19 items) – unadapted Item set D (19 items) – adapted 1 essay, either adapted or unadapted

With this design, student subscores could be calculated using just adapted or just unadapted items, and performance could be compared within subjects, between the two conditions. The one exception involves the essay that forms part of the ELA score; because of time constraints, students were only asked to respond to a single essay, either adapted or not.

Developing the Test Forms

Like the test forms used in the task 2 individual assessments, the task 3 test forms were built from released CAHSEE items. In fact, many of the same items were used in both tasks; this allowed us to take advantage of the information gained in the individual assessments about what aspects of particular items students found difficult. Additional items were selected to round out the task 3 forms, which needed to be longer in order to support the calculation of subscores.²⁸

The items in each subject were chosen to reflect the content strand and b-value distribution on the actual CAHSEE, as shown in figures 3 and 4 below.

²⁸ In addition, nine ELA items used in task 2 were replaced in task 3 due to a variety of reasons. The primary reasons for the substitutions, however, arose from the constraints of building two equivalent half forms with appropriate numbers of items in each content strand.

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Figure 3. Distribution of items by content strand — operational CAHSEE and task 3 test forms

Strand	Number of items in CAHSEE	Number of items in task 3 test forms
Mathematics		
Number Sense	14	7
Statistics, Data Analysis, and Probability	12	6
Algebra and Functions	17	9
Measurement and Geometry	17	8
Mathematical Reasoning	8	4
Algebra	12	6
TOTAL MATHEMATICS	80 multiple choice	40 multiple choice
English Language Arts		
Word Analysis ¹	7	4
Reading Comprehension	18	10
Literary Response and Analysis	20	10
Writing Strategies	12	6
Writing Conventions	15	8
Writing Applications	1	1
TOTAL ELA	72 multiple choice + 1 essay	38 multiple choice + 1 essay

**Figure 4: Distribution of items by equated Rasch item difficulty values (b-values)—
operational CAHSEE and task 3 test forms**

Low b-value	High b-value	Number of items in CAHSEE	Number of items in task 3 test forms
Mathematics			
-2.0	-1.5	0-1	1
-1.5	-1.0	3-6	3
-1.0	-0.5	14-20	12
-5.0	0.0	16-23	10
0.0	0.5	16-23	8
0.5	1.0	9-15	6
1.0	1.5	2-5	0
TOTAL MATHEMATICS		80 items	40 items
English Language Arts			
-2.0	-1.5	0-1	0
-1.5	-1.0	3-6	3
-1.0	-0.5	11-15	7
-5.0	0.0	19-27	9
0.0	0.5	16-25	12
0.5	1.0	9-15	6
1.0	1.5	3-8	2
1.5	2.0	0-2	0
TOTAL ELA		73 items	39 items

Once the source items were selected, experienced AIR item developers constructed adapted versions of all selected items using the principles described above under the conceptual framework. The work was further informed by the results of pilot tests from the OSEP-supported grant also described above. As noted, the adapted items and the source items were then assembled into test forms in which half of the items were adapted and half were not.

Examples of source item/adapted item pairs are given in appendix C. These example items are discussed below in the context of the data analysis for task 3. For the ELA essays, the adaptation involved adding a pre-writing exercise that made use of a graphic organizer.

Test Administration

Students were given test booklets similar to those provided in task 2 and asked to record their answers on a Scantron-style answer sheet.²⁹ Compared with regular CAHSEE test books, the adapted test books had more white space and items were not placed in double columns. Students were told they could write in their test books or on scratch paper, take breaks, and ask questions about test directions. Students received the individual accommodations (but not modifications) that they would normally receive when taking the CAHSEE, including breaks, extended time, or a read-aloud on the mathematics test.

Data Analysis

Because we used CAHSEE items and followed the specifications used for the CAHSEE, we believe that this assessment maintains the same high quality as any other CAHSEE. The primary consideration in this analysis, therefore, is whether the adaptations adequately eliminated the effect of students' disabilities.

We considered various types of evidence consistent with the following hypotheses:

- **Access is improved.** If the adaptations are effective, we would expect to see fewer omitted items among the adapted items than among the unadapted items.
- **Performance exceeds chance.** The targeted students should not only respond to the adapted items, but they should respond meaningfully. That is, they should have a reasonable probability of getting the adapted items correct—at minimum, above chance.
- **Performance improves.** If the adaptations are effective, we would expect to see a boost in performance of the targeted students, both in terms of percentage of items answered correctly and in terms of numbers of students achieving passing scores.

²⁹Although the test book design was similar in the two tasks, the mode of employment was different. In task 2, students read along in their test books as the administrator read the material out loud. Students then responded orally to the questions. In task 3, by contrast, students read the test items to themselves (unless they requested a read-aloud in mathematics only) and marked their answers on a separate answer sheet.

Of course, since not all eligible students have mastery of CAHSEE content, we did not expect that all of the students in the sample would be able to pass the CAHSEE, even if disability-related barriers to performance were removed. The results from task 2 offer some basis for estimating the numbers of students who might achieve passing scores with appropriately adapted instruments. Unfortunately, however, since the task 3 population differed from the task 2 population (as described above) we could only make limited use of the population estimates from task 2 to shape our expectations for task 3.

Mathematics

The results for mathematics were reasonably promising. The comparison of omit rates could not be done because very few students omitted either adapted or unadapted items. However the other comparisons all favored the adapted items.

- Twice as many students (18 students versus 9 students) had scores at chance or below on the unadapted items versus the adapted items.³⁰
- Eleven students obtained passing scores (55 percent or more items answered correctly) on the adapted items, as opposed to seven students on the unadapted items. This is an incremental increase of four out of 50 students, or 8 percent—not too dissimilar from the 5 percent of eligible students judged to have mastery of CAHSEE content in mathematics in task 2.
- Average scores on the two subtests were significantly different based on a repeated-measures t-test comparing subtest scores within student. ($\Pr(T > t) = 0.0009$)

At the item level, 8 of the 40 mathematics items used in task 3 showed p-value increases of 0.2 or greater when adapted. These eight items are reproduced in appendix C in both their adapted and unadapted form. For several of these items, the adaptations follow principles of UDL and would be equally appropriate for use in the general test population as for use with SWDs.

ELA

The results for ELA were considerably less promising. Once again, the comparison of omit rates could not be done because very few students omitted either adapted or unadapted items. None of the other comparisons showed an advantage for the adapted items.

Based on the multiple choice items:

- More students (11 students versus 6 students) had scores at chance or below on the *adapted* items compared to the unadapted items.

³⁰ Since CAHSEE items are four-option multiple choice, chance was defined as responding correctly to 25 percent or fewer of the items.

- Four students obtained passing scores (60 percent or more items answered correctly) on the adapted items, versus five students on the unadapted items.
- Average scores on the two subtests were not significantly different based on a repeated-measures t-test comparing subtest scores within student ($\Pr(T > t) = 0.8070$).

Based on the essay scores:

- Average scores between adapted and unadapted versions of the essays were not significantly different based on a two-sample t-test ($\Pr(T > t) = 0.1477$).

Task 3 Conclusions and Discussion

In task 3 we developed and pilot tested one type of alternative means—a test form composed of regular CAHSEE items that had been adapted to remove the challenges created by deficits in specific cognitive traits or abilities while still maintaining alignment with the content and performance standards of the regular CAHSEE. Unlike our approach in task 2, we chose to employ a group administration and to allow accommodations but not modifications. An alternative means with these attributes—if successful in ameliorating the challenges faced by SWDs—would have the advantage of being less expensive to administer than an individual assessment and more closely conforming to CAHSEE standards than a test that employed modifications.

In mathematics, we achieved a boost in performance that was consistent with our expectations from task 2 concerning the percentage of eligible students who demonstrated mastery of CAHSEE content. In ELA, on the other hand, we did not achieve a significant increase in either average scores or percent passing, despite the fact that our task 2 results led us to believe that higher percentages of eligible students had mastery of CAHSEE content in ELA than in mathematics.

Further discussion of the alternatives means tried out in tasks 2 and 3, as well as discussion of other alternative means used by other states with high school exit examinations, is provided in the concluding section of this report.

Public Comment

As required by the contract, public comment was solicited at two public meetings—one in northern California and one in southern California. The meetings were held at the Orange County Office of Education on April 9, 2009 and at the Sacramento County Office of Education on April 30, 2009. Interested parties were also invited to submit their comments in writing.

No written comments were received. However, 33 people attended one or the other of the meetings, either in person or by phone hook-up. These included school district personnel, parents of SWDs, and one representative from each of the following: an education technical assistance contractor, a legal firm, the Attorney General’s office, the California Teachers Association, and two advocacy groups for persons with disabilities.

After hearing a presentation by the project manager on the study methodology and asking a number of clarifying questions, participants made the following recommendations for the study:

- Take account of students’ social, medical, and educational circumstances in interpreting the results of the student assessments
- Evaluate the students’ opportunities to learn CAHSEE content
- Include a socioeconomically diverse sample of districts in the study. Be sure that the population studied is not overly narrowed by selecting the sample of districts from among those with at least 50 study-eligible cases
- Consider a variety of alternative means, not just tests
- Make stronger efforts to involve parents in the study

Discussion and Recommendations

This section includes a discussion of the two alternatives that were implemented and evaluated during this study: an individual assessment with supporting probes and a written test employing item and form adaptations. We also include discussion of the exit examination alternatives reviewed by AIR, through its partnership in the California Comprehensive Center (CA CC), in a report for the Assembly Bill 2040 Panel; namely: task specific assessment, concordant scores, and collection of evidence. For each of the options we address the degree to which that alternative means can validly assess whether or not students have achieved the same level of academic achievement in the content standards as is required for passage of the CAHSEE. We also discuss the feasibility of each option.

