CALIFORNIA STATE BOARD OF EDUCATION ITEM # 8

NOVEMBER 2003 AGENDA

SUBJECT		
California Assessment System: Test Item Release Plan for the California Standards Tests (CSTs), the California High School Exit Examination (CAHSEE), and the California Alternate Performance Assessment (CAPA)	☒	**Action**
	☒	**Information**
	☐	**Public Hearing**

Recommendation:

Approve the plan for annually releasing subsets of California Standards Test (CST), California Alternate Performance Assessment (CAPA), and California High School Exit Examination (CAHSEE) items and approve releasing subsets of 2003 CST items.

Summary of Previous State Board of Education Discussion and Action

- The State Board of Education (SBE) previously approved releasing CAHSEE items to help students, parents, teachers, administrators, and community members understand the nature of the examination.
- SBE had discussed releasing items from the CSTs, but the item pool for developing the tests did not contain a sufficient number of items to allow for releasing more than a few sample items.
- At its March 2003 meeting, SBE asked for an annual item release plan to include the CSTs, CAHSEE, CAPA, and other tests in the State Assessment System and postponed any item release until a plan is approved.
- Since March 2003, there have been subsequent board discussions related to the item release plan.

Summary of Key Issue(s)

- The California Department of Education (CDE), SBE staff, and SBE testing liaisons have worked with Educational Testing Service (ETS) to develop an Annual Item Release Plan.
- CDE plans to release approximately 20 percent of the 2003 CST items and 25 percent of CAHSEE items in each subject following SBE approval of the Item Release Plan. There are no plans to release any CAPA items until at least 2005.
- The number of items that can be released is based on the quality and quantity of items in the item database for developing subsequent tests that meet the highest professional standards of validity and reliability and the budget available for developing replacement items.
- SBE and CDE will continue working to determine how selected items can be developed as examples that demonstrate how they are used to assess California's

Summary of Key Issue(s)

Academic Content Standards.

Fiscal Analysis (as appropriate)

None provided that funding is available in the program budgets for ongoing item development.

Attachment(s)

Attachment 1: Outlining a Consistent Item Release Strategy for California (Pages 3-15)

**Discussion of Long-Term Item Utilization for the
California Standards Tests and California High School Exit Examination
Educational Testing Service (ETS) DRAFT – 10/23/03**

Background

California State Board of Education and Department of Education wish to develop a long-term plan that will predict the amount of item development that will be required, over the next several years, to sustain the California Standards Tests (CST), California High School Exit Examination (CAHSEE), and other state assessment programs. This document outlines some suggestions SBE and CDE may wish to consider in developing a long-term item utilization plan.

Focus of This Document

This document outlines a possible item utilization plan through 2011, using the English Language Arts tests for the CSTs and CAHSEE as a basis for projections. The general discussion in this draft, however, is applicable to all content areas. By October 31, 2003, ETS will provide similar detailed analyses for each content area in the CSTs and CAHSEE. That is, tables like Tables 1, 5, 6, 7, 8, and 9 will be created for each content area and test, including CAPA.

Phase One of Item Bank Development

The state's approach in both the CST and CAHSEE programs has been to require the development of a large number of items over a three-year period so that a large item bank is quickly created for each content area in each program. Table 1 shows that during 2002 and 2003, ETS field tested 696 and 1488 items for the ELA portion of the CAHSEE. In Fall 2002, ETS field tested 950 ELA items (4 versions X 25 items for grades 3-10 and 3 versions X 25 items for grades 2 and 11). In spring 2003, ETS field tested 1200 items for the ELA CSTs (20 versions X 6 items for grades 2-11). In 2004, approximately 1218 items will be field tested for CAHSEE, and 1356 will be field tested for the CSTs (25 versions X 6 items for grades 6-11, 20 versions X 6 items for grades 3-5, and 16 versions X 6 items for grade 2). For the entire 2002 to 2004 period, ETS will have field tested 3506 items for the ELA CSTs and 3402 items for CAHSEE.