Discussion of Alternative Means Evaluated in this Study

Individual assessment with supporting probes

As implemented for this study, the individual assessment includes a one-on-one, oral administration of CAHSEE items. Each item is presented and read aloud to the student and the student responds to the item. Even if the student answers correctly, a series of scripted probes, based on cognitive interview techniques, are administered to each student. Examples of items together with the probes are provided in appendices A and B. The intent of the probes is to gather additional information from students who answered the question incorrectly that will enable content and special education experts to determine whether the student has the knowledge and skills demanded by the item. One clear advantage of this approach is that it uses CAHSEE items and also permits use of the CAHSEE proficiency standard.

The individual assessments also incorporated a set of item and form adaptations that are discussed more fully below.³¹ We believe that, with further research and development, this combined approach can provide a defensible alternative for assessing students whose disabilities prevent them from demonstrating what they know, while preserving the same high standards as the CAHSEE demands of other students. An important logistical consideration, however, is that approach requires individual administration by a trained examiner and may also require an adjudication process for evaluating student responses to probes to ensure that students' responses are evaluated appropriately.

Adapted Assessments

We use the term "adapted" assessment rather than the terms "accommodated" or "modified." Typically "accommodation" refers to form or item changes provided for an individual student and "modifications" are individual form or item changes that invalidate an assessment. Adapted assessments, as used in this report, are assessments that include item or form revisions to

³¹ Other than general format changes applied to the written forms that students had in front of them during the individual assessments, the adaptations were applied to the item probes only. This meant that students always attempted an item in its unadapted form first. The distinction was useful for research purposes, but would not need to be maintained in operational use.

enhance the accessibility of the assessment, that are provided for all SWDs, and that could also be included in assessments of all students since they conform to the principles of universal design. (We did not, however, include students without disabilities in this study.)

Before implementing adapted assessments, it would be important to conduct additional research to ensure that only construct irrelevant variance is eliminated through the use of adaptations. Normally, adaptations to assessments are evaluated by examining whether the adapted items are easier for non-disabled students. If they are easier, then there is concern that construct relevant variance was altered with the adaptation. Mathematics adaptations used included bolding of important features, simplified language/numbers, graphic revisions, graphic organizers, and scaffolding within a question. In ELA the adaptations included simplified language, bolding, boxing (also known as “embedding”), and vocabulary redundancy. Examples of mathematics adaptations are provided in appendix C, where we present the subset of items that exhibited the greatest gains in p-values after adaptation.

In this study, we found no overall effects for these adaptations in ELA and modest effects in mathematics, which may nevertheless be consistent with expectations about the proportions of students in the eligible population who actually have mastery of CAHSEE content in mathematics. Because of the exploratory nature of this study, we suggest additional research examining adaptations in both mathematics and ELA. Furthermore, since some of these adaptations appeared to produce a greater impact when they were used in the context of the individual assessments than when they were administered in a written test, we would suggest that additional research examine these interactions with test format more closely.

The study of adapted assessments did not include any use of modifications. For example, none of the students had any portion of the ELA examination read to them. The individual assessment protocols did include reading the ELA assessment to students and providing a calculator for mathematics, although examiners reported that the calculator did not seem helpful for most students. Based on the results of this study, we speculate that adaptations for the ELA assessment may only be effective when the assessment is read to students. For many of these students, their word reading fluency may be so limited that no item or form adaptations could have any effect on improving reading comprehension.³²

Alternative means based on item and form adaptations is feasible for CAHSEE since current items and proficiency standards can be used. Item and form adaptations would need to be developed and field-tested. The read aloud modification is currently provided for a significant number of SWDs in the state, but if read aloud is permitted in an adapted assessment,

³² On the other hand, it is important to keep in mind that most of the students who are persistently failing the ELA exam will eventually have the read aloud modification for at least one administration of the CAHSEE. For the students in this study, the standard read aloud modification was not, by itself, sufficient to enable these students to pass CAHSEE.

administration costs would expand to include the preparation of specialized audiotapes to match the adapted forms.

Discussion of Other Alternative Means

In this section, we discuss additional options that we did not evaluate in our research. CDE requested that additional alternatives be discussed in our final report. For this, we draw from the presentation prepared for the AB 2040 Panel by AIR through its partnership with the CA CC. That presentation summarizes approaches taken by other states that provide alternative means for students to satisfy their high school exit examination requirement and receive a fully equivalent diploma. Handouts provided at this presentation are included in appendix D.

Task-Specific Assessment

The CA CC presentation identifies two states that use a task-specific assessment approach. In both states, this is an option that is available to both SWDs and those without. The approach uses open-ended performance tasks (projects) that are aligned with state content standards. One of the states, Maryland, uses end-of-course assessments, and since the state has several end-of-course examinations, constructing multiple alternatives is a major undertaking. The other state, New Jersey, has a single comprehensive exit examination, so development costs are less of an issue.

The task-specific option is widely used in New Jersey and there are reports of concerns about overuse. A New Jersey Department of Education study of 105 schools where 10 percent or more seniors used the alternative test reported that only 4 percent of students failed the examination, so there are naturally concerns about whether passing this examination reflects the same level of mastery of content as passing the regular examination (Rothman, 2009). There appears to be less concern with the Maryland alternative, but the *Bridge Plan for Academic Validation: Update for State Board of Education* reported that 68 percent of students passed the alternative examination (Maryland State Department of Education, 2008). Both states require remedial course work before students can take the alternative, so the significant pass rates may be reasonable.

No information was available on any efforts that these states might have undertaken to ensure equivalence in these examinations, particularly in the performance levels used. If the regular assessment and the alternative assessment measure the same content and reflect the same performance levels, it is not clear why the general education students need an alternative. In the case of SWDs, an argument might be made that the alternative is needed because the general assessment does not sufficiently accommodate their disability. For this option and the ones that follow, the major question is, “Do the alternatives reflect the same level of proficiency on the content being assessed?”

Since California has only a single comprehensive exit exam, developing a task-specific approach would be somewhat less challenging than it would be in Maryland. However, it would not be a trivial undertaking to develop, validate, administer, and score a task-specific assessment. New Jersey’s and Maryland’s task-specific assessments are available to all students, so the administrative and development costs are spread across a much larger pool of students.

Concordant Scores

None of the states in the CA CC presentation appear to rely solely on concordant scores, which are scores from another examination that presumably measures similar content. However, several states include results from other tests as part of the evidence that a student may be proficient in a content area. Many assessments, such as Advanced Placement Tests, are reputable assessments whose validity and reliability would not be questioned. A challenge for a state considering the use of concordant scores is ensuring that the assessment used aligns with state content standards and determining what the proficient cut score on the concordant test should be. Since many other assessments that might be used provide limited accommodations for SWDs, it is not clear that this approach actually provides a more accessible alternative than the CAHSEE.

Collection of Evidence

Seven of the states that are discussed in the CA CC presentation use alternative means that include a collection of evidence as an aspect of their approach. For two states, the presentation identified the alternative method as being a waiver, but in both states collection of evidence is part of the waiver approach.

For most of the seven states, a major element of the collection includes a sample of student work completed over the course of a whole school year. At least one of these states also includes a requirement that students produce some on-demand work as part of the collection of evidence. This may provide an opportunity to structure the work sufficiently that one can be confident that the work is that of the student. Ensuring that the work samples reflect only the work of the student is the major threat to the validity of using a body of work as part of the assessment.

As part of the collection of evidence, one state requires certification that the student has an attendance rate of 95 percent and has maintained a “C” average.

The proportion of SWDs in these states who pursue this alternative approach to pass the exit examinations is extremely variable. In Florida, more than 70 percent of students with learning disabilities pursue the alternative approach. Florida reports that 67 percent students passed the alternative in 2002–03. By contrast, in Mississippi, the state received just 11 portfolios to score, and 6 received passing scores. In North Carolina for 2007–08, 32 SWDs pursued the alternative in Algebra I and 91 percent passed. For English I in North Carolina, 44 students took the alternative and 80 percent passed.

Some of these states indicated that there were concerns about their alternative approach and its equivalence with the regular assessment. An article in the Washington Post reported on the use of portfolios in Fairfax County, Virginia where the proficiency rate for SWDs in one school rose from 34 percent to 100 percent once the county began accepting portfolio assessments in place of the regular assessments (Chandler, 2009).

Research conducted by Gearhart and Herman (1995) demonstrates that student performance, as measured by a portfolio, was to a substantial degree a reflection of the amount of teacher support received. The study provides a justification for questioning the degree to which portfolios reflect student competence. Shavelson (2009) goes further by saying that “portfolios are simply not up to the task of providing the necessary data for making sound assessment of student learning.”

While many states have rigorous procedures for evaluating performance on collections of evidence or portfolios, this does not mitigate the challenge of determining that a portfolio reflects only the student’s work. If it does not, then we cannot be sure that it is an accurate reflection of student knowledge. Our view is that assessments must fairly assess students but also ensure accurate reflections of student knowledge. We do not believe that collections of evidence typically provide sufficient protections so that we can be sure that they reflect what a student knows. The on-demand tasks that some states are using as part of the portfolio may be a step that can address this issue provided that they are administered under standard, supervised conditions. The alternative procedures used by Maryland and New Jersey may be useful models for this approach.