Table 1. Numbers of ELA Items Field Tested, CSTs and CAHSEE 2002 – 2004

ELA Field Tested Items	2002		2003		2004	
Test	CSTs (Fall FT)	CAHSEE	CSTs	CAHSEE	CSTs	CAHSEE
Grade 2	75	-	120	-	96	-
Grade 3	100	-	120	-	120	-
Grade 4	100	-	120	-	120	-
Grade 5	100	-	120	-	120	-
Grade 6	100	-	120	-	150	-
Grade 7	100	-	120	-	150	-
Grade 8	100	-	120	-	150	-
Grade 9	100	-	120	-	150	-
Grade 10	100	696	120	1488	150	1218
Grade 11	75	-	120	-	150	-
Total FT items	950	696	1200	1488	1356	1218
Totals by Year	1640		2688		2574	
Totals by Program:						
CST 3506			CAHSEE 3402			

Similar numbers of items have been field tested for the mathematics tests in both programs as well as the science and history-social science CSTs. Also, in 2002 and 2003, 600 items were approved by the CAPA item review committees, with 80 to be placed on mathematics and ELA operational forms in 2004 and 112 to be field tested in science in 2004. The item bank equilibrium data for these tests will also be provided by October 31.

Phase Two of Item Bank Development

Now that large numbers of items have been developed for the state programs, it is possible to determine the following key aspects of the second phase of item development

and to ascertain how much future development is required to reach and sustain an equilibrium.

1. *Optimal Size of Bank.* The first important consideration in developing a long-term development plan is the size of the item bank that is required to create high quality test forms over several years. With too small a bank, it is difficult to create forms that fulfill the blueprint and also meet appropriate psychometric requirements. That is, without a sufficiently large bank, it is not possible to build operational forms that assess various components of the standards, provide a variety in item types (e.g., with or without mathematical context), and contain items that do not clue each other. ETS also recognizes that there are unnecessary expenses associated with developing and maintaining too large a bank.

Based on our experience with the construction of tests for both CAHSEE and the CSTs, ETS suggests that, with the exception of ELA, the content area item banks should, at equilibrium, contain 4 to 5 times the number of items annually required for building operational forms. For ELA, we suggest 5 to 6 times the number of items annually required for building operational forms. The larger ratio for ELA is based on the fact that most of the items are passage-based. Therefore, releasing or discarding a passage results in a proportionally greater loss of items in the bank.

2. *Field Test Survival Rates.* A second important consideration in determining the best size for an item bank at equilibrium is the percentage of new items that can be expected to survive after being field tested. With one year of CST data and two years of CAHSEE data, ETS has observed that this percentage varies by content area and grade. The data are provided in approximate values in the following tables (note that general mathematics, integrated mathematics, or integrated science CSTs are not included in Table 2 because no field testing was done for these courses):

Table 2. Approximate Percentage of Usable Items from Field Testing
California Standards Tests – 2003 Administrations

60-70%	70-80%	80-90%	90%+
H.S. Math	Biology	Algebra I	Science grade 5
Physics	Earth Science	Algebra II	ELA grade 10
	ELA grade 11	Chemistry	ELA grade 2
	ELA grade 3	ELA grade 4	ELA grade 5
	ELA grade 6	ELA grade 9	ELA grade 7
	Math grade 6	H/SS World	ELA grade 8
		Math grade 7	Geometry
			H/SS U.S. History
			H/SS grade 8
			Math grade 4
			Math grade 3
			Math grade 5
			Math grade 2

Table 3. Approximate Percentage of Usable Items from Field Testing CAHSEE - 2002 and 2003 Administrations

70%	75%
Mathematics	English-Language Arts

These data have been used in the detailed projections of the item bank inventory through 2011, which are summarized in Tables 5 and 6.

3. *Released Items.* A third consideration in determining a long-term item development plan is the numbers of items expected to be released on an annual basis. Recognizing that the percentage and format of the release are policy decisions, ETS has based its calculations for the equilibrium of the item bank on what we believe is the current SBE thinking about item release, as shown in the following table:

Table 4. Annual Percentage and Numbers of Released Items

Year	CAHSEE		CST		CAPA		CELDT	
	%	# per test	%	# per test	%	# per test	%	# per test
1998			0	0				
1999			0	0				
2000			0	0				
2001	75	60	0	0			0	0
2002	75	60	0	0			0	0
2003	25	18-20	20	12-15	0	0	0	0
2004	25	18-20	20	12-15	0	0	0	0
2005	25	18-20	25	15-19	25	2	10	3-8
2006	25	18-20	25	15-19	25	2	10	3-8
2007	25	18-20	25	15-19	25	2	10	3-8
2008	25	18-20	25	15-19	25	2	10	3-8
2009	25	18-20	25	15-19	25	2	10	3-8
2010	25	18-20	25	15-19	25	2	10	3-8