Development costs for a collection of evidence approach should not be extraordinary. However, the effort of scoring of collections of work can be substantial, and CDE would need to develop scoring rubrics that align with standards. The biggest challenge that CDE would face is developing strategies to ensure that student work that is reviewed reflects what the student knows and can do.

Recommendations

We provide the following recommendations for the CDE:

1. The individual assessment strategy, combined with item adaptations and the use of follow-up probes for each question, appears to show promise as an alternative means for demonstrating proficiency on the high school exit standards. However, this or any other alternative approach should only be adopted after further validity research that considers the performance of students who are non-disabled as well as SWDs. Furthermore, the individual assessments used in this study only sampled CAHSEE content so as to reduce the administrative burden. If the approach were used to confer credit to individual students, it would be appropriate to administer the full CAHSEE.
2. We believe that alternative means should only be used for students who have failed to meet the requirements for passing the CAHSEE despite repeated attempts and receipt of remedial interventions. While a substantial number of students demonstrated mastery on the ELA examination in this study using the individual assessment strategy, and two students demonstrated mastery on the mathematics examination, there were many other students who clearly had not mastered CAHSEE content. This leads to a conclusion that there are two challenges to improving the passing rate for SWDs on the CAHSEE. The

first is the issue that this study was designed to address, which deals with how these students can fairly demonstrate what they know. The second issue is the more difficult challenge and that is that we are convinced that most of the students in our sample—who by the standards of their schools and districts purportedly had met all conditions for graduation other than passing the CAHSEE—did not know the content covered by the assessment. We do not believe that any valid assessment of these students' content mastery, regardless of the adaptations or methods of assessment used, would result in passing scores for these students. The implications of this finding lead to our next recommendation.

3. Many students assessed in this study did not have mastery of the content assessed by CAHSEE. One implication of this is that curriculum and instruction for these students need to be reviewed and greater attention should be focused on content mastery when students are initially exposed to the CAHSEE content. In addition, students who fail to achieve passing scores on their first CAHSEE attempt should be considered for diagnostic testing and early interventions to address skill deficiencies. In fact, there could be great value in creating a specific set of diagnostic assessments that could be used to target remediation for these students.
4. The CDE should be cautious in adopting alternative means that include a collection of evidence. Unless this option is pursued carefully, it could have the unfortunate effect of enabling students who really do not know the content assessed by CAHSEE to inappropriately meet the requirements. Of course, with this scenario there is no incentive to address the skill deficits of these students and some students may then complete high school with substantial deficiencies in mathematics and ELA.

References

- Almond, Patricia J., and others. 2009. *White Paper: Cognitive Interview Methods in Reading Test Design and Development for Alternate Assessments Based on Modified Academic Achievement Standards (AA-MAS)*. Dover, NH: Measured Progress and Menlo Park, CA: SRI International. <http://www.measuredprogress.org/resources/inclusive/articlespapers/CognitiveInterviewMethods.pdf> (accessed October 15, 2009).
- Independent Evaluation of the California High School Exit Examination (CAHSEE): Fourth Biennial Report*. 2008. Edited by D. E. Becker and Christa Watters. Alexandria, VA: HumRRO.
- Chandler, Michael A. November 19, 2009. "Alternate Exams Questioned as Test Scores Rise in Virginia," *Washington Post*. <http://www.washingtonpost.com/wp-dyn/content/article/2009/11/18/AR2009111801796.html> (accessed October 15, 2009).
- Del Rio-Parent, Lourdes, and Roger Levine. 2000. *A Cognitive Laboratory Study of Accommodations for Limited English Proficient Students*. Palo Alto, CA: American Institutes for Research.
- A Study of Item Formatting and Delivery Mode from the California Modified Assessment (CMA) Pilot Test*. 2007. Prepared by the Educational Testing Service. Sacramento: California Department of Education.
- Gearhart, Maryl, and Joan L. Herman. 1995. "Portfolio Assessment: Whose Work Is It? Issues in the Use of Classroom Assignments for Accountability," *Evaluation Comment*. Los Angeles: University of California, Center for the Study of Evaluation.
- Godefroy, Olivier, and Marc Rousseaux. March 1996. "Divided and Focused Attention in Patients with Lesion of the Prefrontal Cortex," *Brain and Cognition*, Vol. 30, 155–174.
- Jakwerth, Pamela M.; Frances Stancavage; and Ellen Reed. 1999. *An Investigation of Why Students Do Not Respond to Questions*. A publication of the NAEP Validity Studies Panel, Palo Alto, CA: American Institutes for Research.
- Levine, Roger, and Mette Huberman. 2000. *High School Exit Examination: Cognitive Laboratory Testing of Selected Items*. Palo Alto, CA: American Institutes for Research.
- Reed, Ellen; Roger Levine; and Mette Huberman. 2000. *A Cognitive Laboratory Investigation of the Performance of Learning Disabled Students on VNT Items*. Palo Alto, CA: American Institutes for Research.

- Rothman, Carly. June 15, 2009. "N.J.'s Alternative Test to Pass High School Is Subject of Intensive Study," *The Star-Ledger*. http://www.nj.com/news/index.ssf/2009/06/new_jersey_may_change_states_a.html (accessed October 15, 2009).
- Royall, Donald R., and others. November 2002. "Executive Control Function: A Review of its Promise and Challenges for Clinical Research," *Journal of Neuropsychiatry: Clinical Neuroscience*, Vol. 14, No. 4, 377–405.
- Seidman, L., and others. February 1997. "Toward Defining a Neuropsychology of Attention Deficit-Hyperactivity Disorder: Performance of Children and Adolescents from a Large Clinically Referred Sample," *Journal of Consulting and Clinical Psychology*, Vol. 65, No. 1, 150–160.
- Shavelson, Richard J., Stephen Klein and Roger Benjamin. October 16, 2009. "The Limitations of Portfolios" *Inside Higher Ed*. <http://www.insidehighered.com/views/2009/10/16/shavelson> (accessed March 20, 2010).
- Williams, Diane, L.; Gerald Goldstein; and Nancy Minshew. 2006. "The Profile of Memory Function in Children with Autism," *Neuropsychology*, Vol. 20, No. 1, 21–29.
- Zucker, Sasha; Christy Sassman; and Betsy J. Case. 2004. *Cognitive Interviews*. Pearson Inc. https://www.pearsonassessments.com/NR/rdonlyres/E5CD33E6-D234-46F3-885A-9358575372FB/0/CognitiveLabs_Final.pdf (accessed October 15, 2009).

Appendix A: Examples of Assessment Probes in Mathematics

Test Item:

One hundred is multiplied by a number between 0 and 1. The answer has to be—

- A less than 0.
- B between 0 and 50 but not 25.
- C between 0 and 100 but not 50.
- D between 0 and 100.*

Probing Questions:

P1: “Show/tell me how you came up with your answer.”

If the student answered the question incorrectly, look/listen for an explanation that shows that $100 \times 0 = 0$ and $100 \times 1 = 100$. If the student did not provide a satisfactory response, do probe 2.

P2: “Show/tell me what you get when you do 100 times 0.” (write $100 \times 0 =$ on the student’s test book).

If the student answers by saying 0, go onto the next probing question.

P3: “Show/tell me what you get when you do 100 times 1.” (write $100 \times 1 =$ on the student’s test book).

If the student answers by saying 1, go on to the next probing question.

P4: “Can you give me an example of a number that is between 0 and 1?”

Appendix B: Examples of Assessment Probes in English Language Arts

Test Item:

How does Annie’s father feel when Wei and Frances arrive?

- A sad
- B excited*
- C stubborn
- D threatened

Probing Questions:

If the student answers the item correctly, the interviewer asks this question:

“Explain why you chose answer __./Tell me why you chose answer __.”

If the student answers the item incorrectly, the interviewer asks these questions in sequence:

P1: “Let me reread one part of this passage to you.”

Reread paragraph 21.

“What does Annie’s father do when Wei and Frances get to the house?”

P2: “Based on his actions, how does Annie’s father feel when Wei and Frances arrive?”

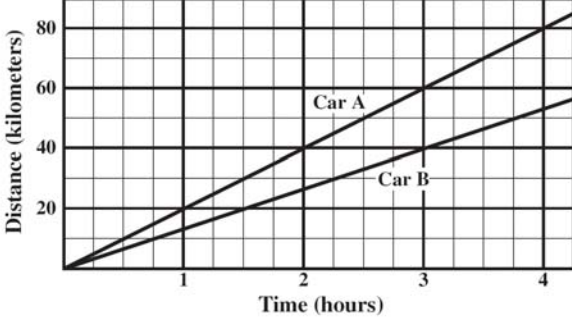
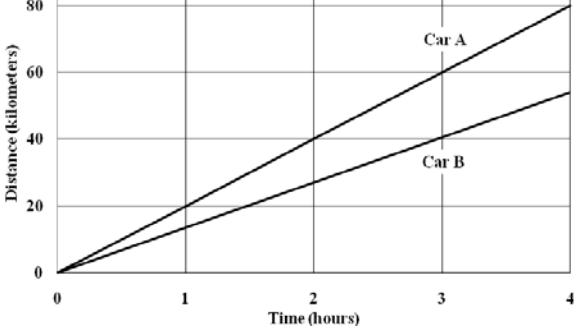
Appendix C: Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (eight items)

Unadapted	Adapted
<p data-bbox="237 474 789 548">Which number has the greatest absolute value?</p> <p data-bbox="237 585 326 617">A -17</p> <p data-bbox="237 655 326 686">B -13</p> <p data-bbox="237 724 326 756">C 15</p> <p data-bbox="237 793 326 825">D 19</p>	<p data-bbox="873 474 1390 548">Which number has the greatest absolute value?</p> <p data-bbox="873 585 979 617">A - 17</p> <p data-bbox="873 655 979 686">B - 13</p> <p data-bbox="873 724 979 756">C 15</p> <p data-bbox="873 793 979 825">D 19</p>
	Adaptation: Bolded “greatest”

Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (continued)

Unadapted	Adapted												
<p>A student asked 50 children to choose between two colors. The results of the survey are shown in the table below.</p> <p style="text-align: center;">Color Survey</p> <table border="1" data-bbox="342 512 652 659"> <thead> <tr> <th>Color</th> <th>Number</th> </tr> </thead> <tbody> <tr> <td>Pink</td> <td>21</td> </tr> <tr> <td>Purple</td> <td>29</td> </tr> </tbody> </table> <p>Based on the data in the table, the student claimed that purple is the favorite color of most of the children. Which reason BEST describes why this is an invalid claim?</p> <p>A Not all of the children chose purple.</p> <p>B More of the children chose pink than purple.</p> <p>C The total number of votes did not equal 50.</p> <p>D The children were only given a choice of two colors.</p>	Color	Number	Pink	21	Purple	29	<p>A student asked 50 children to choose between two colors. The results of the survey are shown in the table below.</p> <p style="text-align: center;">Color Survey</p> <table border="1" data-bbox="987 518 1243 665"> <thead> <tr> <th>Color</th> <th>Number</th> </tr> </thead> <tbody> <tr> <td>Pink</td> <td>21</td> </tr> <tr> <td>Purple</td> <td>29</td> </tr> </tbody> </table> <p>Based on the data in the table, the student claimed that purple is the favorite color of most of the children. Explain why the student should not make this claim.</p> <p>A Not all of the children chose purple.</p> <p>B More of the children chose pink than purple.</p> <p>C The total number of votes did not equal 50.</p> <p>D The children were only given a choice of two colors.</p>	Color	Number	Pink	21	Purple	29
Color	Number												
Pink	21												
Purple	29												
Color	Number												
Pink	21												
Purple	29												
	<p>Adaptation: Adapted stem language and bolded the words “favorite” and “not”</p>												

Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (continued)

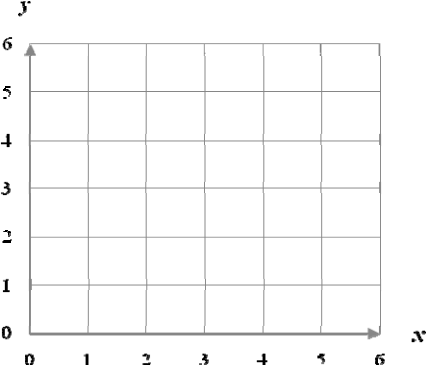
Unadapted	Adapted
 <p>After three hours of travel, Car A is about how many kilometers ahead of Car B?</p> <p>A 2 B 10 C 20 D 25</p>	 <p>After three hours of travel, Car A is about how many kilometers ahead of Car B?</p> <p>A 2 B 10 C 20 D 25</p>
	<p>Adaptation: Deleted extra lines in graph, centered car labels over lines, and bolded the word “ahead”</p>

Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (continued)

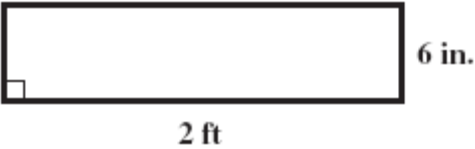
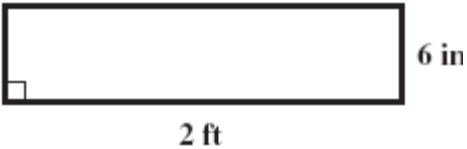
Unadapted	Adapted
<p>Which expression is equivalent to $7a^2b \cdot 7bc^2$?</p> <p>A $14a^2b^2c^2$</p> <p>B $49a^2bc^2$</p> <p>C $49a^2b^2c^2$</p> <p>D $343a^2b^2c^2$</p>	<p>$(7a^2b) \times (7bc^2) =$</p> <p>A $14a^2b^2c^2$</p> <p>B $49a^2bc^2$</p> <p>C $49a^2b^2c^2$</p> <p>D $343a^2b^2c^2$</p>
	<p>Adaptation: Removed words in stem, changed mathematic symbols</p>

Unadapted	Adapted
<p>Marcus plans to buy a Compact Disc (CD) that has a regular price of \$13.99. It is on sale for 10% off, but Marcus will have to pay 7% sales tax. Which is the MOST reasonable estimate of the total cost of the CD including tax?</p> <p>A \$12.50</p> <p>B \$13.50</p> <p>C \$14.50</p> <p>D \$15.50</p>	<p>Marcus plans to buy a Compact Disc (CD) that has a regular price of \$9.99.</p> <ul style="list-style-type: none"> • The CD is on sale for 10% off • Sales tax is 5% <p>Which is the most reasonable estimate of the total cost of the CD including tax?</p> <p>A \$8.50</p> <p>B \$9.50</p> <p>C \$10.50</p> <p>D \$11.50</p>
	<p>Adaptation: Bulleted information, bolded the word “estimate” and simplified numbers.</p>

Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (continued)

Unadapted	Adapted
<p>The points (1, 1), (2, 3), (4, 3), and (5, 1) are the vertices of a polygon. What type of polygon is formed by these points?</p> <p>A Triangle B Trapezoid C Parallelogram D Pentagon</p>	<p>Plot the points (1, 1), (2, 3), (4, 3), and (5, 1) on the grid below. What type of polygon is formed by these points?</p>  <p>A Triangle B Trapezoid C Parallelogram D Pentagon</p>
	<p>Adaptation: Added probe and grid for students to plot points.</p>

Adapted Items with Increases in P-values Greater than 0.2 in Mathematics (continued)

Unadapted	Adapted
<p>The width of the rectangle shown below is 6 inches (in.). The length is 2 feet (ft).</p>  <p>What is the area of the rectangle in square inches?</p> <p>A 12 B 16 C 60 D 14</p>	<p>The width of the rectangle shown below is 6 inches (in.). The length is 2 feet (ft.).</p>  <p>What is the area of the rectangle in square inches?</p> <p>A 12 B 16 C 60 D 144</p>
	<p>Adaptation: Bolded the words “inches”, “feet”, and “square inches”.</p>

Unadapted	Adapted
<p>Sally puts \$200.00 in a bank account. Each year the account earns 8% simple interest. How much interest will be earned in three years?</p> <p>A \$16.00 B \$24.00 C \$48.00 D \$160.00</p>	<p>Sally puts \$100.00 in a bank account. Each year the account earns 8% simple interest. How much interest will be earned in 3 years?</p> <p>A \$8.00 B \$12.00 C \$24.00 D \$80.00</p>
	<p>Adaptation: Simplified number from \$200 to \$100.</p>

Appendix D: Handouts to AB 2040 Panel

Florida: Evidence-Based Waiver	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Complete required courses and obtain 24 course credits, * Maintain a 2.0 GPA, and * Pass Grade 10 FCAT
Exam name & type	Florida Comprehensive Assessment Test (FCAT) (comprehensive)
Tested subjects for graduation	Reading and mathematics
Grade first administered (for graduation purposes)	Grade 10
Number of times students can take exam	Up to five opportunities
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Concordant scores from SAT/ACT (all students) * FCAT waiver (IEP students only)
Alternative described below	FCAT waiver
First instituted	<i>No specific date provided</i>
Eligibility requirements	<p>An eligible student must:</p> <ul style="list-style-type: none"> * Have IEP * Be a senior seeking a standard diploma * Have taken Grade 10 FCAT at least twice, including March of senior year * Have participated in "intensive remediation" in the subject area not passed * Be on track to meet other graduation requirements (as well as any district requirements)
Request process	IEP team decides how student will participate (during fall of student's senior year)
Remediation required?	Yes
Level of administration/ scoring	Local
Administration process and timing	IEP team collects evidence in the student's senior year.
Evidence considered	<ul style="list-style-type: none"> * Other standardized assessments (SAT, CPT, ACT, ASVAB) * Student's performance in core academic courses (classroom work, end of year/semester exams, teacher observations) * Student's performance in higher-level courses (honors, advanced placement) or postsecondary courses through dual enrollment
Scoring process and timing	<ul style="list-style-type: none"> * IEP reviews evidence and makes recommendation * Principal and district superintendent review and verify waiver information prior to the end of the senior school year * Student must still meet other grad requirements
Uniformity in administration / scoring?	While the state provides technical assistance, the districts have the authority to structure their own process for review and approve or deny the IEP's recommendation.
Monitoring	<ul style="list-style-type: none"> * State tracks how many waivers are granted annually * State does not review evidence or monitor process