ETS understands that the percentage of items to be released for CAHSEE is 25 percent annually. For the CSTs, 20 percent will be released in 2003 and 2004, with 25 percent released in subsequent years. Beginning in 2005, it is anticipated that 25 percent for CAPA will be released annually. For CELDT, ETS understands that the release may be as low as 10 percent, with the first release in 2005—subject to SBE and CDE negotiations with the CELDT contractor.

The data in Table 4 have been incorporated in our item bank inventory projections found in Tables 5 and 6.

4. *Attrition.* A fourth consideration in a long-term plan for item development is the natural attrition that occurs in any item bank. Our experience has shown that attrition is caused primarily by three factors:

- Items and/or passages become unexpectedly dated. A passage about student backpacks, while of high interest to students, could become dated, for example, if students turn to another concept for transporting school materials. References in science or mathematics items can also become obsolete over time, despite efforts to anticipate such problems.
- Items or passages become unexpectedly sensitive. For example, items about space shuttles had to be removed after the recent shuttle tragedy. Sensitivities grow and change over a period of years in ways that cannot be anticipated.
- The largest factor influencing attrition is changes in CRP perceptions of item acceptability. In all state programs, there is a slight drift toward more or less rigor in how the standards are interpreted in terms of assessment. It is customary that some percentage of items becomes less acceptable as the state standards become incorporated into instructional materials and become widely used in classrooms. The CSTs experienced a fairly high amount of item bank attrition in 2002, especially in ELA, because when the SAT-9 items were removed from the test

and CRP took the opportunity to examine all the item types in the bank, many items were deemed no longer acceptable.

The 2002 spike in attrition in the CST ELA item bank is reflected in the 2003 data in Table 6. Overall, the attrition rate for ELA is predicted to be 10 percent of the items field tested, with an increase in later years to 15 percent. The expected attrition rate is less for other content areas (mathematics, science and history-social science). ETS's prediction would be 5 percent in the early years and 10 percent in later years. The attrition rate is higher for ELA because items are in groups associated with specific passages.

5. *Item Reuse.* A final consideration in determining a long-term item utilization plan is the number of times an item is used. Traditionally, the CST operational forms have been "refreshed" at the rate of 50 percent, which means that half of all items may serve as equating or linking items. CAHSEE has been refreshed at the rate of 26 to 30 percent, with a linking set of approximately 22 items, as is consistent with industry best practices and also with the need for the security of items on a high stakes test.

In anticipation of the need for a substantial group of release-able items for 2004, CDE gave ETS permission to refresh 60 to 70 percent of the CSTs, using the model of a minimum of 22 anchor items. The purpose of the greater refreshment percentage in 2004 was to create a larger pool of items that had been used at least once operationally, in anticipation of the variety of possible specifications SBE might decide were appropriate for selection of released items. For example, if SBE were to decide that two or three items should be released for a particular standard or that statewide data should be posted for released items, it would be necessary to have a large pool of items that had been used operationally. ETS recognizes that SBE has not elected to use the latter specifications for item release, but both were being discussed last summer, when the 2004 refreshment rate had to be determined.

ETS will return to the 50 percent refreshment model in subsequent years. The 2004 increased rate occurred with no change in the CST test development scope of work. ETS recommends that after items have been refreshed (i.e., removed from operational forms), they remain unused in the item bank for approximately three years. This practice of "resting" items is consistent with industry practices for the security of test items, and it is especially important for CAHSEE, so that the test-taking cohort sees as few repeated items as possible.

With the 50 percent refreshment model, half of the items are used for two or sometimes three years in a row, but then they too should be allowed to rest for at least three years. The number of items ETS suggests for the item banks at equilibrium makes it possible to set aside a large majority of items for three-year periods. This design is reflected in Tables 5 and 6. With this item bank model, half the items typically are used once every four years through the life of the program. Half the items are used twice every five years. The released items are taken from the pool after at least one use, although preference is given to former anchor items that have been used at least twice. As mentioned above, having a sizeable pool of items available for release does not mean that large numbers

must or should be released. A large eligible pool is desirable because it gives flexibility in selection of items according to SBE specifications for the release.