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Florida: Evidence-Based Waiver	
Associated costs	<ul style="list-style-type: none"> * <i>No specific costs identified</i> * Districts encouraged to support remediation for students with disabilities through IDEA Part B entitlement funds.
Challenges/ concerns	<ul style="list-style-type: none"> * <i>No specific challenges identified</i>
Information on the use and results of alternative means	<ul style="list-style-type: none"> * 1,895 waivers submitted by 65 districts in 2002-03 * 2,593 waivers reported by 65 districts in 2003-04 (37% growth) * 2002-03: 42% of waivers were for both subjects; 39% for reading; 19% for math * Majority of waivers were for students with specific learning disabilities (77% in 02-03 and 73% in 03-04) * 2002-03: 1,271 students <u>received</u> waivers; 1,895 forms were submitted (67%). <p><i>[Internal note: It is unclear from the source if the discrepancy is a data issue or whether these students were not approved at the local level.]</i></p> <p><i>Source: "Students with Disabilities Receiving a Standard Diploma through the FCAT Waiver, 2002-03 and 2003-04"</i></p>

Notes/Questions:

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Indiana: Evidence-Based Waiver	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Complete required courses and obtain 40 credits * Pass the English/language arts and mathematics Graduation Qualifying Exams (GQE) or qualify for a GQE waiver * Any additional local graduation requirements
Exam name & type	Graduation Qualifying Exam (GQE) <ul style="list-style-type: none"> * Class of 2011 will be the last class to take one exam in Grade 10 * Subsequent classes will take two smaller End-of-Course exams
Tested subjects for graduation	English 10 and Algebra I
Grade first administered (for graduation purposes)	At whatever grade the student takes a required course
Number of times students can take exam	As many times as needed
Alternative type(s) and eligible students	Evidence-based waiver (any student who did not pass the GQE)
Alternative described below	Evidence-based waiver <i>(the procedures described below are those specifically for students with disabilities)</i>
First instituted	<i>No specific date provided</i>
Eligibility requirements	<p><i>The process for non-disabled students is similar to that for students with disabilities, except that the IEP team is involved.</i></p> <p>An eligible student:</p> <ul style="list-style-type: none"> * Takes and retakes the graduation examination in each subject area in which the student did not achieve a passing score as often as required by the student’s individualized education program. * Completes remediation opportunities provided to the student by the school to the extent required by the student’s IEP. * Maintains a school attendance rate of at least 95% to the extent required by the student’s IEP with excused absences not counting against the student’s attendance. * Maintains at least a “C” average or the equivalent in the courses comprising the credits specifically required for graduation by the State Board of Education. (Students who entered high school between 2000-01 and 2005-06 and applying for waivers need to complete 24 required credits 2000-01/2005-06; students entering high school 2006-07 onwards need to complete 34 required credits)
Request process	Parents make request.
Remediation required?	Yes. Local district determines type/nature of intervention
Level of administration/ scoring	Local

Indiana: Evidence-Based Waiver	
Administration process and timing	The student’s teacher of record, in consultation with a teacher of the student in each subject area in which the student has not achieved a passing score, makes a written recommendation to the IEP team. Process takes place in student’s senior year.
Evidence considered	<ul style="list-style-type: none"> * Tests other than the GQEs * Classroom work * Student still needs to meet the GPA and attendance requirements.
Scoring process and timing	<ul style="list-style-type: none"> * Teacher recommendation to the IEP team must be approved by principal. * A local waiver committee then reviews all submitted waivers (for both disabled and non-disabled students). The IEP team does not make the decision. * Reviews conducted in the spring of student's senior year
Uniformity in administration / scoring?	Indiana is a local-control state; therefore there is a lot of discretion at the local level regarding the level of evidence and the waiver decisions. Although certain criteria must be met (e.g., attendance, GPA, required credits), the state does not have any prescribed components. Districts establish their own procedures and forms (e.g., review a formal letter with supporting evidence or have an informal discussion with the teacher).
Monitoring	<ul style="list-style-type: none"> * The state examines the number of waivers at the district level and could visit/question districts with high waiver use. * State does not approve waivers/evaluate evidence.
Associated costs	Respondent identified the local committee meetings and remediation as local costs. The state provides some fiscal resources to schools for the remediation component of the process; funding has not been established this year due to changes in the assessment.
Challenges/ concerns	<ul style="list-style-type: none"> * The very high attendance rate bar and the fact that GPA and attendance requirements are based on a four-year average. As a result, students struggle in their freshman year may lose out on the opportunity to utilize the waiver option when they are seniors. * Looking into changing the waiver so that it shows improvement over time (growth model) * Difficult to prove competency of students who have been out of high school for several years and then select the waiver
Information on the use and results of alternative means	<i>No data found</i>

Notes/Questions:

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Maryland: Task-Specific Alternative	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Obtain 21 course credits (may be increased by districts) * Pass four High School Assessments * 75 hours or equivalent of service learning
Exam name & type	High School Assessments (HSA) (multiple choice end-of-course tests)
Tested subjects for graduation	Algebra/Data Analysis, Biology, Government, and English
Grade first administered (for graduation purposes)	At whatever grade the student takes a required course
Number of times students can take exam	As many times as needed
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Combined score on HSAs of 1602 (all students) * Concordant scores from other tests (e.g., AP) (all students) * Modified High School Assessment (IEP students) * Bridge Plan for Academic Validation (all students)
Alternative described below	Bridge Plan for Academic Validation
First instituted	2008
Eligibility requirements	<p>An eligible student has:</p> <ul style="list-style-type: none"> * passed the HSA-related course; * did not pass an HSA after two attempts; * participated in locally-administered or approved assistance (such as sample test reviews, tutoring, or after-school or Saturday instruction; the type of assistance is at the district's discretion); * is making satisfactory progress toward graduation (satisfactory GPA and attendance); * has demonstrated satisfactory attendance as determined by the local school system.
Request process	<ul style="list-style-type: none"> * A student/parent, teacher or other school staff make request * Student, parents, and school staff meet to complete an agreement and plan form
Remediation required?	Yes. Districts required to offer assistance to any student failing the HSA
Level of administration/ scoring	Local

Maryland: Task-Specific Alternative	
Administration process and timing	<ul style="list-style-type: none"> * A school staff member selects pre-approved projects aligned with each content expectation (called modules). A sample project for Algebra is: “Design and conduct an investigation to determine the length of time students at a school spend on homework per week. Using the data from that investigation, design and conduct a simulation to estimate the probability that 5 randomly selected students will all spend at least 4 hours on homework per week” (HSAexam.org). The number of projects is determined by the student's highest score. * Student has 7-10 weeks to complete tasks (between 10-20 hours). * Package submitted to district-level review panel must include ethics form
Evidence considered	Project-specific work
Scoring process and timing	<ul style="list-style-type: none"> * District-level panel reviews package and makes recommendation to the district superintendent * Panel comprised of one central office staff familiar with Bridge Plan requirements; two content-certified staff for each content area being reviewed (school staff allowed, but cannot have been student's teacher for HSA-related course); and other designated staff. One SE and ELL professional serve as consultants. * Each school district uses scoring tools developed by the state at the time the projects are written. Each module will be worth a certain number of points; upon successful completion of the module, the student earns its total point value. The student needs to earn enough points to make up the gap between the highest score earned on the HSA and the passing score. Multiple modules may be assigned (one module from each content expectation must be chosen before assigning additional modules). * District superintendent validates findings and recommendations of the local panel
Uniformity in administration / scoring?	<ul style="list-style-type: none"> * Districts use state-developed scoring tools. * The state conducts a review of scoring consistency to ensure uniformity in administration and scoring - see “Monitoring” below
Monitoring	<ul style="list-style-type: none"> * The implementation manual calls for the state to collect and analyze data from local school systems to assess the alternative’s effectiveness . * According to the Bridge Plan’s design document, local panels will be asked periodically to submit project modules for state review to assure consistency across and within school systems. While the findings will be used to modify procedures and training, they will not be used to reverse the panels’ decisions. * In 2011, a stakeholder group will review local panel report and statewide analyses and make a recommendation to the state board as to whether to continue the alternative.
Associated costs	<ul style="list-style-type: none"> * According to the respondent, it costs the state about \$200,000 to bring teachers together in the four content areas to develop the projects. * Locals absorb costs for administration (respondent indicated that this should be minimal).
Challenges/ concerns	Communicating with teachers and school officials across the state regarding appropriate levels of assistance staff can provide to students while completing their projects.

Maryland: Task-Specific Alternative

Information on the use and results of alternative means

68% of local scoring panels passed student projects; results for all students similar to those for ELL and special education students

Source: Bridge Plan for Academic Validation: Update for State Board of Education (undated presentation)

Notes/Questions:

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Massachusetts: Collection of Evidence	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Earn a score of Needs Improvement or higher on the grade 10 state assessment for English Language Arts (ELA) and Mathematics, and beginning with the class of 2010, one high school Science and Technology/Engineering assessments * Any other local graduation requirements
Exam name & type	Massachusetts Comprehensive Assessment System (MCAS) (comprehensive)
Tested subjects for graduation	ELA and mathematics
Grade first administered (for graduation purposes)	Grade 10
Number of times students can take exam	<i>Unknown</i>
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * MCAS-Alt (IEP or Section 504 students only) * Appeals processes (Portfolio and Cohort, for eligible students that do not pass MCAS)
Alternative described below	MCAS-Alt
First instituted	2001
Eligibility requirements	Student must have IEP
Request process	IEP team decides how student will participate in state assessment
Remediation required?	n/a
Level of administration/ scoring	<ul style="list-style-type: none"> * Collection done at local level * Scoring overseen by contractor (currently use teachers overseen by contractor and will be moving to professional scorers)
Administration process and timing	<ul style="list-style-type: none"> * Principal identifies school personnel to collect student's work over the course of the school year. All MCAS-Alt administrators are required to go through training before each administration. * Binders of student work and required forms submitted to the state in the spring.
Evidence considered	<ul style="list-style-type: none"> * Work samples, video clips, and/or photographs documenting the student's performance of tasks based on the standards being assessed * Data charts documenting the student's performance over a period of time during activities based on the learning standard being assessed. Data is collected on at least five different dates and must begin at a level of accuracy and/or independence below 80%, in order to show that the student was taught challenging new skills, knowledge, and concepts. * Supporting documentation, including descriptions provided by the teacher, reflection sheets, and other evidence that indicates how the student was instructed and/or how he or she demonstrated knowledge and skills in the subject being assessed. * For the graduation portfolios, evidence from prior years may be submitted.