The ELA Item Banks at Equilibrium

Cumulative inventories summarizing project item bank growth are shown in Table 5 for the CAHSEE ELA test and in Table 6 for the ELA CSTs. As mentioned earlier, tables for the other content areas and tests will be completed by October 31. For each year from 1999 to 2011, there are actual or predicted entries for the following variables:

- Items field tested during that year
- Usable items from field testing added to the item bank (assuming survival rates of 80% for the ELA CSTs and 75% for the CAHSEE ELA)
- Items needed for operational forms during that year
- Items released during that year
- Items removed by attrition during the year (e.g., dated/sensitive items or due to changes in CRP approach)
- The cumulative inventory of items after additions to and removals from the item bank

As mentioned earlier ETS recommends building all item banks with the goal of reaching an equilibrium number of 4 to 5 times the number of items annually required for building operational forms (5 to 6 times for ELA because of the passage-based items).

As shown in Table 6, for the ELA CSTs, 730 operational items are required per year (130 for grades 2 and 3; 600 for grades 4 through 11). Taking this number times 6 yields 4380 items as the ideal number for the ELA bank. The desirable number of items at equilibrium would actually be about 4500 items, because the CSTs are divided into grades and should have sufficient numbers of passage-based items per grade.

Because of the current California budget crisis, ETS recommends that, beginning immediately, the growth of the ELA CST item bank be slowed, as shown in Table 6. This slower growth would result in fewer items being developed and field tested in 2005 and 2006 than is called for in the current CST Scope of Work. As Table 6 shows, under the new plan the CST ELA item bank would then reach equilibrium in 2010.

For CAHSEE, the growth of the ELA item bank does not need to be slowed, as item development for this contract is almost complete. The existing contract requires field testing of 5880 multiple-choice ELA items. All of these items have been developed, and almost all of them have been approved. Only a few hundred items remain to be reviewed by CAHSEE committees in February 2004. However, because the number of CAHSEE administrations has been reduced, and, primarily, because the number of field test slots on each ELA form has been reduced from 12 to 7, not all of the 5880 items have yet been field tested. Table 5 shows that the remaining items could be field tested at a steady rate between 2005 and 2011, should the state desire.

Additionally, the CAHSEE contract requires that ETS build 24 forms over the life of the contract, 16 for "regular" administrations and 8 for emergency use. These numbers represent a requirement of six operational and six emergency forms each year. Because there are now only five annual CAHSEE administrations, and because it is now clear that emergency forms will not often be used, the number of operational forms built could be reduced to six annually—five operational forms and one emergency form. As of October 2003, ETS has built eleven operational forms and one emergency form. The lower number of forms would mean that the number of CAHSEE ELA operational items required annually would be 438, not the 730 now specified in the contract. This change would permit the CAHSEE ELA item bank to reach optimal size—6 times the annual number of operational items—in 2004. Additional field testing of the already developed items in 2005 through 2011 would increase the number of items in the ELA bank without significant additional expense to the state.

Table 5. Cumulative Item Inventory for CAHSEE ELA

Year	Items Field Tested	Usable Items Added to Bank (75% of FT)	Items on Operational Forms	Items Released	Items Removed by Attrition	Cumulative Inventory
1999	0	0	0	0	0	0
2000	500	375	0	0	0	375
2001	500	375	168	60	38	652
2002	696	522	504	60	38	1076
2003	1488	1116	252	18*	157	2017
2004	1218	914	438	18	112	2801
2005	354	266	438	18	91	2958
2006	354	266	438	18	40	3166
2007	354	266	438	18	40	3374
2008	354	266	438	18	40	3582
2009	354	266	438	18	40	3790
2010	354	283	438	18	40	4015
2011	354	283	438	18	42	4238

* Note that 18 equals 25% of one CAHSEE ELA operation form with 73 items.