Massachusetts: Collection of Evidence	
Scoring process and timing	<ul style="list-style-type: none"> * Student portfolios are centrally scored by licensed MA educators, who receive extensive training and must qualify in order to become scorers. * Each content strand is scored with a universal scoring rubric on the following areas: complexity (1-5), accuracy (M-4), independence (M-4). The entire subject gets an overall score for self-evaluation (M - 4) and generalized performance (1 -3). A score of “M” means there was insufficient evidence or information to generate a numerical score in a rubric area. Performance level for each strand calculated based on complexity and the intersect of accuracy and independence; the overall performance level for the content area is based on the average of the performance levels of the individual strands. * Results for the MCAS-Alt are reported as Awareness, Emerging, Progressing, Incomplete, Needs Improvement, Proficient, or Advanced. The standard MCAS results are reported as Warning, Needs Improvement, Proficient, and Advanced. The first three MCAS-Alt results correspond with the "Warning" level.
Uniformity in administration / scoring?	<ul style="list-style-type: none"> * Scorers are monitored closely for their accuracy and consistency during the scoring process. Table leaders (expert scorers) track each scorer’s consistency in scoring portfolios. High school portfolios are double-scored. The table leader will resolve scoring discrepancies, if any, from the prior two sets of scores. * Each scorer’s rate of agreement with an expert scorer will be based on at least ten scores per content area. Agreement must be maintained at a rate of 80% or higher. If the agreement falls below 80%, scorers will be retrained and double-scored for the remainder of that scoring session. Scorers may be released from scoring when their rate of agreement falls below 80% twice or more in one week.
Monitoring	Scoring is monitored through central location, as described above.
Associated costs	<ul style="list-style-type: none"> * Training and scoring * Locals absorb the cost of collecting the evidence, but it should be a part of their work over the course of the year.
Challenges/ concerns	<ul style="list-style-type: none"> * Outreach and training of the teachers to pull together representative samples of student work that taps the standards and challenges students appropriately. * Changing the mindset to hold students to the same standards. * Amount of time it takes to collect student work. Training helps teachers understand the importance of integrating assessment and instruction instead of viewing the MCAS-Alt as something to do at the end of the year.

Massachusetts: Collection of Evidence

Information on the use and results of alternative means

- * About 1% of students with disabilities statewide take the MCAS-Alt. (1)
 - * In 2008, 4 students met the graduation requirement for English Language Arts (ELA), 13 met the Mathematics, and 3 students in Science and Technology/Engineering, by achieving a score of Needs Improvement or higher on the MCAS-Alt (2)
 - * Since 2001, approximately 75 students taking the MCAS-Alt have met the graduation requirement. (3)
 - * In explaining the low numbers that pass using the alternative: "Massachusetts is one of the only states that allows students who take alternate assessments to meet the graduation requirement, but only if they score Needs Improvement or better." (3)
- Sources: (1) <http://www.doe.mass.edu/mcas/alt/about.html>; (2) 2008 MCAS Alternate Assessment (MCAS-Alt): State Summary of Participation and Performance; (3) <http://www.doe.mass.edu/mcas/alt/about.html?section=faq>

Notes/Questions:

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Mississippi: Collection of Evidence	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Obtain 21 course credits (which can be increased by districts) * Pass specific subject area tests
Exam name & type	Subject Area Tests (end of course)
Tested subjects for graduation	U.S. History since 1877, English II (with writing component), Biology I, and Algebra I
Grade first administered (for graduation purposes)	At whatever grade the student takes the tested courses
Number of times students can take exam	<i>Unknown</i>
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Alternative assessment for subject area tests (IEP students only) * Substitute evaluation process (all students)
Alternative described below	Alternative assessment
First instituted	<i>No date provided.</i>
Eligibility requirements	<p>An eligible student:</p> <ul style="list-style-type: none"> * Must have IEP (and the assessment must be documented on the IEP) * Must have taken the standard assessment (with or without accommodations) at least once * Requires a non-allowed accommodation in the subject for which an alternative is requested
Request process	<ul style="list-style-type: none"> * IEP team decides (for each subject) how student will participate * District submits form to the state for approval, along with copy of student's IEP.
Remediation required?	n/a
Level of administration/scoring	<ul style="list-style-type: none"> * Collection done at the local level * Scoring overseen by the state
Administration process and timing	<ul style="list-style-type: none"> * Teacher collects evidence for each competency goal for each course to be assessed by the alternative. Each work sample follows the procedural requirements laid out by the regulations. * District special education supervisor sends package to the state dept of ed. Package needs to include ethics form. * Packages must be submitted in January for fall courses and April for traditional/spring courses.
Evidence considered	<ul style="list-style-type: none"> * Each competency must have a minimum of five work samples as primary evidence; tests and other assignments. * Optional secondary evidence includes letters of support, descriptions by a teacher, homework, group assignments.

Mississippi: Collection of Evidence	
Scoring process and timing	<ul style="list-style-type: none"> * Conducted at the state level by an approved/external Alternative Assessment Review Committee. The committees meet between 1 - 3 days to review and score the portfolios. * Each subject has its own committee, comprised of 3 -4 subject area and regular education teachers (might have 4-5 teachers depending on number of portfolios) (no fewer than three teachers). These teachers are currently teaching the courses assessed, and cannot be from a district that has submitted evidence. * Members are trained in using the scoring rubric, which provides a score 0 - 3 for each competency goal. Scoring is done independently, then averaged. The sum of ratings for each goal divided by the (number of goals multiplied by the number of raters). * Scoring is pass/fail.
Uniformity in administration / scoring?	<ul style="list-style-type: none"> * The quality of the evidence is inconsistent. State has implemented teacher training to identify some of the pitfalls of the portfolios. * Respondent confident of the scoring uniformity.
Monitoring	<ul style="list-style-type: none"> * Scoring is monitored through central location, as described above. * By approving the forms and overseeing the scoring, the state monitors the number of application forms submitted and the number of students passing.
Associated costs	<ul style="list-style-type: none"> * The state does not pay its teachers to participate in the committee; previously the state provided continuing education units as compensation. * Districts must provide for a substitute while the teachers are scoring (district cost) * Training
Challenges/ concerns	<p>Helping people understand that it is not a different standard. Some people believe it gives an unfair advantage, but the respondent believes it is a very rigorous process.</p>
Information on the use and results of alternative means	<ul style="list-style-type: none"> * According to the respondent, the state received 38 applications to review in 2008-09; less than half (18) were approved. Of those, the dept received only 11 portfolios to score. * Six received passing scores; 5 did not. One student had three portfolios for different content areas (1 passed/2 did not) * It is possible that the number of applications/reviews will increase when the state makes changes to the substitute evaluation process.

Notes/Questions:

New Jersey: Task-Specific Alternative	
Graduation requirements for standard diploma	* Obtain 110 course credits (to be increased to 120), and * Pass the exit exam (for students in the 11th grade on or after Sept 2001)
Exam name & type	High School Proficiency Assessment (HSPA) (comprehensive)
Tested subjects for graduation	Reading, writing, and mathematics
Grade first administered (for graduation purposes)	Grade 11
Number of times students can take exam	Students can re-test three times (October, March, and October in subsequent years)
Alternative type(s) and eligible students	Special Review Assessment (SRA) (any student)
Alternative described below	Special Review Assessment (SRA)
First instituted	<i>No date provided.</i>
Eligibility requirements	Student does not pass one or more of the HSPA content area tests in spring of grade 11 and is expected to pass all other graduation requirements
Request process	n/a
Remediation required?	Yes
Level of administration/ scoring	Local (<i>will be changed to a central location</i>)
Administration process and timing	<ul style="list-style-type: none"> * Before taking the SRA, students must participate in a school-designed SRA instructional program for that content area (typically to begin September of senior year) * Secure test materials are made available by June prior to each school year. * Panel members (see “Scoring process” below) selected and administer vendor-developed, standards-aligned Performance Assessment Tasks (PATs)—using same vendor that develops the HSPA— are administered in the 12th grade over several weeks or months. PATs are untimed, open-ended questions. * Students must complete 2 PATs in each content area cluster/standard (need 4 in language arts and 8 in math) * Students were previously allowed unlimited opportunities to pass
Evidence considered	Task-specific work