Table 6. Cumulative Item Inventory for the ELA CSTs

Year	Items Field Tested	Usable Items Added to Bank (80% of FT)	Items on Operational Forms	Items Released	Items Removed by Attrition	Cumulative Inventory
1999	615	461	350	0	0	461
2000	684	547	350	0	46	962
2001	350	280	350	0	55	1187
2002	950	760	350	0	28	1919
2003	1200	960	730	146	228	2505
2004	1356	1085	730	146	96	3348
2005	678	542	730	183	109	3598
2006	678	542	730	183	81	3876
2007	550	440	730	183	81	4052
2008	550	440	730	183	66	4243
2009	430	344	730	183	66	4338
2010	430	344	730	183	52	4447
2011	430	344	730	183	52	4556

Table 7 gives an overview of the item bank equilibrium for both CAHSEE and CSTs in English-Language Arts. This table shows the expected numbers of items field tested and surviving field testing, the expected numbers of items released and removed due to attrition, and the items remaining in the bank by 2011.

Table 7. Summary of ELA CAHSEE and CST Item Banks 1999 – 2011

Totals	CAHSEE	CSTs
Total Items Field Tested	6880	8901
Total FT Items Surviving	5198	7090
Total Release of Items	300	1573
Total Expected Attrition	678	960
Items Remaining in Bank	4220	4556

The following diagram summarizes the inputs and outputs that create equilibrium in an item bank. The numbers of items in the chart show typical changes during a given year. By October 31, ETS will update this chart to show a specific subject and year.

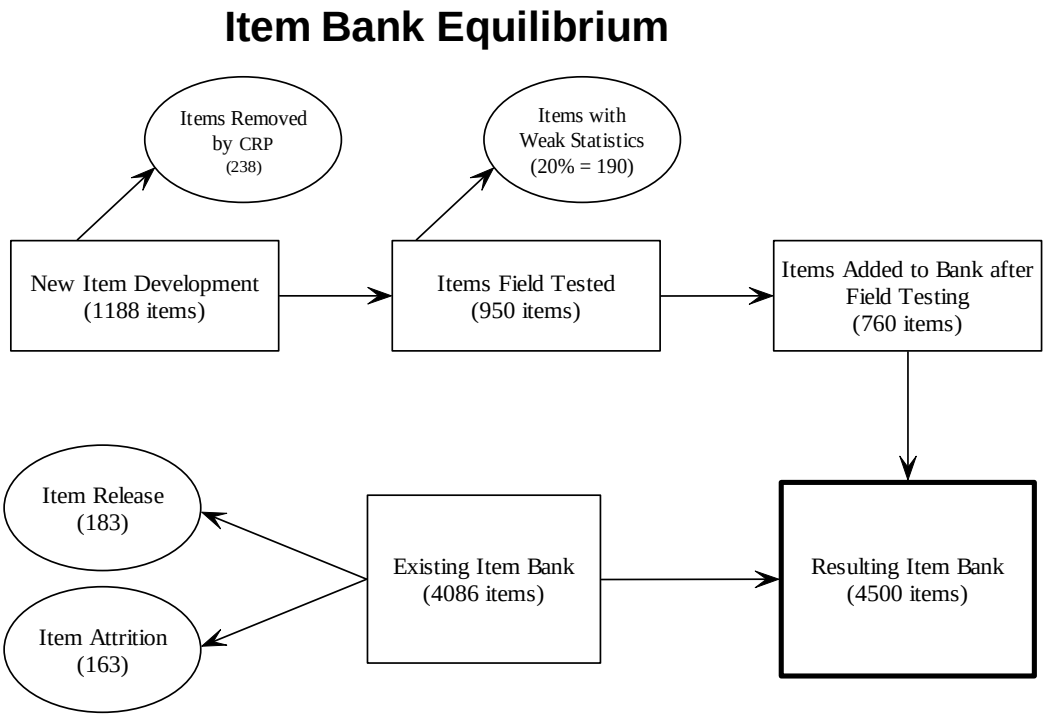

Table 8 shows the number of items that should be developed each year to produce and maintain item bank equilibrium for the CSTs in English-Language Arts. The data are based on the assumption that 80 percent of items taken by ETS to CRP review will be accepted for field testing. ETS has had a 90 percent acceptance rate, on average, in English-language arts, mathematics, and history-social science and a 75 percent acceptance rate in science. The 80 percent figure has been used in this document because it represents an excellent acceptance rate according to general industry standards. Note that the bottom row giving the totals in Table 8 does not calculate to the 80 percent ratio because the numbers of items taken to the CRPs in 1999 and 2000 are unknown.

Table 8. Item Development for ELA CSTs
1999 – 2011

Year	Items Developed for Review by CRP (80% acceptance rate)	Items Field Tested
1999	HEM (unknown)	615
2000	HEM (unknown)	684
2001	438	350
2002	1188	950
2003	1500	1200
2004	1695	1356
2005	1695	678
2006	1188	678
2007	450	550
2008	450	550
2009	300	430
2010	300	430
2011	300	430
Total	9504	8901

Item Release

The following paragraphs summarize ETS's understanding of the plan to be presented to SBE for the release of items, based on the item utilization concepts in this document.

Audiences for the Release

It is ETS's understanding that there are two main audiences to be served by the release of test items. The first is the general public, including the press, who wish to have a better sense of what the tests measure and also want to be assured that the tests are fair to students. The second audience is educators, who wish to understand how the California content standards are measured on state assessments so that standards-based instruction can be improved.

A Plan for the General Public

For the general public, including the press, ETS understands that SBE and CDE may want to see one item from each grade and content area (at least for ELA and mathematics) treated as an "exemplar" item. Each exemplar item would be presented in a context that clarifies the relationship between assessment, standards, and instruction. The context might include, for example, explanations of how the selected item tests the standard, which components of the standard would be tested by other items, how the underlying concept or skills in the standard are expressed at other grade levels, and how the distractors function within the item.

If SBE approves, ETS will ask the CRP members in their first 2004 item review meetings (to be held between January and March) to approve a prototype exemplar treatment and to select items to be given exemplar treatment. The CRP would choose exemplar items

from the sets of items that they approved for release in 2003. ETS will rely on the SBE and CDE providing the CRP with criteria for selection of the exemplars. Once the exemplars are written, ETS will give them to CRP members for review and approval. This plan would allow for development of the exemplars, approval by the CRPs, and publication of the exemplars prior to the next major score release and press event in the state, the CAHSEE data from the March administration.

A Plan for Educators

For educators, ETS understands that SBE and CDE would like to release items in a similar manner to the 2002 CAHSEE release. That is, mathematics, science, and history-social science items would be grouped according to strand or reporting cluster. ELA items would be grouped by passage. Each group would be preceded by a page or half-page of text describing the content of the strand. The language in the introductory text would be taken from the standards and frameworks. Each group of items would be followed by a table giving the answer keys and the standards measured by the items. It is our understanding that statewide p-values (percentage correct) would not be provided in this release. ETS would be pleased to prepare items in this format or in any other SBE-approved format.

Numbers of Items to Be Released

Table 4 in this document shows ETS's understanding of the percentages and numbers of released items for the CSTs and CAHSEE for 2003, as requested by SBE and CDE. Under this plan, a full operational form of each CST could be released after five years. ETS understands that the issue of whether or not the full released form would exactly replicate the blueprint is an open issue. Similarly, whether or not the full released form would replicate the statistical parameters of an actual test is also an open question. Whether or not, over time, items should be presented in relation to the California performance levels (e.g., proficient, advanced) is also to be determined. Finally, ETS understands that SBE may wish to release items according to the depth of coverage in the item bank. Resolution of these questions will be important as the sets of items for 2004 release are selected. Another open question is the treatment of the few standards that are "rotated" annually or biannually.

Selection of Items for 2003 Release

For the CST released items, ETS selected, in January 2003, a draft set of items for potential release. Each set contained approximately 20 percent of an operational form. The criteria for the initial selection included the following:

- At least one item was included from every reporting cluster
- Items represented a range of standards on the operational form
- Items represented a range of difficulties
- Items represented a range of performance levels (e.g., basic, proficient)

ETS presented these items to each CRP at the initial 2003 meeting. At this meeting, the CRP members saw the draft sets as well as the other items eligible for release, and they made changes in the sets as desired. ETS presented the revised sets at the next CRP meeting, where panel members again had the opportunity to make changes. This process

was repeated twice more, so that the CRP members saw the released item sets at each of four meetings held between January and July.

Selection of Items for 2004 Release

We have described the specifications and process used for selection of the 2003 items as context for the SBE to determine what process should be used for 2004 and subsequent years. ETS will be pleased to follow the wishes of the SBE for both the specifications and process to be used.