New Jersey: Task-Specific Alternative	
Scoring process and timing	<ul style="list-style-type: none"> * Scoring done by two district-appointed panels (one for each content area). Comprised of two teachers certified in content area. Most districts have a single high school (therefore their scoring teams will be affiliated with the school). * Scoring rubrics for math and reading are item-specific (based on the specific PAT). Two content-certified team members independently score the PATS. Scores must agree within one point. Final score is average. Proficient score is two. If scores are not adjacent, a 3rd member scores and the final score is the mean of the two highest scores. (2 of the 3 scores must be adjacent) * Generic scoring rubric for math: 0 – 3 points; Reading: 0 – 4 points; Writing: scored holistically using the New Jersey Registered Holistic Scoring Rubric (1-6 points)
Uniformity in administration / scoring?	While the PATs are uniform and released prior to each testing period, there are time period variations and concerns about scoring. See challenges section.
Monitoring	<ul style="list-style-type: none"> * The state conducts onsite verification of selected districts; the onsite team reviews a sample of students’ actual responses to the PATs, the scoring, panel members certification, and other documentation. * Each county is required to send a report to the state, reporting who passed SRA. However, SRA data has been "inconsistent and of limited usefulness." Only successful SRA administrations are reported. ¹ * Monitoring of the process did not receive the needed funding, even though the department knew it was an important component.
Associated costs	<ul style="list-style-type: none"> * Local costs: Instructional/remediation costs, as well as the costs of evaluating to see if the remediation was successful. * Future state costs: Regional scoring centers and reporting the results back to the districts.
Challenges/ concerns	<ul style="list-style-type: none"> * Critics have described the SRA as a backdoor for students to get a diploma without earning it. * New standards issued in June to address concerns: centralized scoring by NJ teachers recruited and trained by state testing vendor; schools have three weeks at a time to administer PATs (once in January/February; once in March/April); students no longer have unlimited opportunities to pass the PATs; PATs will not be available until January. * Some districts’ graduation rates rely heavily upon the SRA. Under new standards, districts graduating more than 15% of its seniors through the SRA are required to submit plan to reduce reliance on the SRA. * Equivalence. The items are so different (multiple choice versus open-ended) and the reliance upon teacher judgments. * There are concerns about the accuracy and security of the tests. * Inconsistent, limited data

New Jersey: Task-Specific Alternative	
Information on the use and results of alternative means	<ul style="list-style-type: none">* The majority of students using the SRA option are general education students.¹* Approximately 12% of the total student population utilized the SRA option in 2007.²* In 2006, 80.4 percent of graduating students met the testing requirement through the HSPA, while 12.5 percent did so through the SRA.¹* At the school level, 19 high schools had 50% or more of their students meet the testing requirement through the SRA option in 2006.¹* A study of 105 schools where 10% or more seniors used the SRA showed just 4% of the 6,620 students failed it; 71% of the students had taken or were enrolled in Algebra II; 63% had taken college prep English; and 9% were enrolled in honors or Advanced Placement English.³

¹Standards and Expectations: Alternative Approaches to Meeting New Jersey's High School Assessment Requirements (2008)

² *Special Review Assessment (SRA): Additional information concerning districts with high SRA usage rates (March 19, 2008)*

³"N.J.'s alternative test to pass high school is subject of intensive study" *The Star-Ledger (June 15, 2009)*

Notes/Questions:

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

North Carolina: Collection of Evidence	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Complete required courses and obtain 20 credits (for freshman entering 2008-09 following the career/college prep track), * Score proficient on five end-of-course tests and a Computer Skills Test, * Successfully complete a Graduation Project, a performance-based component developed, monitored, and scored locally, and * Complete any local graduation requirements
Exam name & type	End of Course (EOC) tests (multiple choice)
Tested subjects for graduation	Algebra 1 (unless exempted by the student's Individualized Education Program), Biology, Civics and Economics, English I, and U.S. History (for students who entered ninth grade in 2006-07 or later)
Grade first administered (for graduation purposes)	At whatever grade student takes required course
Number of times students can take exam	Maximum of three times. Second re-test must be preceded by "focused intervention/remediation" (re-test within three weeks of first test)
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Review process (all students who do not pass an EOC test) * North Carolina Checklist of Academic Standards (NCCLAS) (IEP and LEP students)
Alternative described below	North Carolina Checklist of Academic Standards (NCCLAS)
First instituted	<i>No specific date provided.</i>
Eligibility requirements	<ul style="list-style-type: none"> * Students with disabilities must have an IEP and should be those students who are unable to access the pencil and paper test even with accommodations. * The student's IEP must document that he/she is eligible for NCCLAS for specific assessments.
Request process	* IEP team decides how student will participate in state assessment
Remediation required?	Yes
Level of administration/ scoring	Local
Administration process and timing	* Work samples are maintained throughout the school year by first scorer
Evidence considered	<i>No specific information found.</i>

North Carolina: Collection of Evidence	
Scoring process and timing	<ul style="list-style-type: none"> * Each student has two scorers (assigned by principal) for each content area assessed. First scorer should be student’s primary instructor in assessed content area; second scorer should have some knowledge of student but is not the primary instructor * During the middle of the year, the first scorer completes a Beginning Student Profile * A Final Student Profile is completed for the student at the end of the school year. * Scorers assess work samples independently following a “checklist” of skills based on specific objectives from the standard course of study. * Scorers submit scores via an online system. The final score is calculated by computer program using the goal-level scores. * Scores are reported as Achievement Levels I-IV. Levels III and IV are considered proficient. * If the student does not pass the NCCLAS the first administration, teachers create a Personalized Education Plan (PEP) which focuses the intervention/remediation. The PEP includes the diagnostic evaluation, intervention strategies, and monitoring strategies. (An IEP cannot substitute the PEP). Student is then re-evaluated after the remediation. If the student fails to score a Level III or above on the re-evaluation, then the documentation is reviewed by a committee, which makes a recommendation to the principal as to whether the student has demonstrated proficiency on the EOC standards. Principal makes final decision. (Note: this is the same review committee option that is available to <i>all</i> students failing an EOC. Committee is appointed by the LEA and membership is comprised of teachers, principals, and central office staff. If a student with disability is being considered, the committee must include special education personnel. Respondent noted that there are variations in the principals’ decisions, and some school districts have tried to pressure principals to not pass students.)
Uniformity in administration / scoring?	The evidence is scored locally and the respondent noted there are not enough safeguards in place to ensure validity. Districts may be looking at different types of evidence to make judgment.
Monitoring	The state is currently doing a review of selected school districts of how they complete the checklist.
Associated costs	The process is administered locally; the state doesn’t monitor or document how much it costs. Respondent said the dept estimated that it costs \$39 per student to administer, which includes the time to collect the evidence and scoring. The cost of teacher time is absorbed by the school site.
Challenges/ concerns	<ul style="list-style-type: none"> * Ensuring that IEP teams are making the appropriate decisions about who is offered the NCCLAS (the state offers training). * NCCLAS under close scrutiny by the US DOE, and respondent expects that the NCCLAS will be eliminated soon. Respondent believed the reason for the scrutiny is because the comparability with the standard EOCs is not strong enough.

North Carolina: Collection of Evidence

Information on the use and results of alternative means

- * See attached document
- * Between 0.25% (Algebra I) and 0.5% (English I) of all valid test scores in 2007-08 were from NCCLAS. Between 0.3% and 0.4% of all valid test scores for students with disabilities were from NCCLAS.
- * Approximately 9% of students taking the NCCLAS for Algebra I and 6%-8% of those taking it for English I have a disability. The majority of those students have a specific learning disability; the next largest disability category is "other health impairment."
- * 92% of students with disabilities taking the NCCLAS for Algebra I passed in 2007-08; 80% of students with disabilities taking the NCCLAS for English I passed (higher pass rates than the non-disabled students taking the alternative).

Notes/Questions:

Data on Students Taking the North Carolina Checklist of Academic Standards (NCCLAS)

Source: Reports of Disaggregated State, School System (LEA) and School Performance Data for 2006 – 2008 (<http://disag.ncpublicschools.org/2008/>)

Number and Percent of Students At or Above Achievement Level III in Algebra 1 Students Taking NCCLAS Alternate Assessments

Student Subgroup	2006-2007			2007-2008		
	# At or Above Level III	# Valid Scores	Percent At or Above Level III	# At or Above Level III	# Valid Scores	Percent At or Above Level III
Students With Disabilities	24	33	72.7%	29	32	90.6%
Non-Disabled Students	245	341	71.8%	239	315	75.9%
Serious Emotional Disability	4	6	66.7%	*	*	*
Specific Learning Disability	13	16	81.3%	-	17	>=95%
Other Health Impairment	7	11	63.6%	7	8	87.5%

Number and Percent of Students At or Above Achievement Level III in English 1 Students Taking NCCLAS Alternate Assessments

Student Subgroup	2006-2007			2007-2008		
	# At or Above Level III	# Valid Scores	Percent At or Above Level III	# At or Above Level III	# Valid Scores	Percent At or Above Level III
Students With Disabilities	35	44	79.5%	35	44	79.5%
Non-Disabled Students	335	644	52.0%	339	502	67.5%
Intellectual Disability - Mild	5	6	83.3%	*	*	*
Specific Learning Disability	17	23	73.9%	22	27	81.5%
Other Health Impairment	-	8	>=95%	9	13	69.2%

* Indicates that the student population in the subgroup is too small to report the value
 The percentage and number of students are not shown if the percentage is greater than 95% or less than 5%
 Subgroups with no data are not shown in the table above

Study of Certain Students Who Used Modifications and/or Accommodations on the CAHSEE

Virginia: Collection of Evidence	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Earn at least 22 standard units of credit by passing required courses and electives, and * Earn at least six verified credits by passing end-of-course tests or other assessments approved by the Board of Education (applies to the graduating class of 2004 and onwards)
Exam name & type	Standards of Learning (SOLs) (end of course)
Tested subjects for graduation	English, mathematics, laboratory science, history & social sciences, and a student-selected test in computer science, technology, career and technical education or other areas as prescribed by the State Board of Education
Grade first administered (for graduation purposes)	At whatever grade the student takes a required course
Number of times students can take exam	<i>Unknown</i>
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Concordant scores from other tests (ACT, SAT, AP, etc.) (all students) * Virginia Substitute Evaluation Program (VSEP) (IEP or 504 Plan students only)
Alternative described below	Virginia Substitute Evaluation Program (VSEP)
First instituted	2002-03
Eligibility requirements	<ul style="list-style-type: none"> * The student has a current IEP or 504 Plan (or one that is being developed). * The student is enrolled in a course or has passed a course that has a Standards of Learning end-of-course test and/or the student is pursuing a Modified Standard Diploma and seeking certification to meet the literacy and/or numeracy requirements. * The impact of the student's disability demonstrates to the IEP team or 504 committee that the student will not be able to access the SOLs, even with accommodations, and requires a substitute evaluation. * Accommodations have been considered for use with the student in the course/content submission area.
Request process	<ul style="list-style-type: none"> * IEP team decides how student will participate in state assessment * District submits intent form and plan to the state dept of education for approval
Remediation required?	n/a
Level of administration/scoring	<ul style="list-style-type: none"> * Collection done at the local level * Scoring overseen by contractor.
Administration process and timing	<ul style="list-style-type: none"> * Teachers collect evidence for each standard and string * Materials submitted to the state along with required forms (including affidavit/ethics form), in January for fall courses and May for spring.

Virginia: Collection of Evidence	
Evidence considered	<ul style="list-style-type: none"> * May include, but is not limited to, work samples, audios, videos, anecdotal records, interviews, charts/graphs, and other student-generated work that demonstrates a level of performance in the selected content area submitted. * Multiple-choice evidence will not be scored.
Scoring process and timing	<ul style="list-style-type: none"> * The contractor selects and trains scorers. Evidence must demonstrate the required SOL content identified in the test blueprint for the area being assessed and the level of the student's individual achievement. * Evidence is double-scored using a 0-4 point rubric. * Scoring developed by the state. In the "early days," state did the scoring. They turned it over to a contractor for two reasons: It was not the best practice to have Department of Education employees score work sample based assessments, and secondly, to better manage the volume of collections. * Documentation shows two distinctions in passing scores: pass and pass advanced (determines if student is eligible for advanced degree).
Uniformity in administration / scoring?	Respondent was confident that scoring was uniform. Scoring process involves training, anchor papers, and scorers have to qualify and they are "monitored closely."
Monitoring	Department monitors process through approval of forms/plans; scoring is monitored through central location.
Associated costs	<ul style="list-style-type: none"> * Vendor contract for scoring * Training/workshops in the fall. Because of the budget situation, the state is holding that training through a webinar this year. * Collecting evidence is a local cost. Respondent noted that if done as part of the instructional process, it should be minimal.
Challenges/ concerns	* Helping school districts understand which students are appropriate for this assessment and developing an appropriate justification.
Information on the use and results of alternative means	<ul style="list-style-type: none"> * 184 collections submitted across 11 end-of-course content areas for the 2007-08 school year * 18 submitted for English-reading with 66% passing * 16 submitting for English-writing with 69% passing * 32 submitted for Algebra I with 78% passing <p><i>Source: Virginia Substitute Evaluation Program: The VSEP Process (August 2008 presentation)</i></p>

Notes/Questions:

Washington: Collection of Evidence	
Graduation requirements for standard diploma	<ul style="list-style-type: none"> * Earning 19 course credits in specific subject areas * Completing a high school and beyond plan * Completing a culminating project * Meeting assessment standards in reading, writing, and math
Exam name & type	Washington Assessment of Student Learning (WASL) (comprehensive) <i>(Changing it to the "High School Proficiency Exam")</i>
Tested subjects for graduation	Reading, writing, and mathematics <i>(Math currently suspended due to funding issues)</i>
Grade first administered (for graduation purposes)	Grade 10
Number of times students can take exam	Up to five times
Alternative type(s) and eligible students	<ul style="list-style-type: none"> * Collection of evidence (all students who do not pass WASL) * WASL/grade comparison (all students who do not pass WASL) * Concordant scores from other tests (e.g., AP)
Alternative described below	Collection of Evidence (COE)
First instituted	2006-07 (Summer 2007 scoring)
Eligibility requirements	An eligible student must have taken the WASL at least once
Request process	Testing coordinator at the district identifies students who did not pass the WASL.
Remediation required?	Not required, but as a result of the collection, instruction improved. Emergence of "COE" classes with "very intentional" instruction.
Level of administration/ scoring	<ul style="list-style-type: none"> * Collection done at the local level; * Scoring overseen by the state
Administration process and timing	<ul style="list-style-type: none"> * A designated school-level facilitator is responsible for collecting and maintaining the student work and required forms * Materials submitted to the state in February (for seniors only) or June * Collection must adhere to standardized components and include all required forms, including ethics form. Must have between 8-12 work samples and be aligned to the standards.
Evidence considered	<ul style="list-style-type: none"> * With the intention of minimizing variation in evidence, the state has very specific requirements, such as how many work samples to include, what types of assessments are acceptable as work samples, the minimum number of strands that must be represented in each work sample. * The writing content area required 6 - 8 samples; reading requires 8-12 samples; and mathematics requires 8-12 samples. At least two work samples in each content area must be produced in an "on-demand" setting" (e.g., replicated classroom testing).

Washington: Collection of Evidence	
Scoring process and timing	<ul style="list-style-type: none"> * Scoring done twice a year (March for seniors only and July) at central scoring center; 120 scorers comprised of current/retired high school teachers selected by the dept through an application process. * All scorers receive training (approx 2 days) and are evaluated at the end of the training (must score 80% to qualify) Training includes full set of anchor points. * There are two stages: sufficiency scoring (checks to see if collection has required number of work samples, etc.) and evidence scoring. * Every binder is scored twice; if discrepancy, a third scorer reconciles the scores. Table leaders (expert scorers) monitor the scoring process. * 4-point rubric for each skill within a content area; writing scored holistically * For reading and math, the two highest scores for each strand are used in the final score. In writing, the scores for persuasive, expository, and conventions are added together for the final score. * Proficiency: 78 out of 128 points for math; 72 out of 96 points for reading; 17 out of 24 points for writing
Uniformity in administration / scoring?	<ul style="list-style-type: none"> * Respondent was confident that the scoring was uniform. * Reliability and validity checks are conducted and scoring data analyzed daily during the scoring timeframe. Each scorer has two validity collections a day; data are checked for "scorer drift" to make sure everyone is holding steady. There is also re-training for scorers who consistently score lower or higher than the adjacent score. Scorers are required to have exact scoring with the second scorer 50% of the cases. * Consistency of evidence at the local level is improving over time. Certain parts of the state appear to have different collection quality; respondent believed this was due to lack of training of the teachers and that the teachers have not organized their classrooms appropriately for remediation/quality of collection. For example, some high schools put students using this option in "COE classes" to provide intensive remediation and create the COE; other sites do not offer intervention. The districts make the decision as to whether to offer these interventions. * Some districts require students to participate in the COE; others let the student decide.
Monitoring	<ul style="list-style-type: none"> * Scoring is monitored through central location. * State has done an extensive report on the technical quality of the COE, including evidence for the reliability and validity of test scores. The latest report (2008) presents results on the inter-rater reliability from the 2008 scoring process. <p>[Note: The report presents analyses on correlation between COE and WASL scores (for the same students); evidence supports validity for reading and writing. In math, scores were divergent; trend analyses suggest that some students who passed the math COE were not "motivated" to do well on the WASL math.]</p>

Washington: Collection of Evidence	
Associated costs	<ul style="list-style-type: none"> * If the collections are submitted on time and pass the intake and sufficiency requirements, the district receives a \$300 stipend (these are discretionary funds). Because of the number of math collections, math has been temporarily suspended as a graduation requirement. Respondent noted that the stipend showed that the state was serious about this alternative. * Compensation for scoring teachers (pay, lodging, travel, meals). Districts reimbursed for substitute teachers. * Training (optional for teachers; on-site and web-based); moving to web-based exclusively next year to lower cost
Challenges/ concerns	<ul style="list-style-type: none"> * Scores are not equated; trying to explain to the public how raw scores align with scale scores. * Teachers who have not been trained (they don't give their students an equal opportunity to pass the COE)
Information on the use and results of alternative means	<ul style="list-style-type: none"> * Respondent estimated that approximately 15% of the total student population uses the collection for reading/writing and 25% for math. * In the February 2008 submission, 6% of the 598 students taking the reading COE; 6% of the 617 students taking the writing COE; and 6% of the 1,041 students taking the math COE were special education. 55% of those special education students passed the reading COE; 21% passed the writing; and 62% passed the math COE. <p><i>Source: High School Collections of Evidence Technical Report 2008</i></p>

Notes/Questions: