

Preliminary  
Evaluation Report  
IR-00-27r

June 30, 2000

## High School Exit Examination (HSEE): Year 1 Evaluation Report

Lauress L. Wise  
Carolyn DeMeyer Harris  
D.E. (Sunny) Sipes  
R. Gene Hoffman  
J. Patrick Ford

Prepared for:

**California Department of Education**  
Sacramento, CA

Contract Number: 9234

# High School Exit Examination (HSEE): Year 1 Evaluation Report

## Executive Summary

### Background

California has embarked on a new program to ensure that all students graduating from high school meet minimum standards for verbal and quantitative skills. The California Education Code, Chapter 8, Section 60850, specifies requirements for the High School Exit Examination (HSEE). Beginning with the Class of 2004, students must pass both the English language arts and mathematics sections of this exam to receive a diploma from a public high school in California. Since January 2000, the California Department of Education (CDE) has worked with a development contractor to develop and try out test questions for use in the HSEE. The current schedule calls for testing 9<sup>th</sup> graders on a voluntary basis in March and May of 2001 with mandatory testing of all 10<sup>th</sup> graders (except those passing the exam as 9<sup>th</sup> graders) in 2002. That will be followed by several additional testing opportunities each year for students who have not yet passed the exam.

The legislation specifying the requirements for the new exam also called for an independent evaluation of the HSEE. CDE awarded a contract for this evaluation to the Human Resources Research Organization (HumRRO). Our evaluation will analyze data from a field test of items (test questions) and the annual administrations of the HSEE and report on trends in pupil performance and pupil retention, graduation, dropout, and college attendance rates. The evaluation will include recommendations for improving the quality, fairness, validity, and reliability of the examination. This report describes evaluation activities through June 2000, summarizes the results of these activities, and offers initial recommendations based on conclusions drawn from these results.

### Key Findings and Recommendations

The main conclusions are that a great deal of progress has been made in developing the HSEE and that results to date are quite positive, as indicated by several measures of the quality of the items (test questions). Nonetheless, a great deal remains to be done before the HSEE can be administered operationally. Further, educators surveyed are concerned that students are currently not well prepared to pass the exam. The evidence supporting these conclusions is listed on pages 66–69 in Chapter 6 of this report.

The primary recommendation of this report is based on the evidence that students are not yet well prepared for the HSEE and that many important decisions are needed before the HSEE is ready for operational administration. Our overall recommendation is:

*The State Board of Education, Legislature, and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.*

Three more specific recommendations also are included in this report. First, from discussions with panels of educators, surveys of principals and teachers, and discussions of

the State Board, it is clear that there is confusion about the purposes and nature of California's different high school testing programs. Therefore, our recommendation is:

*Specific Recommendation 1. The Department and the Board need to work together to clarify the relationships and differences among the different high school testing programs, most notably the HSEE, the standards-based STAR assessment, and the Golden State Examinations*

Results from principal and teacher surveys indicate clearly that much needs to be done at the local level to ensure adequate preparation of all students for the HSEE and appropriate remediation for students who do not initially pass it. Our second recommendation is:

*Specific Recommendation 2. The Department and Board should establish, expand, or accelerate processes for communicating with local districts about the HSEE and supporting their preparation for its implementation.*

The HSEE Panel has heard several presentations on testing accommodations for special needs students and English-language learners. The availability and appropriateness of such accommodations is an important legal as well as policy issue. More information may be needed to reach informed decisions. Our final specific recommendation is:

*Specific Recommendation 3. The Department and the development contractor need to gather, review, and discuss more information on the appropriateness and effectiveness of testing accommodations for special needs students and English-language learners.*

More detailed explanations and rationales for each of these recommendations are presented on pages 69–71 of this report. The evaluation activities leading to these conclusions and recommendations are summarized below and described in detail in Chapters 2 through 5 of the report. A brief description of the evaluation plan and Year 1 activities under this plan is presented below, followed by a summary of results as they relate to each of the key evaluation issues.

## **Summary of Activities and Results Leading to the Findings**

### ***Focus of the Evaluation***

The focus of the evaluation is on three key aspects of the HSEE:

- **Quality:** Does the exam provide an accurate and unbiased measure of the knowledge and skills specified in the State content standards to be tested by the exam?
- **Fairness:** Do all students have adequate notice of the new requirement, opportunity to learn the material covered by the exam, and opportunity to demonstrate what they have learned?
- **Consequences:** Do benefits from the exam, in terms of improved student achievement and other positive consequences, outweigh any negative consequences?

## ***Year 1 Activities***

The main body of this report provides a description of the four different activities conducted during Year 1 of the evaluation. The first activity was a *review* of the experiences of 20 states that have preceded California in developing a high school test. The review, described in Chapter 2, provides information on both *consequences* and *fairness* and goes on to discuss issues and concerns with graduation testing programs.

At the outset of the evaluation, a representative sample of 24 districts in California was identified and recruited to participate in a longitudinal study of the effects of the HSEE. Detailed information on programs and results for these districts will supplement statewide data on the key outcomes. The first use of this sample was to recruit two panels of expert teachers and curriculum specialists to *review test questions* as described in Chapter 3 of this report. The panels examined the *quality* of test items and also provided information on whether all students are provided with instruction in the knowledge and skills necessary to answer the test questions, an issue of *fairness*.

The process for developing and reviewing items (specific test questions) yielded a total of 362 English/language arts items and 396 mathematics items that were included in a field test involving more than 7,500 10<sup>th</sup> grade students. Chapter 4 of this report describes preliminary results from an independent *analysis of the field test results* by the evaluation contractor, HumRRO. This analysis assesses the *quality* of the test items and, by implication, the item development process. The analysis also examines the potential accuracy of the HSEE scores and potential passing rates for minority and special needs students and for English-language learners, an issue of test *consequences*.

Chapter 5 describes preliminary results from a baseline survey of teachers and principals from the longitudinal study sample. Issues of *fairness*, in terms of student preparation for the examination, and predicted *consequences* were addressed in this survey. Principals and teachers agree that they are more familiar with state content standards than with the HSEE. Some reported that they had no source of information on the HSEE. Principals believe that students and parents are largely unfamiliar with the exam. Some preparatory activities have already begun, most notably alignment of district curriculum content standards with those of the state.

## ***Results***

The *quality* of the HSEE was best addressed by the item review workshops and the analyses of data from the field test of the HSEE items. Educators participating in the item review workshops concluded that the items were generally well aligned with the targeted standards. Results from the analyses of field test data showed that nearly all test items passed statistical screens indicating appropriate difficulty, the ability to differentiate high and low performing students, and the absence of differential functioning for minority students. Efforts to examine the potential accuracy of the HSEE scores, while very preliminary, are also reasonably positive. Even given conservative assumptions, the error in test scores is small. Roughly 85% of all 10<sup>th</sup> graders would receive a score that is clearly (more than one standard error) above or below the minimum passing score. More than 98% of those students would be classified correctly (passed or failed). Decision errors would be greater for the

15% of students whose achievement is within one standard error of the minimum passing score.

**Fairness** issues were identified in the review of experiences in other states. The court ruling in *Debra P. v. Turlington* (1981) held that all students must be provided instruction in the material covered by the test and that students and parents must be given adequate notice about the test. In a recent case against the Texas Assessment of Academic Skills (*G.I. Forum and Image de Tejas v. Texas Education Agency*, 1997), where the issue was differential impact on minority students due to lower passing rates, the state was required to demonstrate the educational necessity of the exam. Other *fairness* issues include appropriate inclusion and accommodation of students with disabilities and English-language learners. The review suggests that, in implementing a graduation test, a state must establish a structure to support linkages of the exam to instruction at the local level and must attend closely to specific legal requirements.

A key finding on the potential *fairness* of the HSEE was that all students may not have had opportunities to learn the material covered by the test. The item review panels provided an independent assessment of the extent to which students in their districts received sufficient instruction to answer each of the test items correctly. For over half of the mathematics items and more than 90% of the English language arts items, the average ratings across districts suggested that at least one-fourth of the 10<sup>th</sup> grade students had not received instruction that would allow them to answer the test item correctly. These results provide a baseline assessment of the alignment of district curriculum to the contents of the HSEE, prior to adoption of specific test specifications by the State Board. Similar information will be tracked as the HSEE is implemented, to monitor improvements in the alignment of curriculum to the content of the exam. Findings from the item review workshop were consistent with relatively low expected passing rates found in the teacher and principal surveys. Finally, results from the analyses of field test data showed that the test items are relatively difficult for today's 10<sup>th</sup> graders, particularly in mathematics. If these items reflect what we believe students need to know and be able to do, and several panels of reviewers believe that they do, then a significant number of 10<sup>th</sup> grade students are likely to fail this exam. In addition, groups who traditionally score lower on assessments of student achievement will fail at higher rates.

**Consequences** of the HSEE are more difficult to assess at this point, and this was not a primary focus of Year 1 evaluation activities. The principal and teacher surveys provide some insight into concerns and expectations regarding the examination, although these must necessarily be interpreted with caution at this early juncture given that district personnel are not currently well informed about details of the HSEE. Predictions of the impact of the HSEE on student motivation and parental involvement, prior to the first administration and for those who pass, were neutral-to-mildly positive. For those students who fail on the first attempt, some respondents believe that the impact on student motivation will be quite negative, while others believe it could be positive. The respondents expect instructional practices to improve, over time, as a result of the HSEE, but they anticipate that opportunity to learn material covered by the exam will be lower for English-language learners (ELL) and students with disabilities than for the student population as a whole and that fewer of these students will pass the exam.

**High School Exit Examination (HSEE):  
Year 1 Evaluation Report  
Table of Contents**

CHAPTER 1: INTRODUCTION .....	1
The California High School Exit Examination .....	1
Goals for the New HSEE.....	1
The Independent Evaluation of the HSEE.....	2
Contents of the Year 1 Evaluation Report.....	5
CHAPTER 2: BACKGROUND .....	7
Prevalence and Nature of High School Graduation Testing.....	7
From Testing to Higher Achievement.....	10
Anticipated Challenges to California’s High School Exit Exam .....	14
Summary .....	25
CHAPTER 3: ITEM RATING WORKSHOPS.....	28
Item Rating Procedures .....	28
Results .....	31
Conclusions.....	35
CHAPTER 4: ANALYSIS OF FIELD TEST DATA.....	36
Introduction.....	36
Field Test Design .....	36
Item Difficulties .....	38
Item Screening .....	39
Potential Test Accuracy.....	42
Potential Adverse Impact.....	47
Summary .....	48
CHAPTER 5: PRINCIPAL AND TEACHER SURVEYS.....	50
Background.....	50
Survey Development.....	50
Sampling and Administration .....	51
Initial Findings .....	52
Summary .....	64
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS.....	66
General Conclusions.....	66
Recommendations .....	69
REFERENCES .....	73
Appendix A.....	A-1
Explanation of Item Response Theory Concepts Used in Analyzing Potential Test Score Accuracy .....	A-1
Appendix B.....	B-1
Principal Baseline Survey .....	B-2
Teacher Baseline Survey .....	B-9

## List of Tables

Table 2.1. States with high school exit examinations as of 1997–98. _____	8
Table 3.1 Total Items and Number with Low-Standards Alignment Ratings For each Major Language Arts Content Category _____	32
Table 3.2 Total Items and Number with Low Standards Alignment Ratings For each Major Mathematics Content Category _____	32
Table 3.3 Total Items and Number with Low Curriculum Alignment Ratings For each Major Language Arts Content Category _____	34
Table 3.4 Total Items and Number with Low Curriculum Alignment Ratings For each Major Math Content Category _____	34
Table 3.5 Relationship of Item Difficulty to Curriculum Alignment Ratings _____	35
Table 4.1 Average Total Correct Scores by Subject and Field Test Form _____	38
Table 4.2 Average Total Scores by Gender _____	38
Table 4.3 Average Total Scores by Race and Language Fluency _____	39
Table 4.4 Percent of Items Screened Out by Various Statistical Criteria _____	42
Table 4.5 Coefficient Alpha Reliability Estimates for each Field Test Form _____	42
Table 4.6 Percent of Simulated Examinees Scoring at Different Levels For each ELA and Math Pseudo-Form _____	45
Table 4.7 Distribution of Observed Scores from a Single Testing for Students with Different True Scores (Expected Average Across Parallel Forms) _____	47

## List of Figures

<i>Figure 3.1 Five theoretical relationships between test items and their designated content.</i>	29
<i>Figure 4.1 Expected number correct for each simulated examinee on each ELA pseudo-form.</i>	44
<i>Figure 4.2 Expected number correct for each simulated examinee on each math pseudo-form.</i>	44
<i>Figure 4.3a Percent Scoring Below 50 for each True Achievement Level: ELA</i>	46
<i>Figure 4.3b Percent Scoring Below 50 for each True Achievement Level: Math</i>	46
<i>Figure 5.1a Percentage of principals reporting activities already underway to prepare students for the HSEE.</i>	56
<i>Figure 5.1b Percentage of teachers reporting activities already underway to prepare students.</i>	56
<i>Figure 5.2a Percentage of principals reporting plans for remediation of students who do not pass the HSEE.</i>	58
<i>Figure 5.2b Percentage of teachers reporting remediation.</i>	58
<i>Figure 5.3a Principals' predicted impact of the HSEE on student motivation and parental involvement prior to taking the exam for the first time.</i>	60
<i>Figure 5.3b Teachers' predicted impact of the HSEE on student motivation and parental involvement prior to taking the exam for the first time.</i>	60
<i>Figure 5.4 Teachers' prediction of influence of the HSEE on instructional practices over time.</i>	62
<i>Figure A.1. Illustration of IRT concepts of examinee ability distribution and the probability of passing an item.</i>	2



## CHAPTER 1: INTRODUCTION

### The California High School Exit Examination

California has embarked on a new program to ensure that all students graduating from high school meet minimum standards for verbal and quantitative skills. Section 60850 of the California Education Code (EC60850) calls on the State Superintendent of Public Instruction to develop a High School Exit Examination (HSEE) to assess achievement of content standards for English Language Arts (ELA) and Mathematics set by the State Board of Education. The new examination is to be adopted by the Board by October 1, 2000 (EC60850a).

Following provisions in the legislation, a panel of teachers, principals, school board members, parents, and the general public was appointed by the Superintendent and approved by the Board. The HSEE Panel's primary responsibility is to ensure that the exam is aligned with the Board's rigorous content standards for ELA and mathematics (EC60850b). The panel will also consider and make recommendations on a range of test development and administration issues such as frequency of testing, accommodations for students with disabilities, and determination of passing levels.

The legislation specifies a number of steps to ensure that the HSEE will be of high quality. These include review by the Statewide Pupil Assessment Review Panel (see EC60614) and a field test to assess reliability, content validity, and freedom from bias (EC60850c). The exam must conform to testing standards and comply with Title VI of the Civil Rights Act (EC60850e). The legislation requires alignment with both the State's curriculum frameworks (labeled Instructional Validity) and with the content of textbooks used in California schools (labeled Curricular Validity) and must employ criterion-referenced scoring (EC60850f). The HSEE will be offered to students with exceptional needs (as defined in EC56026), with appropriate accommodations where necessary (EC60850g).

Beginning with the graduating Class of 2004 (who enter high school in fall 2000), all students will be required to pass the HSEE to receive a diploma from a public high school in California. School districts may implement additional exams or impose other requirements for high school graduation (ED60850h).

The HSEE will be the third statewide testing program for high schools students. The Standardized Testing and Reporting (STAR) Program, used in computing the academic performance index for elementary and secondary schools, is based on the same academic standards, although the HSEE is designed to assess only those STAR standards judged essential for high school graduation. A third program, the Golden State Examinations, is designed to assess more advanced achievement in particular high school-level courses.

### Goals for the New HSEE

The legislation specifies in detail the development of the HSEE and its use as a new graduation requirement. It does not, however, capture the debate and discussion, in Sacramento and throughout the state, that led to the new graduation requirement.

Presumably the primary goals of this new program are to identify students who are not developing skills that are essential for life after high school (e.g., college and employment) and to encourage districts to give these students the attention and resources needed to help them achieve these skills during their high school years.

The HSEE program may have other, secondary goals. Some Board and the HSEE panel members whom we interviewed suggested that the HSEE will help to ensure that districts provide students opportunities to learn the knowledge and skills specified in the State Curriculum Frameworks (California Department of Education, 1999d and 1999e). Another secondary goal might be to make a California high school diploma more useful for selection decisions made by colleges and employers.

### **The Independent Evaluation of the HSEE**

The enabling legislation also requires the Superintendent to contract for an independent evaluation, beginning in January 2000 (EC60855). Human Resources Research Organization (HumRRO) was selected as the evaluation contractor. The evaluation analyzes data from the field test and annual administrations of the HSEE and reports on trends in pupil performance and in pupil retention, graduation, dropout, and college attendance rates (EC60855a). Passing rates and each of the above trends are examined separately for English-language learners, students with exceptional needs, students qualifying for free/reduced lunch in Title I schools, and other groups identified by the evaluator as being differentially affected by the exam (EC60855b).

The legislation specifies that the evaluation include recommendations for improving the quality, fairness, validity, and reliability of the examination. The evaluator may propose revisions to the design, administration, scoring, processing, or use of the examination (EC60855c). This preliminary report, on the field test, is submitted to the California Department of Education (CDE) and to the State Board of Education (SBE), the state legislature, and the Governor. The report is intended to assist the SBE as it considers adoption of the HSEE. A supplemental report, containing additional analyses of the field test data and of teacher and principal survey data, will be completed prior to the SBE's September 2000 meeting. Subsequently, biennial evaluation reports will be submitted by February 1 of even-numbered years, beginning with 2002.

### **Focus of the Evaluation**

After reviewing the legislative mandate and holding discussions with CDE and SBE staff, HumRRO has organized the independent evaluations around three general topics:

- **Quality:** Does the exam provide an accurate and unbiased measure of the knowledge and skills specified in the curriculum standards to be tested by the exam?
- **Fairness:** Do all students have adequate notice of the new requirement, opportunity to learn the material covered by the exam, and opportunity to demonstrate what they have learned?
- **Consequences:** Do benefits from the exam, in terms of improved student achievement and other positive consequences, outweigh any negative consequences?

In this Year 1 report, the primary focus is on the quality of items developed for the HSEE. The Board has not yet approved the list of specific standards to be covered; no operational forms of the test have been assembled as yet; a number of decisions have yet to be made on administration and scoring procedures; and the passing level has not yet been set. No students have taken the exam under operational conditions. Consequently, it is premature to talk of consequences and we are limited in what we can say on issues of fairness.

## **Evaluation Plan**

Initial plans for conducting the evaluation were specified in our response to the Department's Request for Proposals. During this first year, these plans were updated in response to new and evolving information about plans for developing and implementing the HSEE. Revised evaluation plans were presented to the HSEE Panel in May and to the State Board at its June meeting. A separate report detailing information on the evaluation questions and data sources we will use in addressing these questions has been submitted to CDE (Wise, Hoffman, Harris, Sipes, & Ford, 2000). These plans are summarized briefly here as they relate to the Year 1 activities. Our evaluation plans involve reviewing and analyzing three types of information:

1. We will review plans, proposals, and reports of activities by the HSEE development contractors.
2. We will examine sources of data on all California students and schools, including, but not limited to, data from the HSEE field test and operational administrations.
3. We will gather more intensive data from a representative sample of districts and schools to understand and illustrate the link between the HSEE requirements and outcomes based statewide observations.

**Review of Contractor Plans and Reports.** Since no formal reports were available during this first year, we attended meetings and listened to presentations by the development contractor (AIR) and by CDE. We also monitored various presentations to the HSEE Panel and the State Board and had direct conversations with members of each of these groups.

**Statewide Data Sources.** As specified in the enabling legislation, a primary source of information for our evaluation will be data from the HSEE field test(s) and operational administrations. We have examined 1999 STAR results and will, of course, continue to monitor trends in STAR results over the course of the evaluation. Statewide data on retention and graduation rates and on college attendance will provide primary information on the key outcomes that the HSEE is expected to influence.

**District and School Sample.** In order to understand and document the consequences of the HSEE, we proposed a longitudinal study of a representative sample of high schools. We will collect a great deal of additional information from these schools, primarily through surveys and interviews, that will enable us to understand the impact that the HSEE will have on their programs. During the first year, we recruited teachers and curriculum experts from these districts to review test items and tell us if they covered knowledge and skill not covered for all students in their current curriculum. (See Chapter 3.) We also administered baseline surveys to teachers and principals in the participating schools as described in Chapter 5.

## Selection of Districts and Schools for Longitudinal Study

A brief description of the procedures used to select districts and schools for participation in our longitudinal study is provided here because the schools selected will play an important role as we track changes over the course of the evaluation. The detailed sampling design is described in a separate report. The first step in this process was to select 24 districts to represent California 10<sup>th</sup> grade students as closely as possible. The following stratification variables, listed in order of importance, were used to this end:

- Size (defined by the number of 10<sup>th</sup> graders in the STAR 1999 database). We defined four levels with boundaries set so that the total number of students in each level would be proportional to the number of schools to be sampled from that level.
  1. Very large (more than 10,000 10<sup>th</sup> graders),
  2. Large (2,011–10,000 10<sup>th</sup> graders),
  3. Medium (785–2010), and
  4. Small (less than 785).

There is one very large district, which must be included in the sample. We sampled 7 large districts, one for every 20,625 10<sup>th</sup> graders; 8 medium districts, one for every 15,468 students; and 8 small districts, one for every 10,312 students.

- Proportion of students designated as English-language learners (ELL) (based again on 10<sup>th</sup> grade students). Within each size level, we divided districts into high and low proportion ELL groups. We chose the dividing line so that half of the students in districts of the target size would be in each group. Dividing lines ranged from 3.8% ELL for small districts up to 18.3% for large districts. The one very large district automatically included in the sample, was not further divided.
- Expected Score Level (defined by the STAR Mathematics mean for Grade 10). Within each size level and percent ELL group, districts were divided into two subgroups on the basis of 1999 average scores on the STAR 10<sup>th</sup> grade mathematics assessment. Math was selected over reading to capture variation across districts in the placement of algebra and other key parts of the mathematics curriculum and because reading scores were already correlated with the percentage of ELL students in each district.

After including the one very large district, we put districts in the remaining 12 groups in order (3 size levels by 2 ELL levels by 2 score levels) by:

- Community Type, using a rural, suburban, inner-city classification from the National Center for Education Statistics (NCES) Common Core of Data; and
- Geographic Location, indicated by the ZIP Code of the district headquarters.

Within each of the 12 sampling cells, we selected districts “systematically” (taking every “nth” district, for example, every 10<sup>th</sup> one, so that they would be as evenly distributed by community type and geographic location as possible) with probabilities proportional to the number of 10<sup>th</sup> graders. Replacement districts were identified for all but the largest district,

but all of the districts in the main sample are participating so we did not have to use any of the replacement districts.

The second step was to identify a sample of schools within each of the participating districts. In many cases, the smaller districts have only a single high school. The general plan was to sample 4–5 schools from the larger districts, 3 schools from the medium-sized districts, and 1–2 schools from the smaller districts. We ended up sampling 84 schools, an average of 3.5 for each of the 24 districts.

### **Contents of the Year 1 Evaluation Report**

This Year 1 Evaluation Report covers activities performed on the independent evaluation through June 30, 2000. Chapters 2 through 4 of the report describe different activities conducted during Year 1 and present the results of these activities. The final chapter describes the main conclusions that we draw from these results and our recommendations based on these conclusions.

Chapter 2, Background, contains our review of literature on the experiences in other states where similar graduation requirements have been or are being imposed. Given the high stakes nature of the HSEE, we conducted this background review to assist California in identifying and anticipating possible legal challenges in the implementation process sufficiently early to allow for modifications or course corrections as needed. We then used the results of this review to develop and implement activities in the item review workshops and to survey principals and teachers.

Chapter 3, Item Rating Workshops, encompasses both the design and administration of workshop sessions conducted with California teachers and curriculum specialists as well as the analysis and results of this effort. We selected a representative sample of districts from across California that were not part of the field test sample, and asked them to take part in the workshops. The participants' primary task was to assess how well a sample of field test items covers target standards and to rate whether students in their schools have the opportunity to learn the skills being tested.

Chapter 4, Analyses of Field Test Data, presents our initial examination of data received from the field test, which was conducted in May. We conducted our own independent analyses of these data to determine whether the items developed to date have good statistical properties. The rate at which items survive statistical screens is an indicator of the quality and thoroughness of the item development process and also has a major bearing on whether it will be possible to construct test forms that completely cover the targeted content. In addition to examining "survival" rates, we begin a dialog on test score accuracy for forms that resemble the field test forms. The analyses reported in Chapter 4 are only preliminary as the constructed response items (essay questions) are still being scored at this time. We will issue a supplemental report by mid-August, covering analyses of additional data that become available.

Chapter 5, Principal and Teacher Surveys, describes the development of questionnaires intended to identify any preliminary efforts and concerns related to the pending graduation

exam requirement. The primary areas surveyed included knowledge of the test and other related documents, any preparation thus far, future plans, and expectations of and for the test. Surveys were provided to the same representative sample of districts that were asked to participate in the item review workshops. We shipped survey packets in early May. Preliminary results for surveys returned to date are included, but, again, the more complete analysis and results will be reflected in the Supplemental Report.

Chapter 6 presents our Conclusions and Recommendations based on the existing state of data analyses and results.

## CHAPTER 2: BACKGROUND

### Prevalence and Nature of High School Graduation Testing

Testing is being used by schools and states for a wide range of purposes from diagnosing student deficiencies to holding schools accountable for student achievement. The focus of this paper is on high school exit exams.

High school exit examinations present considerable issues for state policymakers because of their “high stakes” nature. Important consequences are attached to test results, and an exit exam is high stakes in nature because award of a diploma is based on passing the test (NGA Reports Online, 1998). In fact, *High Stakes* is the leading title used by the National Research Council’s Committee on Appropriate Test Use to report on the Congressionally-mandated study on uses of testing for tracking, promotion, and graduation (National Research Council, 1999). The courts have held that a high school diploma is a property interest, and as such, protected by the Fourteenth Amendment. As a result, implementation of an exit exam is likely to present state policymakers and assessment staff with technical and legal challenges (National Research Council, 1999; NGA Reports Online, 1998).

According to the Council of Chief State School Officers’ (CCSSO) annual survey of state student assessment programs, eighteen states had implemented high school exit examinations as of the 1997–98 school year (Olson, Bond & Andrews, 1999). Table 2.1 lists the states and indicates their responses to a survey question asking to what instructional uses these assessment results were put.

The number of states with such a requirement is on the rise; NGA Reports Online (1998) identifies twenty. These range from Florida, which first implemented its test in the 1977–78 school year (then delayed until the 90s as significant legal challenges occurred), to Alaska, whose exam requirement becomes effective in 2000 (NGA Reports Online, 1998). In most of these states, students must pass the examination in order to graduate from high school; other states offer an honors or endorsed diploma based upon test results (Olson et al., 1999; NGA Reports Online, 1998). Passing the test is typically linked with completing proscribed coursework; that is, a student must complete all required coursework and pass the exit examination in order to receive a high school diploma.

Half of the participating states test students on material that is typically taught in 9<sup>th</sup> grade or earlier; these tests can be regarded as minimum competency tests (National Research Council, 1999). Other states have implemented tests containing more advanced material. There is, however, a degree of tension between minimum competency or basic skills approaches and the more recent push for all students to learn at higher levels. Many states initiated state standards after implementing a graduation test, so policymakers need to know that there may be problems of alignment between the content of the exam and the content of the standards (NGA Reports Online, 1998). Even with this caution in mind, Boser (2000) reported preliminary results from a University of Wisconsin study of 10 states that claims there may be little overlap between what state assessments test and what teachers teach—with a range from 5% to 46%. This adds to the existing concerns about standards-based accountability; although teachers are likely to be teaching what they think is appropriate,

their individual lesson plans may not be aligned with state standards or assessments (Boser, 2000).

**Table 2.1. States with high school exit examinations as of 1997–98.**

State <sup>1</sup>	Implementation Dates	Instructional Uses					Program evaluation
		Identification of students at risk	Student diagnosis	Individual student instructional planning	Improvement of instruction for groups of students	Curriculum planning at school/district	
Alabama	March 1999; Class of 2001		✓	✓			
Florida	October 1993; Class of 1995	✓		✓	✓	✓	
Georgia	Fall 1993; Class of 1995	✓	✓	✓	✓	✓	
Hawaii	Spring 1979; Class of 1983	✓	✓	✓	✓	✓	
Indiana	September 1997; Class of 2000	✓	✓	✓	✓	✓	✓
Louisiana	March 2001 & 2002; Class of 2003		✓	✓		✓	
Maryland	1978–79; Class of 1982 <sup>2</sup>	✓	✓	✓			
Mississippi	Fall 1987; Class of 1989						
Nevada	April 1998; Class of 1999	✓	✓	✓		✓	
New Jersey	October 1992; Class of 1994	✓	✓	✓	✓	✓	✓
New Mexico	February 1988; Class of 1990						
New York	1987–88 <sup>3</sup>						
North Carolina	Summer 1994; Class of 1998	✓			✓	✓	✓
Ohio	Fall 1990; Class of 1994				✓	✓	
South Carolina	Spring 1986; Class of 1990 <sup>4</sup>		✓		✓		✓
Tennessee	Phasing out	✓	✓	✓	✓		
Texas	2002–2003	✓	✓	✓	✓	✓	✓
Virginia	Spring 1998; Class of 2004	✓	✓	✓	✓	✓	✓

Policymakers should anticipate that the approach that calls for more rigorous assessments is likely to produce higher failure rates, particularly among educationally disadvantaged

<sup>1</sup> The District of Columbia plans to implement an exit exam starting with the Class of 2003; and Massachusetts' plan starts with the Class of 2003, which must pass 10<sup>th</sup> grade assessment (Argetsinger, 2000).

<sup>2</sup> Maryland functional Testing Program is currently being phased out, to be replaced by a new high school assessment program in January 2002, Class of 2005. However, Maryland recently voted to delay its high school exit exam by at least two years; the decision means the Class of 2007 is likely to be the first to face the test as a graduation requirement (Argetsinger, 2000).

<sup>3</sup> New York's Regents Competency Tests are currently being phased out in favor of Regents Examinations as of June 1999, Class of 2005.

<sup>4</sup> South Carolina's Basic Skills Assessment Program (BASP) is being replaced by Palmetto Achievement Challenge Tests (PACT), Spring 2001, Class of 2003.



students. High failure rates, in turn, may be expected to increase public concern and the likelihood of legal challenges. The Maryland decision to slow the development and implementation timeline of its assessment program points out the public and political concerns that can arise over the consequences of testing rigor (NGA Reports Online, 1998). The Maryland State Board of Education has voted to delay the graduation test requirement by at least 2 years because of added concerns that the state needs to do more to help children prepare for the challenge (Argetsinger, 2000; Solida, 2000). This parallels a caution in the National Academy of Education report (McLaughlin, Shepard, & O'Day, 1995) that before all students are expected to perform to world-class standards, they must be provided a world-class curriculum and instruction. This is likely to be a particular concern in California, where the high standards reflected on the graduation test currently are a challenge to the state's lack of authority to tell districts what courses to teach, and the low per-pupil expenditures are a challenge to upgrading and aligning curriculum and instruction to a level sufficient to meet the expectation the test creates.

For minority children, English-language learners, and students with disabilities, the stakes in appropriate, nondiscriminatory test use are particularly high. Such children tend to do less well on large-scale tests, even where such tests measure only basic skills. In the late 1970s, for example, when minimum competency tests gained popularity, 20% of African American students, compared with 2% of White students, failed Florida's graduation tests and were denied high school diplomas (*Debra P. v. Turlington*, 1979). Similarly, 1998 data from the Texas graduation tests show cumulative failure rates of 17.6% for African American students and 17.4% for Hispanic students, compared with 6.7% for White students (Natriello & Pallas, 1999). For students with disabilities, most of whom now participate in large-scale state assessments, recent data from 14 states consistently show failure rates that are 35–40 percentage points higher than those for nondisabled students (Ysseldyke et al., 1998).

Moreover, states increasingly are raising the standards on their graduation tests to “world-class” levels, such as those embodied in the National Assessment of Educational Progress (NAEP). Based on 1996 NAEP data, 40% of all students would fail tests that reflect world-class” standards, at least initially, and failure rates for minority students would be about 80%. These predictions are consistent with recent data from Massachusetts and New York, where students have begun taking state tests that reflect world-class standards. On such tests, failure rates for students with disabilities would likely be in the range of 75 to 80%.

Regardless of whether the exam content addresses basic skills or more challenging “world-class” levels, statewide high school exit exams are intended to provide a consistent, objective measure of student learning throughout the state. The most commonly tested subject areas are language arts (English reading and/or writing) and mathematics; some states also test science and social studies (NGA Reports Online, 1998). The most common item format in large-scale assessment programs is multiple choice. Most states use multiple choice in conjunction with other assessment types such as extended constructed response, short answer, and performance items (Olson et al, 1999). All of the states with exit exams allow students to retake the test multiple times.

The Individuals with Disabilities Act (IDEA) and the Improving America's Schools Act (IASA, 1994) demand that all students be included in state assessments and that all students be offered equal educational opportunities and be held to the same high standards. According to the IASA, Title I programs must have in place final assessment systems that include all students and provide disaggregated (by component categories such as gender and race/ethnicity) test scores by school year 2000–2001.

Policymakers clearly are going to have to address such requirements and concerns, as well as deal with a number of other policy issues. One concern is that graduation tests may run counter to motivating students, especially high achieving ones. The fear is that students will view the “minimum” level of knowledge and skills as the “maximum” they need to master in high school. Another issue is timing of the test. When administered initially in 8<sup>th</sup> or 9<sup>th</sup> grade, the content cannot include material taught at later grades. Thus the test cannot be as rigorous as if it were to be administered later. However, when tests are administered later, students who do not pass have limited time for remediation. Some states have established “early warning tests” to serve as alerts of any student learning deficiencies. Remediation itself is probably one of the greatest challenges for states. Policymakers, who have a responsibility in overseeing all aspects of the testing program, need to make sure that all appropriate steps are taken to help students who fail pass the test (NGA Reports Online, 1998).

### **From Testing to Higher Achievement**

For proponents of such testing, high school exit exams have an immediate goal: to raise the standard by which a high school diploma is awarded. A long-term implication of this higher standard would be greater achievement for students currently below this level. Consistent with this implicit goal, the 1999 Council of Chief State School Officers' (CCSSO; Olson et al., 1999) survey reports that the most common reason cited by states for their various assessments, not limited to the high school exit exam, was to improve instruction. School accountability, student accountability, and staff accountability were cited with far less frequency.

There are two routes by which a high school exit exam may impact student achievement. First, student motivation may be affected. Second, instruction can be modified, presumably for the better.

The effect of a high stakes exam on student motivation is a double-edged sword. The beneficial aspect is that student motivation may increase when students and parents realize that the exam must be passed in order to earn a high school diploma. High achievers may increase effort in order to pass the exam as soon as possible, while students who fail the exam on the first attempt may see this as a wake-up call and try harder. On the other hand, some argue that the exam may be a demotivator for students. They warn that high-achieving students who are able to pass the exam early may lose motivation to continue to strive for further learning. Since many states tie their exam contents to 9<sup>th</sup> grade curricula (e.g., Algebra), this is a viable concern. Further, they claim that students who fail the exam may become disillusioned with education and their abilities (CCSSO; Olson et al., 1999).

The link between mandatory testing and improved instruction, and by implication, higher achievement, is not a direct one. By themselves, tests improve education no more than a thermometer reduces fever—the existence of a test ensures nothing other than that the level of achievement will be measured. In order for a high stakes statewide testing program to foster a higher level of student achievement, several intervening connections must be in place. Like the links in a chain, each connection is an integral part of a fully-functioning system.

For example, if test results are to accurately reflect achievement, the test must meet standards of appropriate test use (e.g., validity, psychometrics, content). In order for test results to subsequently affect achievement, test scores, or the framework used in building the test, must provide information that guides instructional changes; therefore, the test should lead to the identification of specific areas of instruction that need shoring up. Further, in order for this diagnostic information to be put to use, the score reports must be interpretable. Even more fundamentally, the relevant parties must be aware of the information available in the score reports. Not only must test results be diagnostic, understandable, and publicized, but educators must be able to determine what the results imply for current practice and enjoin for future practice. At a broader level, the aggregation of test results across schools, districts, and student demographic groups should provide useful information regarding equity of successful instruction. Tools and procedures to monitor and rectify any inequities must be in place before any high stakes student impact is enacted.

The following sections address each of these links between high stakes testing and improved instruction.

### **Standards of Appropriate Test Use**

The National Research Council's committee on high stakes testing relied upon three broad principles of appropriate test use (National Research Council, 1999). First, the test should have measurement validity (i.e., be valid for its stated purpose), defensible test psychometrics, and appropriate content. In order for an examination score to be a mandatory component of a high school diploma, the content must be included in curricula throughout the state. Any inconsistency opens the door for legal action from students claiming not to have been taught the requisite material. Before or after the fact, the presence of a mandatory statewide exam presupposes commonality among all schools throughout the state. If not already present, one of the consequences of such an exam is the alignment of curricula with exam contents. The *Debra P. v. Turlington* case (1981) established, among other things, that graduation test contents must reflect what students have been taught. In reporting on a forthcoming study from the University of Wisconsin, Boser (2000) reiterates the concern on basic premise of implementation: "States should delineate what students should know and be able to do, teachers should match instruction to those standards, and state tests should measure how well students meet those expectations." However, the Wisconsin researchers found some criticism of states for not doing enough to explain to teachers what is on the exam, given the penalties and rewards systems that are increasingly accompanying the tests (Boser, 2000). In California this is likely to be exacerbated by the state's limited authority to tell districts what to teach.

Second, test users must make valid attributions of cause. Poor scores can be the result of students simply failing to learn what they've been taught, or confounded by other factors such as disabilities or language barriers that are not related to the construct, or a denial of opportunity to learn.

Third, the test should address consequences (consequential validity). In other words, the test should result in better education.

### **Diagnostic Information**

The specificity of a test score impacts the utility of that score. For example, a low score in mathematics may be the result of any of a number of specific deficits. If only a single mathematics score is reported, this may not provide sufficient detail to instigate appropriate remedial measures. In the absence of diagnostic scores for individual students, school faculty, parents, and students are left guessing as to what material or knowledge is lacking in a less-than-satisfactory score. Not only might this result in inappropriate or unnecessary remedial instruction for the individual student, but it might also fail to illuminate school- or districtwide curriculum gaps, as well as differential results for particular groups of students (e.g., minority students, low-socioeconomic (SES) students, English-language learners, students with disabilities). Without diagnostic-level information, an unintended consequence may be the ill-informed adoption of ineffective “quick fixes” to raise scores.

This is an issue in California as discussion continues regarding whether the exit examination is intended as a certification exam or as a basis for remediation. If the plan is to allow students four times to pass the test, it would seem that the year between each testing window is intended for students and teachers to work on the problem areas to increase the likelihood of passing in the future. If the time between test administrations is not meant for teaching and learning to promote passing, but rather only for certification, then this process must be carefully and clearly explained.

The specificity of reliable assessment stipulates test length, administration time, and cost, among other things. These factors may preclude the generation of a full level of diagnostic detail by the statewide exit exam. For practical purposes, the exit exam score may be limited to aggregate levels (i.e., pass/fail math). If so, other sources of information (e.g., other test scores, teacher evaluation) may be recruited at the schoolhouse level to augment the score for diagnostic purposes.

### **Meaningful Score Reports**

In order to be put to good use, the score reports must be clear to educators. At both the individual student level and the aggregate classroom/schoolhouse level, teachers must be able to understand the implications of the reports. This requires, at a minimum, a well-designed report; it may additionally call for professional development training and score interpretation guides.

### **Awareness and Communication**

An easily overlooked aspect of a new testing system is the propagation of information, including counsel about what data are available. For example, if a school-level diagnostic

report is provided along with individual student results, educators must be aware of the information provided by the school-level report and how to use it, in order for it to be utilized effectively. If data and tools are available to permit schools/districts to aggregate data at various levels, personnel must be trained and procedures put in place to facilitate their use.

The 1998 CCSSO survey asked of each state, “What kinds of [successful] materials has your state or others in your state developed for assessment program publicity, explanation, or training?” Frequent responses included training manuals, newspaper inserts, brochures, web site information, videos, public access television, and interpretive guides. Most states also reported various forms of professional development to educate teachers and other educational staff on the statewide assessment program. These ranged from paper documents to hands-on workshops (Olson et al, 1999).

### **Identification of Appropriate Instructional Changes**

Raising awareness about and understanding of test results are necessary steps toward eliciting appropriate instructional changes, but they are not enough. School personnel must be able to determine what various patterns of results imply for instruction, such as which instructional units are succeeding and which are falling short. Areas needing modification may include changes such as additional instructional time for certain concepts, inclusion of additional concepts, or altered pedagogical approaches. These analyses should be conducted at the individual level, as well as at various aggregate levels (e.g., English-language learners, students with disabilities, schoolwide, districtwide).

### **Implementation of Appropriate Instructional Changes**

The application of test results to appropriate instructional changes relies upon teacher training and availability of adequate resources in a sufficient timeframe. Teachers must know what adjustments to make, how to make them, and have the necessary time, money, and facilities to implement those changes. This is a two-pronged process: remediation of students who failed the exam already, and modification of first-time instruction to students who have not yet taken the exam.

In 1994, 7 of the 18 states with graduation examinations provided funds to schools or districts specifically for remedial education (Bond & King, 1995). Another approach to remediation is to provide additional educational material directly to students in need. For the Texas Assessment of Academic Skills (TAAS), the state provides *TAAS Study Guides* free of charge to school districts, concurrent with test results. These are distributed to students who fail any parts of the TAAS. Connecticut, Kentucky, and Maryland are other states that have spent heavily to bring teaching and learning into line with state standards.

Modification of first-time instruction presumes feedback to the teachers of subjects being tested. For example, if the students in a particular high school have difficulty with the Algebra portion of the exam, feedback should be provided to the middle schools that feed that high school, as some of the students would have completed Algebra in the 8<sup>th</sup> grade. Algebra instruction may require modification at both the middle and high school levels. In order for this to happen, there must be a feedback mechanism within the high school, as well as from the high school to its feeder schools.

Modification of instruction to enhance test scores brings with it short- and long-term consequences, some good and some bad (Hess & Brigham, 2000). Long-term benefits include increased equity, clarity, and focus in curriculum, and an efficient use of resources. These benefits come hand in hand with consequences such as limited local decision-making and a narrowed curriculum. Despite the long-term equity advantages, in the short term there are negative consequences for poor, minority, and special needs students. Chudowsky and Behuniak (1998) also stress the negative consequences of a narrowed curriculum, as well as the loss of instructional time to preparation and administration of the test. Their teacher focus groups also revealed classroom benefits such as the availability of test practice materials, new instructional approaches, and the purchase of more lab equipment and textbooks.

### **Anticipated Challenges to California's High School Exit Exam**

The history of public education in the United States is marked by numerous legal challenges. The U.S. Constitution, federal civil rights statutes, and other judicial decisions have been invoked to determine whether specific tests are discriminatory or otherwise inappropriate. In recent years, high stakes testing programs have faced complaints of intentional discrimination, carryover effects of prior discrimination, and disparate impact upon various demographic groups, including English-language learners. In addition to discriminatory claims, other due-process issues have included insufficient notice of the test requirements and lack of opportunity to learn the test content. These allegations have been put to the legal test in several states, including Alabama, California, Florida, Georgia, Illinois, and New York—with varying outcomes.

The requirement that a student pass an exit examination in order to receive a high school diploma is perhaps one of the highest stakes for which an exam can be used. The importance of a high school diploma practically guarantees that such a requirement will be tested in the courts. In fact, several states have preceded California in the implementation of high school exit examinations and their experiences can provide some insight as to the likely legal challenges that can be anticipated.

The decision in the *Debra P. v. Turlington* (1981) case in Florida established three standards that high school exit examinations must meet to stand up to constitutional scrutiny:

1. The test must measure knowledge and skills that are taught in the state's schools (referred to as "curricular validity")
2. Students must receive adequate notice of the test, the requirements for passing, and the consequences of not passing; and
3. The test must not intentionally discriminate against a protected group or class.

These areas also provide guidance for the bases of potential challenges, which include claims of inadequate content coverage, misalignment of curriculum to standards, insufficient notice, adverse racial/ethnic impact, and lack of accommodations for students with special needs and English-language learners.

## Students Who Have Not Been Taught the Knowledge and Skills Measured By the Test

Clearly, examination content should be representative of what students have been taught. Alignment of curriculum standards and instruction to the examination is an essential first step in preparing for an exit exam. Curricular validity is considered to mean consistency between the test content and both what is found in the curricular materials and what is being taught in classrooms (NGA Reports Online, 1998). In *Debra P. v. Turlington* (1981), Florida demonstrated curricular validity through an outside study that surveyed curricular materials, teachers, district educators, and students and showed that the content on the exam was being taught in the classrooms.

California's *Guiding Principles for the High School Exit Examination Standards Panel* (California Department of Education, 1999a) describes the responsibility of school districts to prepare students to succeed. Alignment is the first item, indicating that California is aware of experiences in other states (e.g., Florida and Texas). The changes that must occur in the content and teaching methods require time. Thus, states need to provide for a sufficient timeframe between the introduction of the new test and the high stakes decisions to allow for the necessary alignment (suggested, for example, by Bond & King, 1995). This time period needs to be several years during which the districts and schools can work out and implement approaches that target fully preparing students for the examination.

The other prevalent concern is providing students with the opportunity to learn (OTL) what is tested. In *Goals 2000* (1994; 1996), "OTL standards" are defined as the "criteria for, and the basis of assessing the sufficiency or quality of the resources, practices, and conditions necessary at each level of the education system to provide all students with the opportunity to learn the material in voluntary national content standards or state content standards" (§3(a)(7)). The recommendation at the voluntary national level includes the following:

- Curricula, instructional materials, and technologies
- Teacher capability
- Continuous professional development
- Alignment of curriculum, instructional practices, and assessments with content standards
- Safety and security of the learning environment
- Non-discriminatory policies, curricula, and instructional practices
- Other factors that help students receive a fair opportunity to achieve the knowledge and skills in the content standards.

In efforts that have followed, the Joint Standards Panel (AERA, APA, & NCME, 1999) asserts that promotion and graduation tests should cover only the "content and skills that students have had an opportunity to learn" (146, Standard 13.5). *High Stakes* (National Research Council, 1999) recommends that "tests should be used for high stakes decisions about individual mastery only after implementing changes in teaching and curriculum that ensure that students have been taught the knowledge and skills on which they will be tested."

In addition to the overall implications of OTL, states must consider what this means for students with disabilities and English-language learners (ELL). The 1997 National Research Council study, *Educating One and All* edited by McDonnell, McLaughlin, and Morison, and *High Stakes* edited by Heubert and Hauser (National Research Council, 1999), both state that if students with disabilities are required to pass state graduation tests, then educators must modify the students' IEPs so that the students get taught the relevant knowledge and skills. This is no small task, and a state needs to allow sufficient time to review and modify all its IEPs. For English-language learners, having been taught the relevant knowledge and skills takes on a different meaning when the test is in English. OTL for ELL means that these students have been given the chance to acquire not only the relevant subject matter but also the necessary levels of English proficiency. Again, a state must factor in adequate time for meeting these students' need. Other issues must be considered alongside alignment and OTL. One of these is equal educational opportunity. Even if alignment of the content standards, instruction, and test content are in place and OTL actions were well established, a possible area of challenge to the examination is "adverse impact" based on claims of differential access to test preparation materials, substantial variations in setting and physical conditions during testing, and major disparities in quality of instruction (National Research Council, 1999). Questions about adverse impact should be included in validity studies, and the department of education, as the test user, should implement an evaluation component to keep track of the intended as well as unintended consequences of the examination (Messick, 1989).

Another important consideration related to the appropriateness of the content skills measured by the test is that it does not include items that are irrelevant to a construct or that underrepresent a construct. The issue is one of giving an unfair advantage or disadvantage to any particular subgroup. Adverse impact is a particularly relevant concern for subgroups such as minority students, students with disabilities, and English-language learners. These students will be at a disadvantage if the test measures facts or concepts that are not relevant to the content. As a guide to addressing these issues, Messick (1989) maintains that identification of either of the practices signals a validity problem, but absence of such identification indicates a policy problem. Thus, the test developer should provide evidence of the appropriateness of the content and cognitive processes to be measured. The department of education, as the test user, should include this in its evaluation component.

One of the biggest challenges for states is helping students who do not pass the graduation test, or who are in danger of not passing. Some states give funds to local school districts for remediation, which most often includes tutoring and extended learning time. Some states also fund summer school programs to assist students (NGA Reports Online, 1998). California has an initial statement related to this challenge in its *Guiding Principles for the High School Exit Examination Standards Panel* (California Department of Education, 1999a), where it posits that districts must provide alternatives for low-performing students, to include the reduction of electives to permit supplemental instruction during the academic year.



## Insufficient Notice and Implementation

As a result of legal challenges in several states, the necessity to provide adequate advanced notice of the high stakes requirement of the test has become a fairly straightforward condition that all states now know must be an integral part of any graduation exam planning. In the case of the decision in *Debra P. v. Turlington* (1981), the court deemed four years as “sufficient notice.” The basic language emphasizes that both students and school personnel should receive clear indications of the content and performance for which they will be held accountable. They also should be provided with general scoring guidelines and examples of proficiency in a standard. Other related materials (e.g., curricular frameworks, sample tasks, assessment specifications, and model responses) may be helpful in conveying expectations (National Research Council, 1999).

California has included such language in its law, requiring districts to provide early and adequate notice to students and parents about the test requirement. Further, the test becomes a requirement for graduation in the 2003–04 school year. The intent is that districts will be able to align curriculum and instruction to the state standards in all grades so students will be prepared to pass the exam prior to graduation (“SB 2X,” 1999).

In addition to covering the advance notification requirement, the *High Stakes* report (National Research Council, 1999) stresses that states should take a number of steps before, during, and after test development to help ensure the appropriateness of the content and skills being measured. It recommends that the domain to be assessed on the graduation test should be clearly and widely publicized. These are important considerations that seem related to “notice,” although they may not have been targets of direct challenges. The law is interested in whether tests are used inappropriately for student promotion, tracking, or graduation, and it is imperative that states be unambiguous about the purpose (National Research Council, 1999). States must fully and clearly define the function of an assessment and allow adequate time and resources to work out technical problems, give the test a fair trial, and gain support for it (Smith, Heinecke, & Noble, 1999c).

According to the policy researchers involved with the Arizona Student Assessment Program (ASAP), Arizona’s experiences provide an interesting story, and perhaps guiding lessons, about the political and educational interplay on high stakes testing and the need to explicitly describe the function of an assessment (Smith, Heinecke, & Noble, 1999b). On the surface—that is, the official descriptions of the program and each of its elements at its inception in 1990—ASAP seemed clearly delineated. ASAP’s original aims were grounded in educational reform, which involved using the state’s existing standardized testing more as a diagnostic tool rather than as an accountability measure and altering instructional practices in classrooms. However, the legislative language did not mention ASAP, any particular form of testing (e.g., performance, authentic, alternative), or any principles of practice (Smith, Heinecke, & Noble, 1999d). According to the policy researchers, what happened is that the Legislature heard “accountability” (although this was minimized publicly), and the Department of Education heard “instructional reform.” No funds were committed for teacher training in the reform principles and practices, so generally only school districts with an interest in this new direction and with sufficient wealth undertook staff development for ASAP. Consequently, the program moved away from a reform premise to a legislative one, and key legislators pressed the Department to start producing data for accountability

purposes. This set in motion a very constrained timeline that did not allow for building the capacity of schools and teachers, developing sound psychometric instruments, ironing out administrative wrinkles, conducting a study that compared ASAP with prior testing programs, or establishing an independent evaluation of the performance assessments (Smith, Heinecke, & Noble, 1999d). Thus, ASAP was viewed differently by various groups, who formed their perspectives only from direct experience. This conceptual confusion led to a range of viewpoints, including seeing as a set of performance tests, a measure of how students learn, a set of reform ideals, an authentic assessment, integrated learning, and a way out of the problems of high stakes standardized testing. Awareness of the accountability aspects of ASAP arose later (Smith, Heinecke, & Noble, 1999c), and the high stakes function was revealed fully when the state newspaper published the 1993 results by school and grade level and ranked them similarly to the way it had reported standardized test results (Smith, Heinecke, & Noble, 1999d). The ambiguity of the purpose of the ASAP and the multiple meanings led to the suspension of the test in 1995, just a week before it was scheduled to be administered (Smith, Heinecke, & Noble, 1999a).

### Students of Color

Adverse impact is at the heart of challenges under civil rights laws, and was the basis for a case against the Texas Assessment of Academic Skills (TAAS) (*G.I. Forum and Image de Tejas v. Texas Education Agency*, 1997). A large gap exists in pass rates between white and minority students on the TAAS (Natriello & Pallas, 1999). Graduation rates for all students dropped precipitously after implementation of the TAAS. During the 8 years of its use, however, the White graduation rate has returned to its previous level, while graduation rates for Latino and African American students have not (“The case,” 1999; Shannon, 1999). The judge recently ruled in favor of the defendants in this case, concluding that the state had demonstrated “educational necessity” (“Court rules,” 2000; “Decision,” 2000). Although the judge did not dispute plaintiffs’ factual claims and acknowledged that the TAAS did have legally meaningful disparate impact against African American and Latino students, he concluded that the high stakes test approach is the only way to force improved outcomes in Texas education.

Although it is difficult to argue with the notion of school accountability for students’ learning of basic skills for graduation, the test scores often take on such importance that negative practices may occur. In efforts to achieve higher scores, schools may retain students who are having difficulty in the grade level preceding the target test year. While the intended consequence is to provide the students with an added year to build their skills, the unintended consequence is that these students may be more inclined to drop out, which reduces the number of students in the test-taking pool and, consequently, the number who are likely to fail the test (“The case,” 1999). Important considerations are setting the passing score based on accepted measurement procedures and using multiple measures to determine graduation (National Research Council, 1999; American Educational Research Association et al., 1985:54, Standard 8.12).

Natriello and Pallas (1999), who suggest that the growth of testing programs may be viewed from a social perspective as a control on awarding high school diplomas and on the related upward class movement of the disadvantaged, state, “the current testing boom may be

viewed both as an attempt to control the educational bureaucracies and to slow the rate at which new groups enter positions of power and influence in U.S. society.” (p. 3) The essence of this problem is captured by Richard Jaeger, who testified in a North Carolina case that grade retention does not lead to improved learning but does increase the dropout rate (“N.C. lawsuit,” 1997). Whether such a social view is acknowledged, there certainly is evidence that the recommendations of the Joint Standards Panel (AERA, APA, & NCME, 1999) are not always followed when tests are used in decisions about tracking (Darling-Hammond, 1991; Shepard, 1998), promotion (Shepard, 1991), and graduation (Stake, 1998). Such validity problems are a civil rights issue for minority students in particular, who are disproportionately assigned to lower track courses, retained in grade, and denied diplomas on the basis of test scores (Heubert, 2000). Over and over, these effects have been shown to have greater impact on minority students of color. Ultimately, the strongest practice to inoculate against such effects is to provide minority, and all, students with a high-quality education.

Actual charges of racial bias in graduation tests often focus on the method used for question or item selection. In the case of the TAAS, a type of correlation is used that typically selects questions (or items) with the highest correlations (“Racial bias,” 2000). These are items with the greatest differences between high and low scorers. This method is frequently used to increase consistency and reliability. An unintended consequence is that this approach is likely to increase bias against minority students because, as a group, they do less well on the test overall.

When consistency is a primary factor used to screen test questions (or items), it is often the middle-to-upper-class experiences that are reflected, and certain other social and cultural perspectives are often eliminated (“Racial bias,” 2000). One alternative is to select items via a different procedure such as an approach that reduces bias by using the smallest racial gaps from a particular subject area. This is more complex, however, when there are more than two racial groups to balance. Another possibility is to reject the premise of “unidimensionality” of this type of correlational selection. This assumes that test-takers use only one cognitive approach to solve a problem rather than multiple ways of thinking. If the basis of test development “assumes” unidimensionality but instruction “allows” for multi-cultural approaches there likely is cultural bias in the test (“Racial bias,” 2000).

Score differentials between white and non-white students have come to be expected as a challenge in assessment programs. On most achievement and attainment indicators there are gaps. The usual, and most obvious, explanations point to inequalities in income and school resources. States are concerned about how to best address the difficulties without raising other charges leveled at graduation tests such as that they encourage “teaching to the test,” which, in turn, is considered to have detrimental effects on disadvantaged and high minority schools and districts. An interesting direction in this area may be offered by Bempechat (1998), whose five-year study examined successful students to identify actions schools and families can take to help at-risk students reach their full potential. The factors behind the success of at-risk students are related to their beliefs about effort and ability, which form a fundamental foundation for academic success. Her study found that that “the overall pedagogical philosophy of Catholic schools is one that leads children to believe in their intellectual abilities and strive for academic excellence.” This benefit was noted particularly

in increased academic success of African American and Latino students in Catholic schools. The key issue for educators is to identify strategies that will positively affect the way children think about success and failure (Bempechat, 1998).

## **Students with Disabilities**

Federal and state laws now mandate inclusion of all students, including those with disabilities and English-language learners, in large-scale assessments (National Research Council, 1999). *Goals 2000, Educate America Act* (1994; 1996), Title I of IASA (1994), Individuals with Disabilities Act (IDEA, 1988, 1998), Section 504 of the Rehabilitation Act (1973), and Title II of the Americans with Disabilities Act (ADA, 1990) influence how students with disabilities are educated as well as included in assessment programs. They also establish legal rights for special needs students that can affect exit examinations.

Traditionally, students with disabilities have not participated in large-scale assessments. Parents' and educators' concerns have ranged from the stress of testing, confusion about modifications and accommodations, and the lowering of schools' average scores, to the possible mismatch between test content and curricula and difficulties in administering tests to students with severe disabilities (National Research Council, 1999). However, the importance of a high school diploma to one's future in our society has become increasingly clear and points to the need for more research to address the divided camps on the practice of issuing an alternative or modified diploma or certificate of completion to students with disabilities. Evidence is needed to help sort out charges of stigmatization to the students on the one hand and devaluation of the credential on the other (National Research Council, 1997).

Parents and educators of special needs students have indicated that they want special needs students to meet the same high standards established for general students (Ysseldyke et al., 1998). Given a greater push toward this goal, states will need to consider even more closely the "sufficient notice" component in their planning for graduation tests. As they struggle with testing systems that accommodate students, educators are recognizing that preparing for and establishing the changes to a student's individualized education program (IEP) that will provide the necessary support for inclusion in a state testing program, requires even more time than implementing other types of support programs. The Indiana Civil Liberties Union filed a suit in 1998 alleging that the state provided neither sufficient notice to students with disabilities of the requirement for graduation nor adequate accommodations in administering the tests to special needs students (Solida, 2000).

Nevertheless, with current graduation exams, as with most state assessment programs, students receiving special education services frequently may be exempted from the test. For example, in Texas, the student's special education committee determines if the TAAS or end-of-course examinations are appropriate measures of the student's progress. Students who are not exempt based on their special education status (e.g., learning disabilities such as dyslexia) may be eligible for testing accommodations (Natriello & Pallas, 1999). In New York, students with disabilities also may be exempted from the Regents Examinations, or may be eligible for testing modifications. In Minnesota, students with disabilities are eligible for accommodations or modifications when such specification is included in the individualized

education program (IEP) or 504 plans. Modifications that involve adjustment in the test that actually changes the standard result in a notation on the transcript that indicates “Pass-Individual” rather than “Pass-State” (Natriello & Pallas, 1999).

Among the 12 recommendations made in the study group report to the National Research Council, *Educating One and All*, (National Research Council, 1997; pp. 197–210), the following seem particularly worthy of consideration in California:

- States and localities that decide to implement standards-based reform should design their common content standards, performance standards, and assessments to maximize participation of students with disabilities.
- The presumption should be that students with disabilities will participate, but an individual student may require alterations; such decisions must have a compelling educational justification.
- States and localities should revise policies that discourage maximum participation of students with disabilities and provide incentives to encourage widespread participation.
- When content and performance standards are altered for a student with a disability, (a) the alternate standards should be challenging yet achievable, (b) they should reflect the full range of knowledge and skills that the student needs to live a full, productive life, and (c) the school system should inform parents and the student of any consequences of these alterations.
- Accommodations should be justified on an individual basis and should be unrelated to the knowledge and skills being measured.
- States and localities should provide information to parents to help them make more informed choices about their children’s participation and to understand the consequences of the choices.
- Students with disabilities should be given sufficient opportunity to learn the knowledge and skills expected of them before high stakes consequences are attached to their performance.
- Policymakers should monitor the unintended consequences of participation in standards-based reform for all students including those with disabilities.
- States should design standards policies that reflect realistic timelines and the resource levels needed to implement standards-based reform.

The validity of an exit exam that allows for accommodations for students with disabilities is a major concern, but one for which there is little guidance from research.

Accommodations are meant to correct for score influences that are caused by a disability unrelated to the construct being measured. However, the threat is that the correction may be the wrong one, irrelevant, or excessive. Many approaches to assessment accommodations do assume that disabilities are not directly related to the construct being tested, and such relationships are very important to judging the validity of inferences based on the scores. Inconsistent descriptions and classifications of disabilities and lack of clearly segregated skills have resulted in difficulties in maximizing validity when accommodations are involved (National Research Council, 1997). More research on the validity of scores from

accommodated testing is needed. The 1995 field test of the National Assessment of Educational Progress (NAEP) in mathematics and science provided some evidence about the relationship of accommodations to standards-based reform. The results revealed that it is feasible to assess a greater number of students with disabilities than previously. However, given small sample sizes and lack of clarity and consistency between disability and corresponding accommodations, the results are only suggestive of the effects of accommodations on score comparability (National Research Council, 1997).

## **English-Language Learners**

Failure to graduate from high school has tremendous consequences on a student. Hauser (1997) reports evidence that such failure is linked to problems of employment, earnings, starting and maintaining a family, civic participation, and health. The Mexican American Legal Defense and Educational Fund (MALDEF) pointed to such consequences when it filed a lawsuit over the Texas graduation exam, alleging that it discriminates against Hispanic and African American students, who fail the test in disproportionate numbers (Smith, 1999; Herrick & Walt, 1997). Whether failure to gain a high school diploma is due to graduation tests or other reasons, Hauser (1997) aptly captures the bleak outlook for these students: “Failure to obtain at least a high school diploma looks more and more like the contemporary equivalent of functional illiteracy. High school dropout indicates a failure to pass minimum thresholds of economic, social, or political motivation, access, and competence” (p. 154).

In the 1998–99 school year, approximately 37% of California students were not native-English speakers; almost 25% were classified as ELL (English-language learners) (California Department of Education, 1999b). Thus, the preparation and performance of these students on the graduation exam are very important considerations for California.

Several current federal laws stipulate the participation of *all* students in assessments of performance (e.g., Goals 2000, Title I (Helping Disadvantaged Children Meet High Standards), and Title VII (Bilingual Education)) (National Research Council, 1999). The primary difficulty with assessment of English-language learners (ELL) is obtaining accurate reflections of their knowledge in a given subject, because deficits in language may impede the students’ understanding of—and response to—an exam question. Therefore an incorrect response may not reflect a lack of content knowledge, but rather a language problem. Consequently, the first step of the “decision” is whether the ELL student should participate in the assessment at all—that is, whether the student should be exempt. Such decisions are usually based on some indicator of English proficiency.

California decided to allow fewer exemptions for the SAT9, which is required of all its students grades 2 through 11. It requires all ELL who have been in the state for more than 1 year to take the test in English (National Research Council, 1999). However, the superintendent of San Francisco Public Schools filed a suit against the policy, and prevailed, maintaining that it takes more time than has been allowed before a student becomes sufficiently proficient to take the test in English (“S.F. files suit,” 1998).

For the high school exit examination, California will require all students who do not have waivers to pass in order to receive a diploma. ELL, as a designation, is not a reason for a waiver; thus, it would be more accurate to use a term such as “temporarily exempt” or

“deferred” for ELL (as clarified during meeting of California Board of Education, March 9, 2000). Nevertheless, an essential consideration for testing ELL is whether the students have been taught the knowledge and skills that the test measures, including oral and written English.

When an ELL student is included, the state must address the best method of administering the test or modifying testing procedures to obtain accurate measures. Florida, Texas, and New York, which also have large ELL populations, have addressed some of the challenges in the following ways (National Research Council, 1999).

**Florida**

- ELL in approved program fewer than 2 years may be temporarily exempted from taking the high school competency test, but they cannot receive a standard diploma until the test is passed.
- Students not receiving special language assistance may not be exempted solely on basis of ELL designation.
- ELL who do not pass the competency test can return for a 13<sup>th</sup> year to focus on sections that were not passed; those who do not pass can opt for a certificate of completion rather than a diploma.

**New York**

- Students must pass state Regents exams or approved alternative assessments in four core courses—English, mathematics, social studies, and science—plus other courses, to demonstrate achievement of state standards.
- ELL who enter U.S. schools in 9<sup>th</sup> grade or later can take all required Regents exams, except English/Language Arts, in five different languages.
- All ELL must pass the English Regents exam to receive a diploma.
- State is implementing a comprehensive strategy to help ELL pass the exam, which includes (a) identification of model programs; (b) professional development for all teachers of ELL on integration of English language arts standards into instruction; (c) alignment of ELL programs with state standards; and (d) other added instructional programs and resources, publications, and statewide symposia.

**Texas**

- Students must pass an exit exam in reading, writing, and mathematics, taken in 10<sup>th</sup> grade, to graduate.
- School-based committee uses six criteria to determine whether each ELL is tested on the English TAAS or Spanish TAAS, or is exempted and given an alternative assessment.
- ELL entering U.S. schools in third grade or later are required to take English TAAS after 3 years.

Aside from commonly expected validity problems of bias in testing (e.g., small number of ELL in sample, content bias, and linguistic and cultural biases), when a test is developed in a non-English version, other concerns arise, including translation and score equivalence.

A basic truism is that to some extent all assessments are measures of language (see for example, National Research Council and Institute of Medicine, 1997: 120–122 and The National Education Goals Panel, 1998: 6). However, in spite of potential concerns and problems with native-language testing, given California’s demographics, this is an important consideration for its graduation exam. Lack of proficiency in the language of a test can severely underestimate the test taker’s knowledge. Further, scorers may introduce errors on extended response or performance-based items through unwarranted influence of linguistic elements of responses (National Research Council, 1999).

There are clear calls for more research that would inform decisions about obtaining valid scores for ELL in large-scale assessments (National Research Council and Institute of Medicine, 1997; Olson & Goldstein, 1997). Such research would help address existing difficulties in various strategies being used or considered to enhance ELL participation in large-scale assessments (National Research Council, 1999). For example, use of native-language assessments requires more research on equivalency, validity and reliability before a state incorporates them in its assessment program. Information from New York, which is trying native-language exams in all subjects except English, will help inform this area.

A pressing need in the area of English-language learning and assessment is for a definition of English proficiency that can be applied across the country (Olson & Goldstein, 1997). Issues related to this include the following:

- Use of an English-language assessment of proficiency rather than years in English-only instruction; years of instruction may not accurately predict literacy.
- A proficiency measure should include both oral and written language.
- An individualized approach to match the testing to a student’s English literacy level requires development, validation, and adoption of a standard procedure to determine levels of literacy; then, a threshold level would be established to indicate when a student could take the standard English assessment.
- For subjects other than English, when a student is proficient in more than one language, and tests are available in those languages, the student’s more/most proficient language should be the basis for testing.
- The amount of instruction in a particular subject and in the native language should be considered in deciding the language of the assessment. (Texas’ procedure for this type of decision-making will help inform this area.)

Currently, California statute (Senate Bill 2X) includes language that directs the school district to make a determination of English-language proficiency and whether a given student should be assessed by the graduation exam or deferred for up to 24 months to allow the student at least 6 months of English-language instruction. Under the law, a district must offer summer school programs to any student who does not demonstrate sufficient progress toward passing the graduation exam (using STAR results or grades and other academic indicators). Funding for this initiative includes ELL who need to build their English-language proficiency (“SB 2X,” 1999). This is an important step; however, based on experiences in Texas, legal challenges may invoke inequities in the entire school system for



students of color and ELL (Herrick & Walt, 1997). This suggests the need for a state to examine its educational infrastructure and to begin a standard of instruction from the beginning of all students' education.

### **The Legal Milieu**

These groups of students—students who have not been taught the knowledge and skills measured by the test, those who claim insufficient notice, students of color, students with disabilities, and English-language learners—are all groups who may eventually feel that they have a legitimate legal challenge to the California High School Exit Exam. Certainly proactive steps can be taken to minimize the likely challenges. Unfortunately, there have been inconsistent and even contradictory legal decisions in the past, and there is no single set of rules or guidelines that can guarantee successful legal defense.

In the face of probable legal challenges and decisions, well-informed test implementers would do well to design and build a test framework with these issues in mind. The educational community has vast experience with considering issues of adverse impact and the needs of various student groups. Psychometric, legal, and education expertise all play critical roles.

### **Summary**

As California undertakes the development and implementation of its high school exit examination, policymakers and assessment staff are interested in using the experiences of states that preceded them in this endeavor. This background chapter focused on the most common issues and concerns and the current “state-of-the-art” in high stakes graduation testing. There are two important paths that a state must traverse in building its exit exam program. It is clear that a state must consider critical links, or logical steps, among the parts of its program in wisely planning and establishing the foundation and structure that will support its graduation test. A state can also benefit by paying attention to legal decisions in other states, which have been a predominant influence in setting the requirements a state must meet in creating its exit examination.

### **Critical Links**

A strong system of supporting and integrated “links” is necessary to achieve the desired outcomes of graduation testing. The most common reasons for establishing an exit exam program are to promote standards for awarding a diploma and improved classroom instruction. However, these improvements do not grow from the decision itself; they require planning and nurturing. The following test criteria are critical links for the state to consider:

- *The Standards of Test Use...*
  - Have measurement validity, defensible test psychometrics, and appropriate content aligned with standards and taught throughout the state;
  - Address causes of poor results with valid explanations; and
  - Lead to educationally beneficial outcomes.
- *The Test Scores...*

- Consider use of test scores for diagnostic purposes, or
- Couple “pass-fail” scores with program of diagnostic information and appropriate remediation.
- *The Score Reports...*
  - Take into account individual- and aggregate-level test results;
  - Reflect audience needs and uses clear, straightforward language; and
  - Provide a well-designed format.
- *Public Awareness and Communication by the state...*
  - Attend to the dissemination of explanatory and promotional information about the examination and its related components; and
  - Use a variety of dissemination and distribution formats and strategies appropriate to the audience.
- *The Instructional Changes...*
  - Involve relating patterns of test outcomes to appropriate instructional approaches;
  - Take into account individual- and aggregate-level results;
  - Address instructional needs of students who fail the exam;
  - Deliver appropriate and timely first-time instruction to students in advance of taking the exam;
  - Establish sound feedback mechanism between results and educators; and
  - Provide adequate resources for materials and teacher training to facilitate desired instructional requirements and modifications.

## **Legal Influences**

Policy decisions for any educational area are first and foremost established and manipulated through laws and statutes. Every day, education policymakers deal with and interpret federal, state, and local mandates in regard to numerous aspects of educating children. These regulations act as a foundation and scaffolding that must be solidly in place to support implementation of new policies and programs, such as graduation testing. For example, Titles I and IX have long functioned to specify education requirements, including those that apply to testing in general. Civil rights regulations also must be considered in educational decision-making because students in identified groups retain their rights in all aspects of their lives, including education. Policymakers must ensure adherence to these laws and safeguards in all educational areas. Thus, any of these laws can serve as the basis for legal challenges in education.

In the area of testing, the issues focus on whether the test may be discriminatory or used inappropriately for promotion, tracking, or graduation. Unfortunately, there is no absolute judgment on what constitutes discriminatory or inappropriate practices. This results in decisions on analogous legal challenges that may be diametrically opposed, or that finish in similar conclusions but are based on different justifications. However, the Improving America’s Schools Act (IASA) of 1994, which made major changes to Title I, is nearing its

2000–2001 school year date for implementation of the new testing system. States must interpret and operationally define the requirements of the act, and it seems likely tests used in this area will become a target of legal challenges.

Another side of legal considerations is how courts have applied legal principles to issues related directly to students and high school graduation. Any state contemplating implementation of an exit examination is keenly aware that legal challenges will most likely follow. It makes sense, then, to study what has happened in other states in order to prepare for or try to avoid legal action. Legal decisions in other states have served as one primary source of establishing basic requirements of graduation tests. Based on the decision in Florida, there are three standards that high school exit examinations must meet to pass constitutional muster:

1. The test must have curricular validity, that is, measure knowledge and skills that are taught in the state's schools.
2. Students must receive adequate notice of the test, the requirements for passing, and the consequences of not passing.
3. The test must not intentionally discriminate against a protected group or class.

Although these represent results in a particular state, they are currently the accepted “starting point” for other states in their planning, development, and implementation of an exit examination. Texas has experienced the most recent decision related to its graduation test, and the state education agency prevailed on the basis of supporting each of the areas above. Therefore, it would seem that the most prudent recommendations to California and other states regarding graduation testing are to (a) adhere to all the applicable laws related to education, students, and programs generally, and to testing specifically; (b) follow the related professional standards for developing, implementing, monitoring, and use of the test; and (c) document the processes and results related to each aspect of the laws and the test.

Perhaps the most useful guidance for California and other states comes from U.S. Secretary of Education, Richard W. Riley, a strong supporter of standards-based reform. In his February 22, 2000 “State of American Education” address, Riley called for a “midcourse review” of the standards movement, a step he said was needed “because there is a gap between what we know we should be doing and what we are doing” (Riley, 2000: 6). Specifically, Secretary Riley said that state standards should be “challenging but realistic.... [Y]ou have to help students and teachers prepare for these [high stakes] tests—they need the preparation time and resources to succeed, and the test must be on matters that they have been taught” (Riley, 2000: 7). He also advised states not to rely on any single measure of students’ knowledge in making high stakes decisions: “All states should incorporate multiple ways of measuring learning” (Riley, 2000, 6).

## CHAPTER 3: ITEM RATING WORKSHOPS

Two panels of educators were convened (one in northern California and one in southern California) to help evaluate items being developed for California’s new High School Exit Exam (HSEE). Each panel participated in an item review workshop designed to provide two types of information: alignment of the field test items to California’s Content Standards (California Department of Education, 1999d, 1999e) and alignment of the curriculum being taught in the classroom with each of the field test items.

### Item Rating Procedures

#### Panel Members

The panel members were recruited from our longitudinal evaluation of 24 districts selected to be representative of the state sample as described in Chapter 1. We asked districts to identify individuals who were highly knowledgeable about the district’s language arts or mathematics curriculum and instruction and were currently serving in either a teaching role or a district curriculum specialist role. The rating workshops were held on Saturdays and because of time and cost constraints, we were not able to include staff from all districts. We ended up with 44 panelists representing 13 of the 24 districts. Large, medium and small districts, as defined on the basis of the number of 10<sup>th</sup> grade students in our sampling procedures, were represented in the workshops as were a roughly equal number of low-ELL, high-ELL, low-mathematics, and high-mathematics districts. Large districts and districts with low 1999 STAR Math means were slightly underrepresented, but there were at least two districts from each category.

For the standards-alignment ratings, we expected general agreement across participants, regardless of the districts they came from. For the curriculum-alignment ratings, we did expect differences across districts and thus sought to be sure that different types of districts were included in our workshops. For both ratings, the primary focus was on differences among items. We were not yet trying to make inferences about differences among districts and so exact representation of all districts in the state was not a primary goal.

#### Standards Alignment Ratings

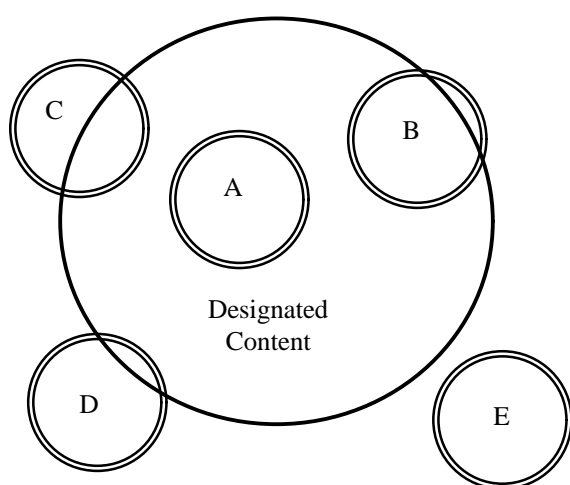
Mathematics or English Language Arts Content Standards<sup>5</sup> “Blueprints” for reading and mathematics have been recommended by the High School Exit Examination Standards Panel; they specify which particular components of each content strand should be included in the exam. Educators in the item review workshops judged the extent to which the items measure their intended content as designated by the item developers.

No single item can completely cover its targeted content. Using Venn diagrams, any given item can be represented as a small circle that is placed in some overlapping relationship to a larger circle representing the item’s target content. Five placements are

---

<sup>5</sup> Reading and Mathematics are “domains,” which are divided into “strands,” which are further divided into “substrands,” which are finally divided into “standards” (CDE, 1999). Items target “standards” in the blueprints.

depicted in Figure 3.1, representing five potential relationships between an item and its designated content.



Items A and B both show good content match. Item A is completely related to the content. For Item B, only a small portion of the item is outside the designated content area, perhaps indicating that it has some necessary language/communication components to convey the mathematical problem.

Items D and E both show clearly unacceptable content match. These items measure something other than designated content.

Figure 3.1 Five theoretical relationships between test items and their designated content.

Item C shows a partial, ambiguous match. This kind of item includes competencies or other factors (beyond necessary language/communication skills) in addition to the designated content. These items are “iffy” in the sense that psychometrically their overlap with the content could allow them to contribute valid test information, but their extraneous content could cause problems. These will also be “iffy” in the minds of the judges about whether or not they have a sufficient content match to be useful in the way intended by the test blueprint. For each field test item, panel members indicated whether each item was more like A, B, C, D, or E.

The items for Mathematics Reasoning (MR) represent a special case. Each of the MR items is obviously intended to assess mathematical reasoning and the reasoning problem for each item is intended to target one of the other mathematics strands. For example, an item may be intended to assess the MR strand “make and test conjectures by using both inductive and deductive reasoning” and attempt to do so by presenting a measurement and geometry problem. For such an item, raters made the additional judgment about whether or not the reasoning requirement involves the designated measurement and geometry standard.

Standards alignment was analyzed in terms of the proportion of items rated as A, B, C, D, and E (recoded to 5, 4, 3, 2, and 1 respectively) by content standard, by substrand, by strand, and by domain. A- and B-rated items clearly signal that, in the opinion of skilled educators, the item covers the intended content. To the extent that high proportions of these items are included in the field test, the potential for content validity of the assessment will be enhanced. C-rated items are more problematic, depending on the sources of extraneous content in the items. Some C-rated items could be used in an operational test if, as a set, they do not introduce systematic bias. High proportions of C-rated items signal the need to closely examine differential item functioning when field test results become available. High proportions of D- and E-rated items signal clear problems with test content or potential

balance in content, at least from the perspective of the panel members and their interpretation of the content frameworks.

### **Curriculum Alignment Ratings**

Although panel members represented only a small proportion of the state, they were drawn from schools that were carefully selected to be representative of the state as a whole. Therefore, the panels provided an initial appraisal of the extent to which the content represented by the field test items is being taught to California students. The information is intended as a baseline against which we will assess changes in the alignment of curriculum toward the standards embodied in the HSEE as it is implemented over the next several years.

For each item, panel members were asked two questions:

1. What percent of your school's 10<sup>th</sup> grade students have been provided sufficient instruction to correctly answer this item?
2. What percent of your school's 12<sup>th</sup> grade students have been provided sufficient instruction to correctly answer this item?

The panelists provided answers on the following scale:

- (4) More than 95%
- (3) 75% to 94%
- (2) 50% to 74%
- (1) fewer than 50%

In analyzing curriculum alignment, we focused on the ratings for 10<sup>th</sup> grade students, since all students will be required to take the HSEE in either the 9<sup>th</sup> or 10<sup>th</sup> grade. We flagged items with an average rating below 2.5 on this 4-point scale as this implied that fewer than 75% of students (across all districts) had been provided sufficient instruction to correctly answer the item. Put another way, for items with these low ratings, more than 25% of all 10<sup>th</sup> grade students may not have had an opportunity to learn the material tested by the item.

### **Rating Booklets**

For each subject, four field test forms with 99 to 102 items each were prepared. We divided each of these forms into two rating booklets of about 50 items each and asked panelists to rate as many of these booklets as possible. Booklets were assigned to panelists in a spiraled fashion so that we had approximately the same number of raters for each booklet and thus for each item.

### **Rater Training and Calibration**

After raters were presented descriptions of their rating tasks and allowed to ask questions, all raters rated the same set of approximately five reading or math items. The group then reviewed the ratings and discussed rationales. This review process helped the raters better understand the rating task by clarifying differences, particularly between B and C rated items.

## Group Discussion

In addition to the rating tasks, educators also engaged in a discussion period at the end of the day, after they had rated a sizable number of items. They were asked first to rate the proportion of their districts' students who they think would have been exposed to each content strand as characterized by the test items. A general discussion about their districts' plans or need for plans to facilitate students' preparation for the exit exam immediately followed.

## Results

### Standards Alignment

The standards alignment ratings used a 5-point scale, with items rated as 4 and 5 (denoted B and A in the discussion above) classified as having good alignment. The mean ratings for each item had an estimated reliability of .69 for each subject (.693 for ELA ratings and .687 for mathematics ratings). While not high enough to use as a basis for making important decisions about individual items, these values indicate a general agreement among the ratings as to differences in standards-alignment across items. The standard error of the item means was .29 for ELA and .38 for math. At this level of accuracy, differences of about three-quarters of a point were statistically significant. Our primary concern, however, was the average of the ratings for all items and for major groupings of items, and not with rating differences for specific items. The accuracy of these statistics was much greater, with standard errors less than .15 for the overall average.

Tables 3.1 and 3.2 show the results for the alignment of the items to their targeted standards. In general, the judges agreed closely with prior panels of experts who had screened these items. Nearly 80% of the ELA items and well over 90% of the mathematics items were judged to be "on target." Within ELA, the greatest concern was with the Writing Application items. These were all extended constructed response (essay) items, but scoring rubrics and anchor papers were not yet available for these items. These items will need to be re-examined when more information on scoring procedures becomes available. There were also concerns about a number of reading comprehension items, particularly items having to do with synthesis or comparison.

For mathematics, the mathematical reasoning category was the most problematic, followed by statistics, data analysis and probability at the Grade 7 level. We recommend more careful analysis of the mathematical reasoning standards and how they are assessed. The only standard with more than two low-rated items was "Make and test conjectures by using both inductive and deductive reasoning." As with reading, it appears that it is more difficult to write items for standards involving higher levels of synthesis and analysis. In all, however, only 35 of the 396 mathematics items were flagged for potential problems with alignment to the targeted standard.

Table 3.1 Total Items and Number With Low-Standards Alignment Ratings for Each Major Language Arts Content Category

SUBJECT/STRANDS	TARGET NO. ITEMS PER TEST FORM	NO. ITEMS IN THE FIELD TEST	NO. WITH LOW ALIGNMENT RATINGS*	PERCENT FLAGGED
<b><i>Reading Vocabulary:</i></b>				
Word Analysis, Fluency, and Vocabulary Development (RV)	10	42	3	7.1%
<b><i>Reading Comprehension:</i></b>				
Focus on Informational Materials (RI)	30	121	43	35.5%
Literary Response and Analysis (RL)	30	68	16	23.5%
<b><i>Writing:</i></b>				
Writing Strategies (WS)	12	58	7	12.1%
Written and Oral English Language Conventions (WC)	18	61	2	3.3%
Writing Applications (WA)	2	12	12	100.0%
<b>TOTAL ALL ITEMS</b>	<b>102</b>	<b>362</b>	<b>83</b>	<b>22.9%</b>

\* Note: Items were flagged if the average standards-alignment rating was less than 3.5 indicating significant differences between item content and the content of the targeted standard.

Table 3.2 Total Items and Number With Low Standards Alignment Ratings for Each Major Mathematics Content Category

CONTENT AREA (STRAND)	TARGET NO. ITEMS PER TEST FORM	NO. ITEMS IN THE FIELD TEST	NO. WITH LOW ALIGNMENT RATINGS*	PERCENT FLAGGED
Statistics, Data Analysis, and Probability (Grade 6)	6	41	1	2.4%
Statistics, Data Analysis, and Probability (Grade 7)	8	23	4	17.4%
Number Sense	14	63	3	4.8%
Algebra and Functions	17	77	8	4.8%
Measurement and Geometry	20	80	5	6.3%
Mathematical Reasoning	8	43	9	20.9%
Algebra 1	26	69	5	7.2%
<b>TOTAL ALL ITEMS</b>	<b>99</b>	<b>396</b>	<b>35</b>	<b>8.8%</b>

\* Note: Items were flagged if the average standards-alignment rating was less than 3.5 indicating significant differences between item content and the content of the targeted standard.



## Curriculum Alignment

Tables 3.3 and 3.5 show the numbers of items with relatively low curriculum alignment ratings for ELA and math respectively. We relied on the 10<sup>th</sup> grade ratings, since all students will be required to take the HSEE in either 9<sup>th</sup> or 10<sup>th</sup> grade. We used a cutoff of 2.5 to separate items where more than 25% of the students were judged not to have been prepared to answer the item. In analyzing the curriculum-alignment ratings, we excluded items that had been flagged for low standards-alignment. The excluded items were judged to be poor measures of the target standards and so the alignment of the item to the district's curriculum was not relevant.

For both ELA and math, our panelists believed that, for a majority of the items, there are a significant number of students who have not had the opportunity to learn the skills tapped by these items. The ELA panelists, in particular, judged that over 90% of the standards-aligned items could cause problems for their 10<sup>th</sup> grade students. Writing Conventions was the only category where fewer than 90% of the items were flagged.

For Mathematics, the panelists rated just over half of the items as having potential curriculum-alignment problems. There was considerable variation across the different math content categories in these ratings. Nearly 80% of the Algebra 1 items and 64% of the Algebra and Functions items were flagged. The percentage of items flagged in the other content categories was significantly less.

Since each panelist rated alignment to the curriculum of a different district, we could not use inter-rater agreement as a measure of the accuracy of the curriculum-alignment ratings. Instead, we used information on the difficulty of the test items in the field test (see Chapter 4) as an indicator of the validity of the ratings. Test questions may be difficult for several reasons, one of which is that a significant number of students have not been taught the content measured by the item. We would expect that, on average, items that were not aligned to the curriculum of a significant number of students would be more difficult than items that were aligned to the curriculum of most students. Table 3.5 shows that this was, in fact, the case for both the ELA and the math item reviews. The difference in passing rates between high and low alignment items was nearly 30 percentage points for both subjects.

Table 3.3 Total Items and Number With Low Curriculum Alignment Ratings for each Major Language Arts Content Category

SUBJECT/STRANDS	NO. ITEMS IN THE FIELD TEST	NO. ALIGNED TO TARGET STANDARD	NO. WITH LOW CURRICULUM ALIGNMENT*	PERCENT FLAGGED
<b><i>Reading Vocabulary:</i></b>				
Word Analysis, Fluency, and Systematic Vocabulary Development	42	39	36	92.3%
<b>Reading Comprehension:</b>				
Focus on Informational Materials	121	78	74	94.9%
<i>Literary Response and Analysis</i>	68	52	48	92.3%
<b><i>Writing:</i></b>				
Writing Strategies	58	51	47	92.2%
Written and Oral English Language Conventions	61	59	48	81.4%
Writing Applications	12	0	0	---
<b>TOTAL ITEMS</b>	<b>362</b>	<b>279</b>	<b>253</b>	<b>90.7%</b>

\* Note: Items were flagged if the average for the 10<sup>th</sup> grade student ratings was less than 2.5 indicating that more than 25% of the 10<sup>th</sup> graders had not received sufficient instruction to prepare them to answer the item correctly.

Table 3.4 Total Items and Number With Low Curriculum Alignment Ratings for Each Major Math Content Category

CALIFORNIA CONTENT STANDARD	NO. ITEMS IN THE FIELD TEST	NO. ALIGNED TO TARGET STANDARD	NO. WITH LOW CURRICULUM ALIGNMENT*	PERCENT FLAGGED
Statistics, Data Analysis, and Probability (Grade 6)	41	40	15	37.5%
Statistics, Data Analysis, and Probability (Grade 7)	23	19	7	36.8%
Number Sense	63	60	18	30.0%
Algebra and Functions	77	69	44	63.8%
Measurement and Geometry	80	75	36	48.0%
Mathematical Reasoning	43	34	14	41.2%
Algebra 1	69	64	51	79.7%
<b>TOTAL ITEMS</b>	<b>396</b>	<b>361</b>	<b>185</b>	<b>51.2%</b>

\* Note: Items were flagged if the average for the 10<sup>th</sup> grade student ratings was less than 2.5 indicating that more than 25% of the 10<sup>th</sup> graders had not received sufficient instruction to prepare them to answer the item correctly.

Table 3.5 Relationship of Item Difficulty to Curriculum Alignment Ratings

SUBJECT	CURRICULUM ALIGNMENT	PERCENT ITEMS WITH MORE THAN 50% PASSING IN THE FIELD TEST
Language Arts	High	93.9%
	Low	67.7%
Mathematics	High	59.4%
	Low	30.9%

### **Conclusions**

The results of our analyses confirmed that most of the items developed for the field test are reasonably aligned to the state standards. Even though our panelists were not necessarily experts in these standards, they reached reasonable agreement among themselves on differences in alignment across the pool of field test items and they agreed as well with the HSEE Panel who had examined the alignment of these items. A modest number of items were flagged for further review. Nearly all of these items had standards-alignment ratings near the margin of acceptability.

With respect to the alignment of the test items to the current curriculum in different districts, our results suggest that curriculum specialists need to bring the curriculum into alignment with targeted standards and items tested on the HSEE. Current results are only baseline information. The State Board has not yet adopted these standards and the students who will have to meet the standards have not yet reached high school (they have now completed the 8<sup>th</sup> grade). It will be important to repeat this exercise next spring to see whether the alignment of the curriculum to the items and hence to the standards is increasing.

## CHAPTER 4: ANALYSIS OF FIELD TEST DATA

### Introduction

As stated in Chapter 1, one focus of this preliminary report is the quality of the items that have been developed for use in the initial form of the HSEE. We have reviewed the process used in developing and editing new items and assembled two panels of educators who also reviewed the roughly 750 items that have been developed. (See Chapter 3 above.) In late May and early June, the development contractor (AIR) conducted a field test of these items, administering them to a reasonably large and representative sample of California 10<sup>th</sup> graders. As part of our evaluation, we have conducted independent analyses of the data collected from this field test. We describe these analyses and their results in this chapter.

Our analyses addressed three general issues. First, *what proportion of the items have good statistical properties?* The answer to this question provides an indication of the soundness of the development procedures and also will determine whether there are enough high-quality items to begin assembling one or more operational forms of the exam. Second, we want to provide a preliminary *assessment of the likely accuracy of scores* from the items that were field tested. Accuracy is one of the important issues that the State Board must consider as it reaches a decision about adopting the test in September or October. The third issue is whether we can estimate the *possible impact that the exam would have on different demographic groups* if it were administered today (or, more precisely, last Spring).

It is important to point out that our analyses are necessarily preliminary. We do not yet have results for the essay questions that are still being scored and we have not yet had a chance to review primary analyses being conducted by AIR. We plan to issue a supplemental report on August 25, after we receive additional information and conduct further analyses. Some of the counts and estimates in this preliminary report may change slightly in our final version or may differ slightly from results of analyses conducted by AIR. It seems likely that any such differences will be minor and will not affect the main conclusions that we draw from our preliminary analyses.

### Field Test Design

#### Test Booklets

AIR constructed four test booklets (forms) of English language arts (ELA) items. Each form contained 100 multiple choice (MC) items followed by two constructed response (CR) essay items. A total of 59 different reading passages with MC questions (items) were tried out. Many of these passages were included in more than one test booklet with differences in some or all of the questions asked about the passage. The purpose of this repetition was to avoid asking too many different questions of any one student, but still allow the contractor to pick the best items for each passage when it is used in an operational form. In all, 338 unique multiple choice items were tried out in the field test, with 62 of these items included in two different forms (bringing the total MC items printed to 400, or 100 per booklet). Three different versions (subforms) of each booklet were created with the same MC items

but different CR items. In this way a total of 24 different CR items were tried out (2 for each of the 3 versions of each of the 4 forms).

AIR also constructed 4 forms of Mathematics (Math) items. Each form contained 99 MC items. There were no CR items for math and there was no overlap across the 4 different math booklets.

## Field Test Sample

Details of the Field Test Sampling plan will be presented in AIR's report on the field test. Their basic goal was to ensure that the sample of students completing each test booklet covered a wide range of abilities and was generally representative of 10<sup>th</sup> grade students in California. The field test was not intended to provide normative information, as operational forms have not yet been assembled, so exact representativeness was not a primary concern. For each of the two exams, AIR sorted California schools by their level of performance on the corresponding 1999 STAR test and then picked 10 schools from the lowest performing tenth (decile) of these schools, 10 schools from the next lowest performing tenths, and so on up to 10 schools from the top performing 10<sup>th</sup>. This approach appears to be an effective way of obtaining samples of schools that span the full range of ELA and Math abilities.

For each of the selected schools, AIR requested up to 66 10<sup>th</sup> grade students. Some of the schools were too small to be able to comply with this request and others could not supply the total requested students because of end-of-year scheduling problems. Fortunately, AIR had been reasonably conservative in planning for this contingency and the resulting sample sizes appear adequate for most or all of the intended analyses. Within each school, the four different ELA or Math booklets were assigned to roughly one-fourth of the students tested. This provided "randomly equivalent" samples of students for the different booklets (the same ability levels except for random factors in the assignment to booklet that become negligible with large sample sizes).

Table 4.1 shows the total number of students completing each booklet. In these and the tables that follow, a small number of students with missing form codes or no valid item responses (86 in the ELA sample and 84 in the Math sample) were deleted from our analyses. Even though the tests were long, nearly all students responded to all of the items. Only 6% of the ELA sample and 5% of the Math sample failed to respond to (omitted) more than five of the 100 or 99 items.

Table 4.1 also shows the average "total correct" scores for the 100 (ELA) or 99 (Math) items presented to each student. For both subjects, these averages and the standard deviations (which show how much the scores varied across different students) were very similar across the four test forms. Assuming the random assignment of students to booklets worked as intended, this similarity in number correct scores suggests that the items in each of the different booklets were of comparable average difficulty.

**Table 4.1 Average Total Correct Scores by Subject and Field Test Form**

Subject	Form	Sample Size	Average (%)	Standard Deviation
ELA–MC	1.x	998	58.9	20.6
	2.x	1017	59.3	20.8
	3.x	906	58.9	21.0
	4.x	836	58.5	19.6
	ALL	3757	58.9	20.5
Math	1	1009	46.1	19.2
	2	922	47.3	18.0
	3	1020	48.1	17.9
	4	969	46.7	17.0
	ALL	3920	47.1	18.1

### Item Difficulties

The results in Table 4.1 above also provides important information on the average difficulty of the HSEE items for California 10<sup>th</sup> Graders. Each ELA form had 100 MC items, so an average score of 59 means that, across items, the average percent answering correctly was also 59. These items were all 4-option multiple-choice items. Because of guessing, the percentage of students answering correctly is greater than the percentage that actually know the right answer. For example, suppose 45.3% of the students knew the correct answer and the other 54.7% guessed randomly. All of the students knowing the answer (45.3%) would answer correctly and one fourth of the students who did not know the answer (13.7%) would answer correctly through random guessing, so the expected percent answering correctly would be 59% (45.3 + 13.7). This example suggests that these are relatively difficult items, with fewer than half the students knowing the answer for the average item.

For Math the items appear even more difficult. An average score of 47.05 on a 99-item test translates into an average of 47.5% correct on the average item. This result would be obtained if only 30% of the students knew the correct answer and one-quarter of the other 70% (17.5%) answered correctly through guessing. Thus it seems that less than one-third of the students knew the correct answer for the average math item.

We also examined the distribution of number correct scores for different demographic groups as shown in Tables 4.2 and 4.3. These results provide a preliminary indication of the relative difficulty of the HSEE items for different groups of students. In a later section of this chapter, we will use this information, along with data from the 1999 STAR administration, to provide very preliminary suggestions about the possible differential impact (passing rates) of the HSEE for these different groups of students.

**Table 4.2 Average Total Scores by Gender**

Subject	Gender	N	Mean (%)	Standard Deviation
ELA–MC	Female	1835	62.9	18.9
	Male	1895	55.2	21.2
Math	Female	1915	46.8	17.3
	Male	1988	47.4	18.8

**Table 4.3 Average Total Scores by Race and Language Fluency**

Subject	Race/Language Status	N	Mean (%)	Standard Deviation
ELA–MC	African American (1)	197	50.2	20.0
	Asian (3)	266	68.8	18.6
	Hispanic (5)	1314	50.9	18.3
	White (7)	1610	65.7	19.4
	L.E.P	430	40.5	14.4
Math	African American (1)	300	41.0	15.4
	Asian (3)	318	57.6	19.8
	Hispanic (5)	1108	38.6	13.8
	White (7)	1800	52.0	18.1
	L.E.P	318	35.6	14.8

### **Item Screening**

A total of 396 Math items were organized into 4 field test booklets of 99 items each. Four test booklets of ELA items were also included in the field test. Each of the ELA booklets contained 100 multiple choice (MC) items. Each booklet was further divided into 3 “subforms” with two distinct constructed response items in each subform. Across the forms and subforms, a total of 24 constructed response items was administered to about 350 students each. Scores for these items are not yet available, so their analyses will be included in a supplemental report.

The 100 MC items included in the 4 forms were not all unique. There were 338 unique MC items with 62 of these items included in two different forms each to provide a basis for linking statistics for item in the different test forms. In our item screening analyses, we used statistics from the first occurrence (lowest form number) of each of these duplicated items. To the extent feasible, our supplemental report will include analyses of differences in statistical results across forms for these duplicated items and alternatives for pooling results across forms.

We made a very preliminary effort to estimate the number of field test items with statistical properties that suggest they would need to be dropped or revised (and re-tested) before being used in operational forms. Statistical indicators were used to assess: (1) whether items were inappropriately easy or difficult, (2) whether the item score provided information that was at odds with (did not generalize to) the information provided by the other items, and (3) whether the item appeared to function differently for different demographic groups (females, Hispanics, or African Americans).

### **Item Difficulty**

We computed the percent passing (p-values) for each item. In subsequent analyses, it might be possible and desirable to adjust these p-values for differences between the field test samples and the total population of California’s 10<sup>th</sup> grade students. As noted above, the procedures used in drawing the sample should have been sufficient to ensure that any such adjustments would be minor. Item difficulty screens are used to weed out items, which, although they could be perfectly valid, provide little or no useful information. More often than not, extreme item difficulties also reflect item flaws so that most of the items screened out are not valid measures of the intended standards as well as being inefficient. For

example, if nearly all students pass an item, it may well be that the distracters (incorrect options) are not plausible or that something in the item text “gives away” the correct answer. Similarly, if the percentage answering correctly is at the guessing level (suggesting that no one really knows the correct answer), the item provides little information about student skills and is likely to be flawed. In this case, the item could be incorrectly keyed or have no correct option or have some problem in the text that leads even able students astray. We flagged items with passing rates above 95% as too easy and those with passing rates below 25% (the guessing level for 4-option items) as too difficult.

### **Item-Total Correlation**

Another indicator of potential item problems is when results from the item disagree with (fail to generalize to) the scores on other items. The item-total correlation coefficient measures the extent to which students who answer the item correctly also score well on the rest of the test. Because the item score is dichotomous (scored pass or fail) and the total score has a continuous (more normal) distribution, the range of the item-total correlations is limited, particularly when the percentage of students passing the item is much different from 50. We computed a Clemans-Brogden biserial correlation coefficient (Lord & Novick, 1968, page 341) that corrects for differences in item difficulty. Possible values range from  $-1.0$  to  $+1.0$  with positive values indicating agreement between the item score and the total score. We flagged all items with values less than 0.2 as having a generalizability problem. Often these items are mis-keyed or have ambiguities in the text or options that limit their validity as a measure of achievement of the targeted standards.

### **Differential Item Functioning (DIF)**

It is common practice to look for differences in the way an item functions across different groups of students. In most analyses of differential item functioning (DIF), a focal group is identified that is of specific concern. The rates at which members of this group answer an item correctly (pass) are compared to passing rates for a second reference group. In our analyses, Hispanics, African Americans, and Females were the focal groups of interest. In each case, statistics for these students were compared to statistics for all other students in the field test.

The issue is not just whether there are different passing rates for these different groups. The question addressed in DIF analyses is whether group differences in passing rates for some items are significantly larger than the differences in passing rates for the other items. Another way of framing this issue is to ask whether students from different groups who are at the same overall level of achievement (usually indicated by the total test score) have the same probability of answering the item correctly.



We computed DIF statistics<sup>6</sup> for females, Hispanics and African Americans—the groups of most common concern in test bias studies. The sample sizes for females and Hispanics (more than 400 and 300 per test form respectively) were large enough to detect moderate and large DIF reliably. The sample size for African Americans was much smaller, generally 40 to 50 per item. Only a few items were flagged as having potentially significant DIF for this group, in part because the sample size was not large enough to allow detection of items with only moderate DIF.

Note that a finding of significant DIF does not necessarily mean that an item is not a valid measure of the intended standard. Group differences in preparation can lead to greater group differences on some items than on others. For example, suppose that male and female students took algebra at the same rate, but many more male students went on to take geometry by the 10<sup>th</sup> grade. We would expect larger gender differences in passing rates for geometry items than for algebra items, even if all items were perfectly valid measures of their intended content. A common practice is to flag all items with significant DIF values for further content and sensitivity review. Many of these items would then be subsequently accepted and used without further changes. We used a relatively high cut-off (the .01 level) to estimate the proportion of items that would eventually be screened out because of DIF concerns.

## Item Screening Results

Table 4.4 summarizes our item screening results. It should be noted that these are preliminary estimates based on statistical criteria only. AIR will end up with somewhat different results using somewhat different statistical criteria and incorporating editorial, as well as statistical, review of flagged items.

Overall the results show the tests to be highly effective in correlating items to standards and asking questions correctly. In many programs, half of the items or more are screened out on the basis of initial field test results. We flagged only 1 out of 4 of the Math items and 1 out of 8 of the ELA items. The very high survival rates for the HSEE items shows a high degree of effectiveness in the item development and review procedures and reflects the fact that some of these items have been previously screened.

---

<sup>6</sup> A commonly used DIF statistic, the Mantel-Haenszel log odds ratio (Mantel & Haenszel, 1959), compares the odds of passing the item (percent correct/percent incorrect) for focal and reference group members at each different total score level. An odds ratio is computed for each total score level (indicating comparable overall ability). If the odds of passing for the focal group are the same as for the reference group, the ratio of the odds values is 1.0 and the logarithm of this ratio is 0.0. To the extent that the log-odds values (across all of the score levels) are different from 0.0, the item is said to function differently (be disproportionately hard or easy) for the focal and reference groups. We computed a chi-square statistic (see Dorans & Holland, 1993, page 40) that tests whether the Mantel-Haenszel statistic is different from 0.0. We flagged cases where the statistic was greater than 7.8794. This corresponds to the .005 level for a one-degree chi-square, meaning that there was less than .01 chance of getting a value this large (or a correspondingly small one) by chance alone.

Table 4.4 Percent of Items Screened Out by Various Statistical Criteria

Subject/Statistic	ELA-MC	Math
Total field test Items	338	396
Number passing all screens	294	307
Percent passing all screens	87.0%	77.5%
% Too easy*	0.0%	0.0%
% Too hard	2.1%	7.6%
% Low Item-Total Correlation	3.6%	6.6%
% DIF–Female	7.4%	9.6%
% DIF–Hispanic	3.0%	0.8%
% DIF–African American	0.0%	1.0%

\* Note: Percents add to more than 100 because some items were flagged for more than one reason.

## Potential Test Accuracy

The second issue that we sought to address in our preliminary analyses of the field test data was how accurate HSEE forms might be. This will necessarily be a technically dense discussion. Accuracy involves concepts of each student’s “true” score (generally defined as the average of the scores across an infinite number of parallel forms) and of error (the difference between the score from a single testing and this true score). Models and estimates of the relative size of errors in test scores are used to construct confidence bounds in score reports, similar to the “margin of error” figures now commonly reported for political polls.

Most of our analyses here are based on models that come from Item Response Theory (IRT). We provide a brief description of the key IRT concepts that underlie these analyses in Appendix A. Readers are referred to more standard texts (Lord & Novick, 1968; Lord 1980; or Hambleton and Swaminathan, 1985) for a more detailed presentation.

## Method

**Simulated Test Forms.** Table 4.5 shows Coefficient Alpha reliability estimates for each of the field test forms. These values are all quite high, which is not surprising given the length of the tryout booklets (which mirrored the target length for operational forms). Each of these tryout booklets contained some items that are likely to be screened out, so further analyses of the accuracy of these particular forms was not judged to be useful.

Table 4.5. Coefficient Alpha Reliability Estimates for each Field Test Form

Form	ELA	Math
1	.96	.95
2	.96	.94
3	.96	.96
4	.95	.94

We used the MULTILOG program (Thissen, 1991) to estimate IRT item parameters for each of the field test booklets. MULTILOG was chosen because it will handle CR items with more than 2 score levels together with MC items scored dichotomously. For items with negative item-total correlations, we fixed the parameters in advance (setting the slope to 0 to estimate a flat-line ICC). Each of the runs “converged” on the first try and there were no extreme values for the parameter estimates (other than for the “fixed” items).

In our further analyses of test score accuracy, we created “pseudo-forms” by selecting two (for ELA) or three (for Math) distinct sets of items from the pool of 294 ELA items and 307 math items with no statistical flags.<sup>7</sup> Note that, except for screening out questionable items, item selection was based on content classification only and not on any statistical properties. Specifically, there was no attempt to match item difficulties across forms as there would be in selecting items for operational test forms.

**Simulated Examinees.** For each of the pseudo-forms, we computed test accuracy information for 100 different levels of ability (simulated examinees). The IRT item parameter estimates were based on the assumption that there was a normal distribution of achievement. We sliced the area under the normal curve into 100 equal-area bars (each bar thus representing 1% of the examinee population). We took the midpoint on the ability scale for each slice as the level of ability we wanted to analyze. These levels of ability thus correspond generally to percentiles, except that percentiles are defined by the boundaries between each slice rather than the midpoints.

**Simulated Decision Points.** While recommendations for test content appear close to final, there has been little specific discussion, let alone any recommendations, of how well students will be expected to perform to pass each of the two tests. Normative information from STAR and NAEP suggest that California students are further behind in reading than in math. The field test data, however, show that the items assessing the Math standards appear to be more difficult than the items assessing the ELA standards.

The decision about how high students must score to pass these tests is a policy judgment. There are no right or wrong judgments (although there may be incorrect interpretations of the results), and it is not possible to predict, in advance, what the final decisions will be. There are, however, common conceptions based on the idea that a pool of items represent evenly some target domain that would suggest some plausible decision points. In less formal testing, it is common to think that about 70% correct is passing. No one can be expected to master 100% of any domain of knowledge or skill, but something above half of the domain would be reasonable. Seventy percent is between those two extremes.

Expecting students to answer 70% of the current HSEE items correctly would appear to be a reasonably high standard, given the difficulties estimated for the field test items. We used 70% correct as the upper end of the range of plausible passing standards and chose 50% to anchor the lower end. If the passing level were set much below 50% of the current items, it might be important to go back to the drawing board and develop some easier items. We used 60%-correct as an intermediate decision point in some of our analyses.

## Results

**Potential Percent Passing.** For each of our 100 simulated examinees (ability levels) we computed their expected number correct score on each of the simulated forms. Figures 4.1

---

<sup>7</sup> To identify items for each pseudo-form, we grouped all of the items by the specific standard they were designed to test and then selected every “kth” item, where k was 294/100 for ELA and 307/99 for Math. This approach ensured as complete coverage of the different standards as possible. We used different starting points for each of the two ELA and three math pseudo-forms so that there would be no overlap in these forms.

and 4.2 show how the expected number correct scores would vary across the different forms and ability levels.

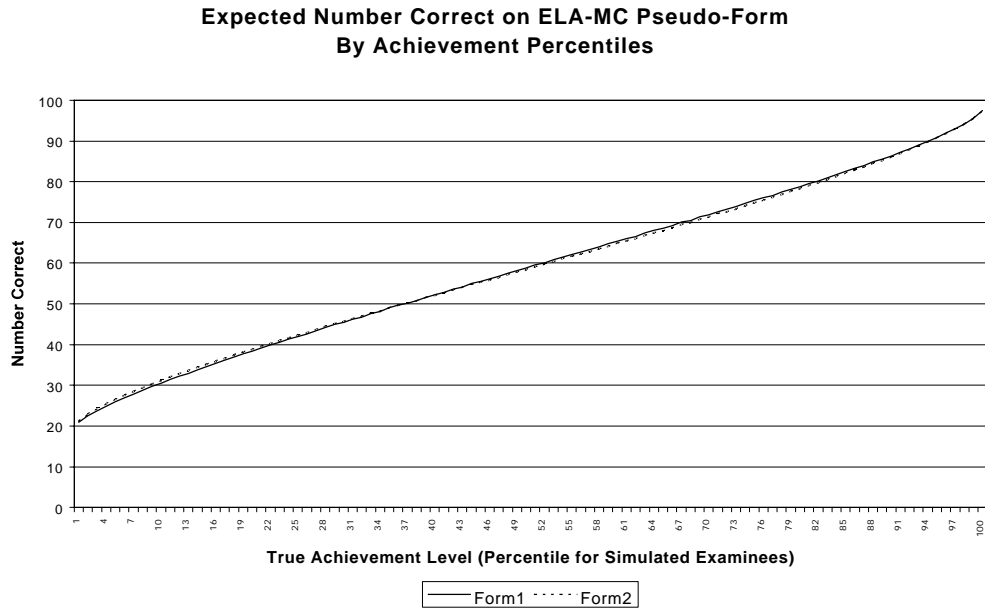


Figure 4.1 Expected number correct for each simulated examinee on each ELA pseudo-form.

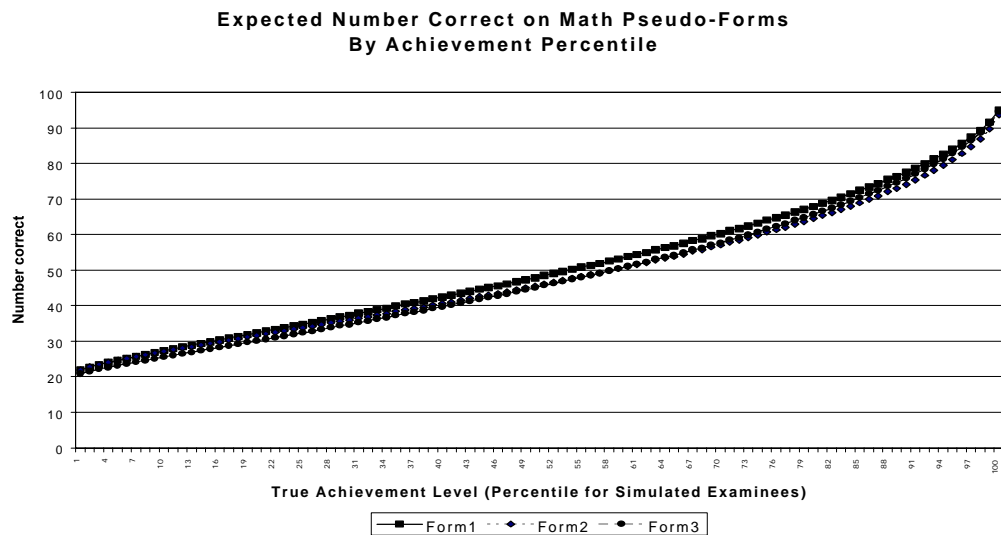


Figure 4.2. Expected number correct for each simulated examinee on each math pseudo-form.

Table 4.6 shows the percentage of students who would be expected to fall in the score ranges defined by decision points at 50, 60, and 70 out of 99 or 100 items correct. For ELA,

about one-third of the students would score below 50, one-third between 50 and 70, and one-third above 70. For Math, more than half of the students would score below 50, while only about 16% would pass if a total score of 70 were required. There was close agreement across the different pseudo-forms, even though no attempt was made to balance their statistical properties. Consequently, we pooled results across pseudo-forms in the remaining analyses.

Table 4.6 Percent of Simulated Examinees Scoring at Different Levels For Each ELA and Math Pseudo-Form

Subject / Pseudo Form	Percent of Students Expected to Score:			
	0–49	50–59	60–69	70 & above
ELA/1	36%	15%	15%	34%
ELA/2	36%	16%	16%	32%
ELA–average	36%	15.5%	15.5%	31%
Math/1	53%	16%	13%	18%
Math/2	58%	16%	12%	14%
Math/3	58%	15%	11%	16%
Math–Average	56%	16%	12%	16%

**Decision Error.** In characterizing the importance of test accuracy, we looked at decision errors as the most critical concern. Decision errors are made if a student passes the exam even though his or her true achievement level (defined on the expected number correct scale) is below the decision point (false positives) or if a student fails when his or her true achievement level is above the decision point (false negatives). Given the relatively low proportion of students expected to reach scores of 70 or higher, we chose 50 items correct as the decision point to focus on.

For each true achievement level, we computed the probability (based on the Item Response Theory (IRT) parameter estimates) that a student at that level would score below the decision point. Figures 4.3a and 4.3b show the proportion of time students would be expected to score below (fail) for each true achievement level.

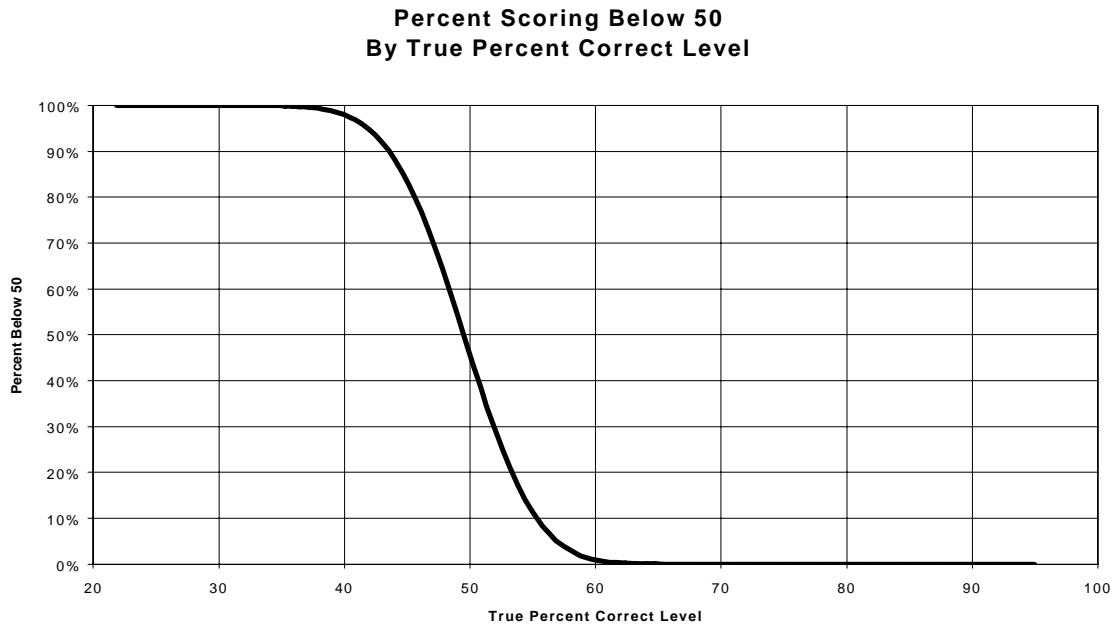


Figure 4.3a Percent Scoring Below 50 for Each True Achievement Level: ELA

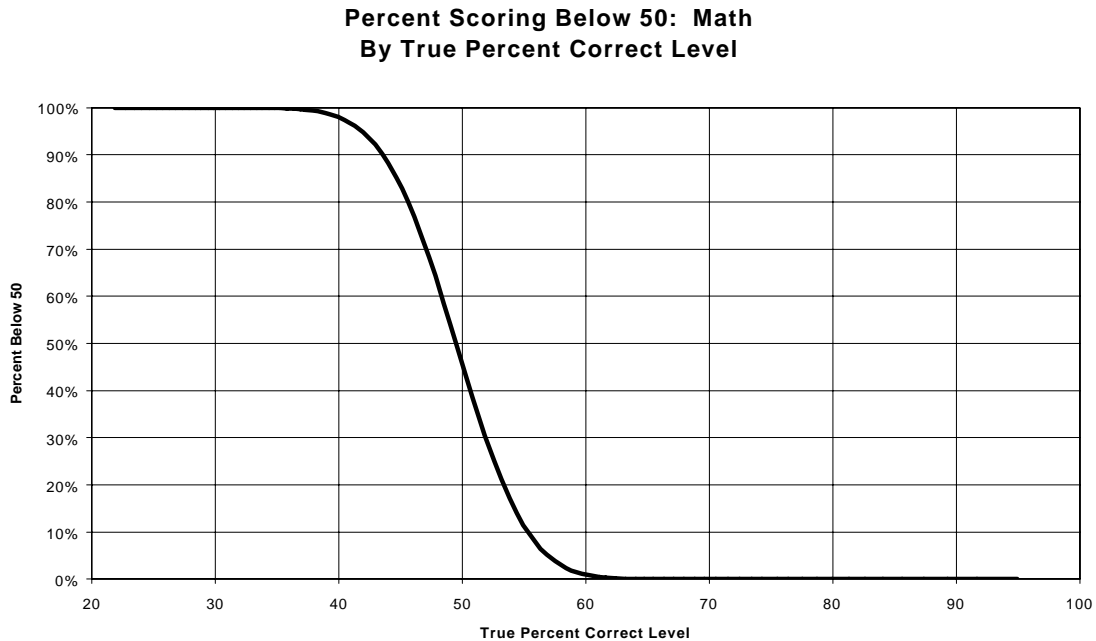


Figure 4.3b Percent Scoring Below 50 for Each True Achievement Level: Math

Nearly all students whose expected score is below 40 will fail and all students whose expected score is above 60 will pass. Between 40 and 60 there is a significant possibility of decision error. Of course, a student whose expected score is right at the decision point will pass or fail about half the time.

Table 4.7 summarizes the percent of the students in each expected number correct range (true achievement) whose single-test scores would fall in each of the decision-point ranges. Fewer than 9% of the students with expected scores below 50 would score in the ranges above 50 and fewer than 8% of the students with expected scores above 70 would score below 70. There is considerably greater uncertainty for students with expected scores between 50 and 70.

**Table 4.7 Distribution of Observed Scores From a Single Testing for Students With Different True Scores (Expected Average Across Parallel Forms)**

Subject	True (Expected) Number Correct	Percent of Students Who Would Actually Score:			
		0–49	50–59	60–69	70 & above
ELA	00–49	91.4%	8.5%	0.1%	0.0%
	50–59	16.0%	63.9%	19.8%	0.2%
	60–69	0.1%	14.8%	65.5%	19.5%
	70 +	0.0%	0.0%	6.0%	94.0%
MATH	00–49	93.2%	7.7%	0.1%	0.0%
	50–59	16.0%	64.0%	19.7%	0.2%
	60–69	0.1%	14.9%	65.9%	19.1%
	70 +	0.0%	0.0%	7.7%	92.3%

The conditional standard errors near the decision ranges were between 4 and 5 points. Table 4.8 shows classification results for a decision point of 50 correct for students whose true scores are more than 5 points (about one standard error) below or above the decision point and for students within 5 points of the decision point. For students outside this standard-error range, 98 to 99% were correctly classified. This seems quite reasonable and is about as far as we can go with statistical analyses. Beyond this point lie policy judgments about the adequacy and consequences of different rates of decision errors.

**Table 4.8 Number of Students Scoring Below/Above 50 by True Score Level**

Subject	True (Expected) Number Correct	Number of Students	Percent of These Students Who Would Actually Score:	
			< 50	50+
ELA	00.00–44.99	28.5	98.0%	2.0%
	45.00–54.99	15.5	45.9%	54.1%
	55.00+	56.0	0.7%	99.3%
MATH	00.00–44.99	47.0	98.5%	1.5%
	45.00–54.99	17.7	47.7%	52.3%
	55.00+	35.3	0.9%	99.1%

### **Potential Adverse Impact**

It is not surprising that average item passing rates varied by demographic group since other indicators of student achievement, most notably recent STAR results, vary similarly by demographic group. We would like to know how passing rates for the test, as a whole, will vary for these different groups, but at this point, passing standards have not yet been set for the HSEE. Ultimately, judgments about how many of the standards must be mastered and to what degree need to be translated into a minimum performance level for a set of test items.

To provide a basis for interpreting performance difference on HSEE items across various groups, we examined results from the 1999 administration of STAR. In the principal surveys described in Chapter 5, nearly half of the principals estimated that fewer than half of last year's 10<sup>th</sup> graders would meet the HSEE requirements. The 1999 STAR results include comparisons to National norms. We looked at the percent of students in key demographic groups scoring above the 25<sup>th</sup> percentile for the nation as a whole (meaning that 75% of the nation's students would pass). For ELA, the passing rate for California would be about 50%. We also looked at what would happen if much higher standards, corresponding to the 50<sup>th</sup> percentile for the nation as a whole, were implemented. Table 4.9 shows the results.

**Table 4.9 Percent of Students Above National Norm Values by Demographic Group**

Demographic Group	Reading		Mathematics	
	National 25 <sup>th</sup> Percentile	National 50 <sup>th</sup> Percentile	National 25 <sup>th</sup> Percentile	National 50 <sup>th</sup> Percentile
All California Students	55	33	70	44
ELL Students	13	3	50	20
African American	40	17	52	22
Hispanic	36	15	57	25
Low SES (Parents Ed < HS)	28	10	54	23

As suggested both by the results in Table 4.9 and the field test score statistics in Tables 4.2 and 4.3, there will almost certainly be lower passing rates for Hispanic, African American, English-language learners (ELL), and low socioeconomic (SES) students than for students in general. The potentially great negative impact of denying more minority, ELL, and low SES students a diploma should be carefully considered. If the program works as intended, this negative consequence would be offset by programs that reduced the number of these students, with or without a high school diploma, who fail to develop language arts and mathematical skills that are critical to success beyond high school.

### **Summary**

Overall, the results from the HSEE field test were quite positive. Notwithstanding the long test length, nearly all the students answered all of the items. The sample sizes, while less than hoped for, were adequate to provide stable estimates of both traditional and IRT item parameters. One limitation was the relatively modest number of African American, special needs, and ELL students who were tested, making it difficult to determine whether the items functioned differently for these groups.

Relatively few items had obvious statistical problems. This result confirmed results of direct observation that the item development and review process was thorough and effective.

Efforts to examine the potential accuracy of the HSEE scores, while very preliminary, were also reasonably positive. Even if the decision point were near the middle of the achievement distribution, 90% of the students taking an ELA form and 84% of the students taking a Math form would score more than one standard error above or below the decision point and these students would be classified (passed or failed) correctly at least 98% of the time.



One concern raised by the field test results was the relative difficulty of the items, particularly in mathematics. If these items reflect what we believe students need to know and be able to do, and several panels of reviewers believe that they do, then a significant number of 10<sup>th</sup> grade students are likely to fail this exam. Groups who traditionally score lower on assessments of student achievement will fail at higher rates. It will be important, therefore, to ensure that there are effective programs to help students at risk, both before and after their initial experience with this exam. It is possible that students will perform at higher levels during operational testing than they did on this field test where the results do not count. However, the very high completion rates suggest that nearly all students took the field test seriously.

## CHAPTER 5: PRINCIPAL AND TEACHER SURVEYS

### Background

Educational reform such as California's high school exit examination will exert an impact beyond just the receipt of a standards-based diploma. The reform will stimulate many changes throughout districts and schools, and will serve as a catalyst for those changes by providing feedback about students. In addition to the performance information, the assessment is seen as a way to influence and improve teaching and learning. Consequently, a key research issue is the relationship between the exit exam and teaching practices advocated by reform standards. One purpose of a thorough evaluation, then, is to find out about what is going on in the classrooms.

Surveys are one component of the evaluation method to examine such consequences and assess the impact of the HSEE over time. Two surveys were administered to capture baseline data: one for principals and another for teachers in the same schools. The principal survey requested demographic and background information about the school, students, and parents. The teacher survey emphasized classroom practices. Given administration of these surveys early in the HSEE development and implementation process, both principal and teacher surveys contained several open-ended items to allow the respondents to clarify their responses and to inform HumRRO of any misunderstandings or omissions we might have about the operation of California schools and their relationship to district and state operations.

The information collection and review conducted for the background report for the HSEE (see Chapter 2) were critical to formulation of guiding issues and questions for the surveys. The background report helped to establish the context for developing and implementing a graduation test by examining other states' experiences. Given the nature of this baseline data collection, using a small sample of California schools at a time when the exit examination is just being developed and pilot tested with another sample of schools, the surveys required direction for asking anticipatory types of questions. Because there is no immediate requirement for schools to implement the exit exam, the survey needed to allow for low levels of planning and preparation without attaching negative connotations to such levels. However, the researchers needed to provide a means to describe any early planning and preparation they did find. Based on HumRRO's prior experience during the pre-implementation stages of some major educational initiatives, an understanding of the process of "early and late planners and implementers" also was used to develop survey items.

### Survey Development

The following are preliminary questions used to collect baseline data and to address issues of interest:

1. What are current graduation and college-going rates for different demographic groups?
2. What specialty education programs are currently offered?

3. What is the extent and type of current preparation for the HSEE?
4. What degree of familiarity do schools currently have with the HSEE?
5. How familiar are schools with the State Content Standards?
6. What plans are underway at schools to prepare faculty, parents, and community for the first administration of the HSEE?
7. What activities have schools undertaken to prepare students for the first administration of the HSEE, including students with special needs and English-language learners?
8. How do schools anticipate addressing failures on the HSEE?
9. What are schools' predictions for first administration pass rates?
10. What are schools' predictions for the impact of the HSEE?
11. What are schools' predictions for influence of the HSEE on instructional practices?
12. What are schools' predictions for opportunity to learn and opportunity to demonstrate knowledge and skills by various student groups?

### **Sampling and Administration**

The goal for the sampling plan was to select districts for inclusion in the HSEE evaluation data collection efforts that would be as representative as possible. A complete description of the sampling procedure is presented in Chapter 1. The resulting sample for the principal and teachers surveys, as well as for the item review workshops, comprised 24 districts. An introductory letter from the State Superintendent of Public Instruction and a project “fact sheet” were sent to each district superintendent to provide information about the evaluation and to request cooperation with the effort. In HumRRO’s follow-up with the superintendents, they were asked to identify the principal, or other point-of-contact (POC), at the sample schools in their districts. Based on this information, principal and teacher survey packets were shipped in early May to 83 schools. Packets, which were sent to the attention of the principal or POC, included the following:

- Cover letter and instructions to principal
- One principal survey
- Cover letter and instructions to teacher
- Four teacher surveys—two labeled for English language arts and two labeled for mathematics
- Fact Sheet for California High School Exit Examination Evaluation
- Instructions and packaging for returning evaluation materials

Principals were asked to complete their questionnaire (or to designate someone to do so). They also were asked to identify, based on faculty size, up to two teachers of algebra 1 or other appropriate course, and two 9<sup>th</sup> or 10<sup>th</sup> grade language arts teachers to complete the teacher surveys. Each survey was contained in a sealable envelope to be returned to the principal for shipment to HumRRO. The cover letters to both the principal and the teachers encouraged respondents to contact a HumRRO project member if there were questions or concerns. A copy of the survey instruments is included in Appendix B.

Return of evaluation materials was requested by the end of May. Follow-up telephone calls were initiated the first full week of June with schools that had not responded, to check on the status of completing and returning their evaluation materials. This chapter reports preliminary findings based on returns up to June 19, 2000.

### **Initial Findings**

Surveys were completed<sup>8</sup> by 33 high school principals and 96 teachers. Preliminary results<sup>9</sup> are reported in the following areas:

- Background
- Knowledge
- Preparation Thus Far
- Future Plans
- Expectations
- Other

### **Background**

Principals were asked to provide some demographic information on themselves. Two-thirds of the respondents (22 of 33) were male; 70% were White; 8% Hispanic; and 6% declined to specify; 94% reported education beyond a bachelor's degree (6% some graduate school, 82% master's degrees, 6% doctorate degrees) and 6% responded "other." They were asked to identify their primary subject area when they were teaching; the responses varied widely. The most common subject was English (24%). The respondents reported between 1 and 30 years of experience as a principal (mean = 12.34, standard deviation = 7.79) and 4–33 years teaching (mean = 14.50, standard deviation = 7.78). They have worked 1–23 years in their present school and 5–38 years in public schools.

Teachers were also asked to provide demographic information. Over half (55%) of the respondent teachers were female; 84% were White; 7% were Hispanic; 5% were Asian/Pacific Islander; 1% were black; and 2% were other or declined to specify; 9% reported having only a bachelor's degree; most respondents reported education beyond a bachelor's degree (36% some graduate school, 43% master's degrees, 5% doctoral degrees); 6% indicated other education; 48% indicated that the primary subject area they taught was English or language arts; 45% specified mathematics as their primary subject area; and 7% indicated "other." Eighty percent indicated that their college training was in their primary subject area.

Principals were asked to provide background school information. The current number of teachers on staff ranged from 3 to 200, with a mean of 83 (standard deviation = 53). Principals reported that the percentage of teachers with advanced degrees ranged from 1% to 75%. Counselor-student ratios ranged from 1:1 to 1:1000, with a median of 400:1. Forty-

---

<sup>8</sup> These counts include all surveys received as of 06/19/00. Surveys received after this date will be included in the final version of this report. Open-ended comments made by principals were analyzed, but those made by teachers were not included in this preliminary analysis.

<sup>9</sup> All percentages will not total to 100% due to omitted responses to individual survey questions.

two percent of the responding schools currently have a testing coordinator; an additional 6% reported plans to have one by September 2000. Most schools (79%) operate on a semester basis; 12% configure their school year in quarters and 6% operate year-round schools. The majority of principals (73%) reported that their schools hold 6–7 academic periods per day. They reported, on average, a graduation rate of 80%, with varying rates by racial/ethnic group. Post-graduation attendance in two-year colleges averaged 29% and four-year colleges, 28%.

Principals were asked to indicate whether their schools offered various specialty education programs. Forty-eight percent offer remedial courses; 24%, magnet programs; 67%, special education; 46%, English-language learners; 15%, multicultural/diversity-based; 48%, Advanced Placement; 3%, International Baccalaureate; 33%, school/community/business partnerships; 30%, targeted tutoring; and 9%, other.

Teachers were asked to provide some information about their own classes. Asked to provide average enrollment per class period, responses ranged from 1 to 40, with a mean of 27 (s.d. = 6.3). Seventy-eight percent report that they create groups within classes for instruction. Of these, 65% assign students to these groups randomly; 11% use ability grouping; 7% allow students to choose their groups; and 18% indicated that they assign students to groups on some other basis. Twenty-five percent of teachers reported that 100% of their students were fluent English speakers; 47% indicated that 90–99% were fluent in English; 18% reported 75–89%; 6% reported 50–74%; and 1% indicated that less than 50% of their students were fluent in English.

Teachers were asked about various instructional practices. Thirty-eight percent of teachers require students to maintain a portfolio; an additional 12% indicated that they require another product in lieu of the portfolio. Three-quarters of teachers (77%) reported that students spend  $\frac{1}{2}$  hour or more of class time working with a partner or in a small group, on a weekly basis.

Teachers were asked to estimate the amount of time, on average, they believed students spend working on assignments outside the classroom each week. Half of the respondents (53%) estimated  $\frac{1}{2}$  to 3 hours; 21% estimated more than 3 hours; 18%, less than  $\frac{1}{2}$  hour; and 6%, none.

Teachers were asked to indicate the importance of specific instructional techniques. Techniques frequently endorsed as “very important” were: developing students’ abilities to make connections among content topics (78%), using questioning techniques to promote interaction and discussion (78%), using problem-solving as a means and a goal (77%), and using direct instruction (69%).

Teachers were asked to estimate how often they plan for students to participate in specific types of activities. The activities rated most frequently (once or twice a week or almost every day) were: do work from textbooks (86%), do work from supplemental materials (78%), apply subject area knowledge to real-world situations (74%), write a few sentences (70%), and work in pairs or small groups (69%).

## Knowledge

Principals and teachers were asked to report their familiarity with the HSEE and state content standards. The majority of principals (76%) responded that they had only general information about the exam. Eighteen percent reported that they were very familiar with the exam, while 3% expressed no familiarity. Teachers reported less familiarity with the exam than the principals: 14% claimed to be very familiar, 62% generally familiar, and 24% reported no familiarity. Because we asked principals to identify a small number of teachers to complete this survey, we wanted to determine whether these teachers were representative of teachers at the school. To this end, we also asked the teachers to estimate how familiar other teachers at the school were with the exam. Indeed, other teachers were rated as less familiar: 3% very familiar, 60% generally familiar, and 34% not at all familiar. This is an indication that the respondents may be more involved with the HSEE than typical teachers.

It is unsurprising that the level of familiarity with extant state content standards was higher than with the as-yet-unimplemented exam. Fifty-eight percent of principals said they were very familiar with the state content standards, while 36% reported general familiarity, and 3%, none. Teachers reported more familiarity with state content standards than did principals: 69% very familiar, 25% generally familiar, and 3% not at all familiar. As was the case with the question on familiarity with the HSEE, these teachers rated their own familiarity with state content standards as higher than they did other teachers' familiarity: 36% very familiar, 46% generally familiar, and 5% not at all familiar.

One possible source of information on the HSEE and state content standards for teachers could have been the HSEE Educator Panel Item Rating Workshops. We asked teachers whether they had participated in either of the May 2000 workshops; only 9% indicated that they had.

Respondents were asked to identify the source(s) of their information regarding the HSEE. Most principals indicated that their information came through official channels. Principals reported receiving information from: district-provided information (94%), state-provided information (73%), newspaper (54%), professional associations (46%), education organizations (33%), computer-based sources (24%), and other (6%). Three percent of principals indicated that they had no sources of information on the HSEE. Teachers reported that their information came from: school-provided information (62%), district-provided information (40%), newspaper (28%), state-provided information (24%), education organizations (14%), professional associations (12%), computer-based sources (5%), and other (8%). Ten percent of teachers indicated that they had no sources of information on the HSEE.

Principals were also asked to estimate how familiar their students and parents are with the exit exam. Responses indicated a belief that the exit exam is virtually unknown outside the educational community. No principals responded that students/parents were very familiar or familiar. Only 12% said that they were somewhat familiar; 48% indicated that they were not very familiar; and 36% replied that students/parents were not at all familiar.

## Preparation Thus Far

Although the HSEE has not yet been administered at any of the schools surveyed, we asked about preparation that has already been initiated. One precursor to a successful program is to align school curricula with the state content standards, to ensure that the test is testing what is being taught. Thus respondents were queried about alignment with state content standards. In short, principals indicated that most are already moving in the direction of alignment, but still have a way to go. Principals reported that 100% of their districts/schools encourage use of the content standards to organize instruction, and 79% said their schools are in the process of aligning their curricula to the standards. Fifty-two percent said that their schools/districts have plans to ensure that all students receive instruction in each of the content standards. Twenty-seven percent stated that their textbooks do not align well with the content standards; 36% report that they can cover all the content standards with a mix of textbooks and supplemental material.

Along similar lines, respondents were asked to compare their district standards and the state content standards. Most principals (73%) responded that their districts have adopted the state standards, and another 21% reported that their district standards include more than the state content standards. Thus, a total of 94% indicated that their district standards encompass all state standards. However, 3% reported that the state standards include more than the district standards, and 3% indicated that they could not judge. No respondents indicated that the two sets of standards were different, nor that their districts had no official standards.

Respondents were asked how much time they personally spent during the 1999–2000 school year in activities related to the HSEE (e.g., meetings, discussions, curriculum review, professional development). Most principals reported spending 6–15 hours (52%) or 16–35 hours (27%). Eighteen percent reported fewer than 6 hours; 3%, more than 35 hours. Most teachers reported fewer hours than principals: 19% none, 62% fewer than 6 hours, 9% 6–15 hours, 4% 16–35 hours, and 4% more than 35 hours. Teachers were also asked to estimate the total 1999–2000 time they spent on classroom instruction activities related to the HSEE (e.g., department planning, student preparation, curriculum review). A greater amount of time was reported for these activities: 19% none, 45% fewer than 6 hours, 16% 6–15 hours, 7% 16–35 hours, and 10% more than 35 hours.

Respondents were asked to identify the specific activities they have undertaken to prepare students for the first the HSEE administration. Although the students who will participate in the HSEE have not yet entered the 9<sup>th</sup> grade, most principals reported initiating some activities; only 18% indicated that they have implemented none. Figure 5.1a indicates the percentage of principals who reported implementing each activity; Figure 5.1b indicates teachers' responses. In general, fewer activities were reported by teachers; 33% indicated that none had taken place. This may mean principals were aware of some individual teachers implementing activities even though implementation was not schoolwide.

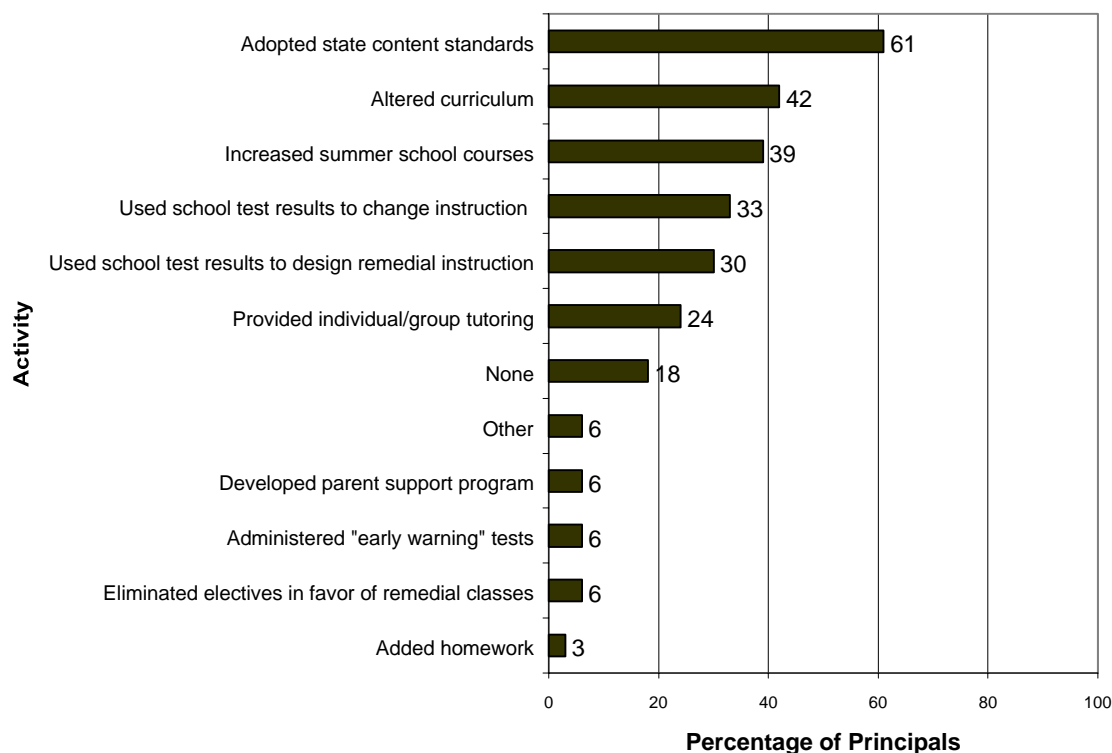


Figure 5.1a Percentage of principals reporting activities already underway to prepare students for the HSEE.

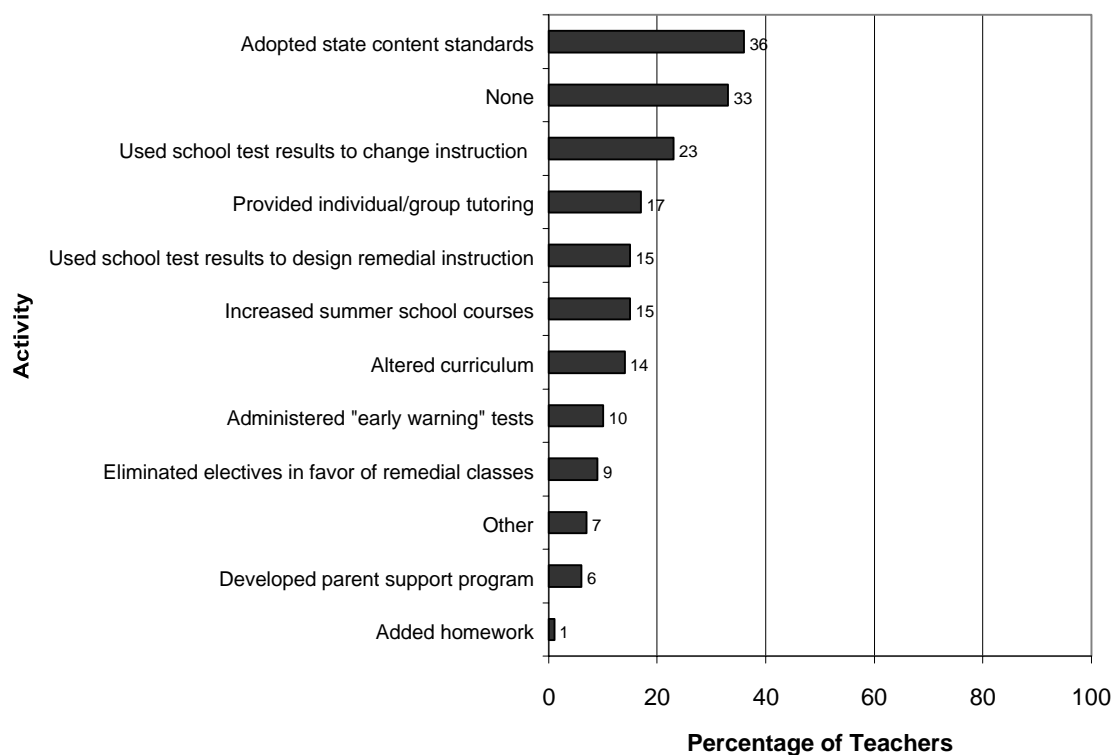


Figure 5.1b Percentage of teachers reporting activities already underway to prepare students.



## Future Plans

In addition to any preparatory steps taken thus far, the surveys inquired about future plans to deal with this new requirement. In particular, efforts to prepare teachers and others for the exam, to prepare Individual Education Plans [IEPs] for special education students, and remediation plans subsequent to the first exam administration were probed.

Principals were provided a list of possible remedial practices and asked which they planned. Figure 5.2a lists the percentage of principals who endorsed each activity and Figure 5.2b reflects teacher responses to the same question. Similar to the pattern of preparatory steps, more principals reported activities than did teachers. For example, 15% of principals indicated that no plans had been made for remediation, compared to 36% of teachers.

Approximately half of the 40 open-ended responses on “plans to prepare staff, parents, and the community for the initial exam administration” cited plans for staff-related efforts such as department and faculty meetings, inservice training, and content and curriculum workshops. A third of the responses mentioned public outreach, parent communications, and general dissemination of information about the exam. Several respondents (8%) stated that they are waiting for direction from CDE—specifically to rule on staff development days that are not “buy back days.”

For principals, almost 30% of the 34 open-ended responses on “plans to work with students who fail the initial exam administration” reiterated that no plans had been formed yet, or that the schools were waiting on district plans or were waiting for the exam itself to plan. Half of their comments mentioned plans to notify parents and to offer tutoring or other practice, expanded summer school and reading programs, and development or modifications of remedial and exam support courses. Among the remaining responses were some specific plans such as (a) revising a student’s four-year high school plan to improve the areas of weakness, and (b) implementing a Fall 2000 mandatory parent and student orientation and administration of diagnostic tests in mathematics, reading, and writing.

Over half of the principals’ 21 open-ended response on “plans or strategies to prepare for IEP changes that will allow participation of students with disabilities” stated they had made no plans yet or that they will develop a plan according the law. A fourth of the comments said they would continue to follow the IEP recommendations for accommodations. Among the remaining responses were some specific plans such as (a) implementing a Fall 2000 plan to identify special needs students who are likely to participate in the exam and noting what accommodations will be needed, (b) starting to expose special needs students to algebra, and (c) including special needs student in other HSEE efforts.

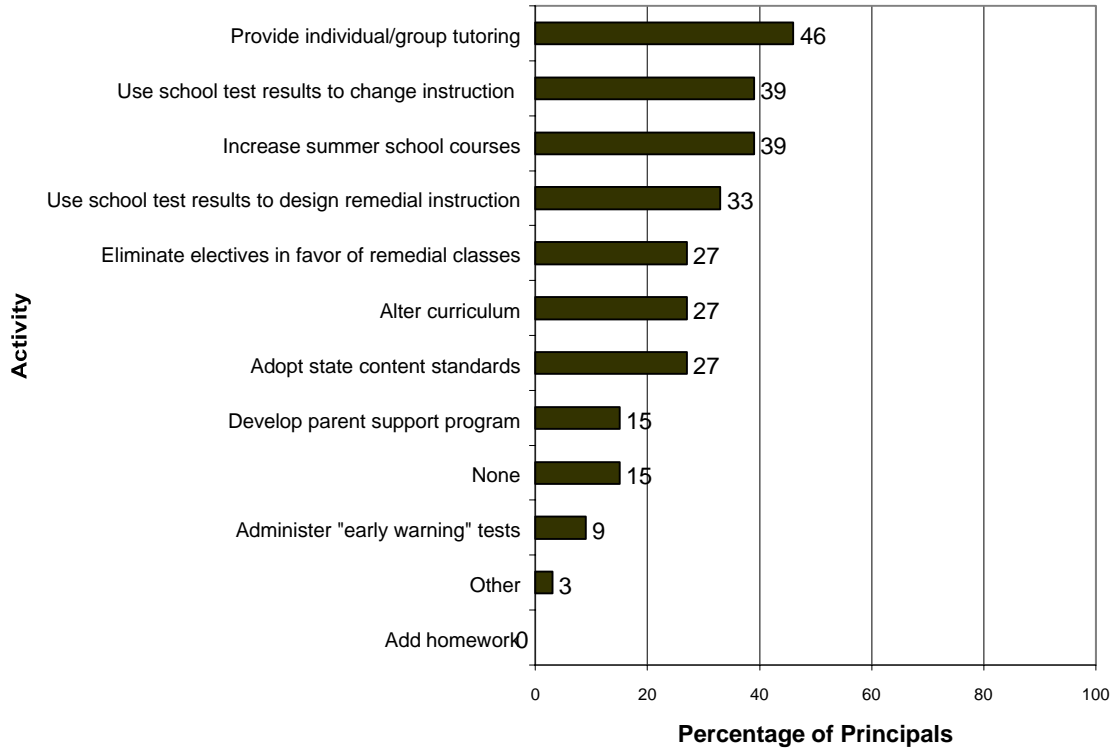


Figure 5.2a Percentage of principals reporting plans for remediation of students who do not pass the HSEE.

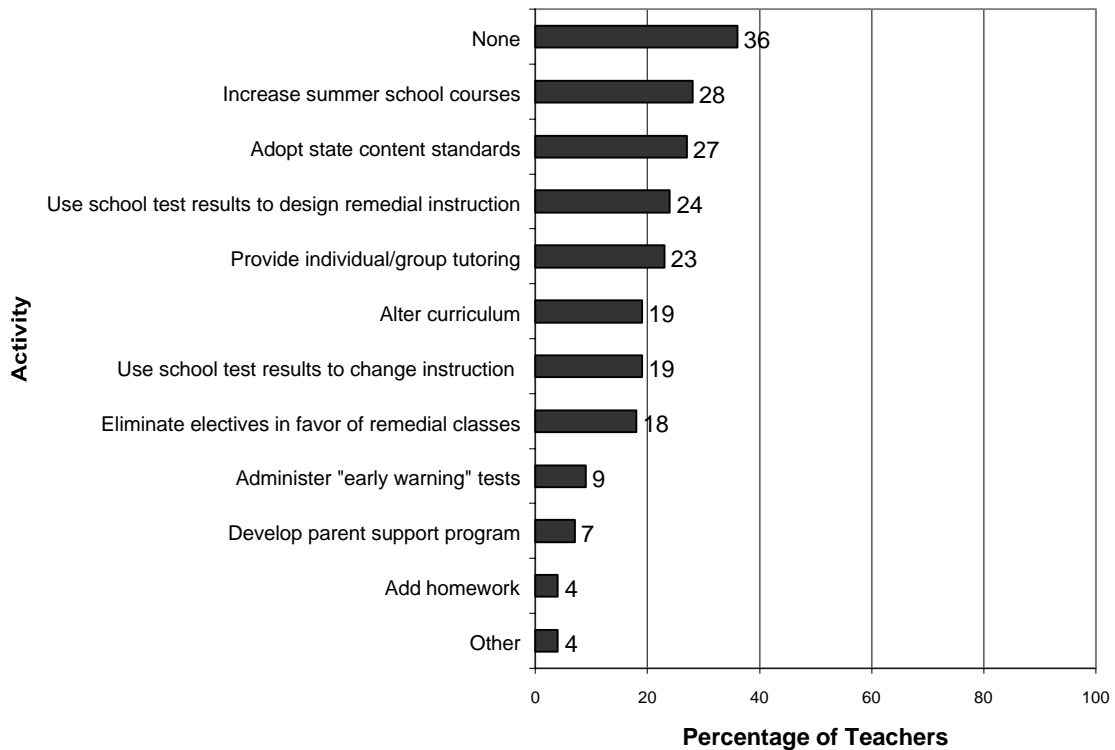


Figure 5.2b Percentage of teachers reporting remediation.

## Expectations

Several survey items queried the respondent's expectations for the exam: anticipated pass rates, impact of the exam on student motivation and parental involvement, and so on.

Principals were asked to estimate the percentage of current 10<sup>th</sup> grade students (Class of 2002) who would deserve a passing grade on the upcoming exam.<sup>10</sup> Responses were generally guarded. Nearly half (48%) of principals predicted that fewer than 50% of students would pass the exam; 30% predicted 50–74% of students would pass; 15% predicted 75–95%; and 3% of principals predicted that more than 95% of students would pass.

Teachers were asked two variants of the same question. They were asked to estimate the preparedness of students to pass the HSEE in the 9<sup>th</sup> grade and in the 10<sup>th</sup> grade, based upon the teacher's knowledge of the feeder schools. Twenty-one percent of teachers responded that students were prepared (or better) in the 9<sup>th</sup> grade; 46% indicated that students were prepared or better in the 10<sup>th</sup> grade. The responses were similar to those of the principals.

Principals and teachers were also asked to predict the impact of the HSEE on student motivation and parental involvement, under various circumstances. Figures 5.3a and 5.3b reflect the impacts anticipated prior to administration of the exam. Principals predicted a wider variety of impact on student motivation than on parental involvement. Some negative impact on student motivation was predicted prior to the exam, but largely neutral or positive effects were posited for parental involvement prior to the first administration. Comparison of Figures 5.3a and 5.3b indicate that teachers are somewhat more pessimistic than principals about the impact of the HSEE on student motivation and parental involvement.

---

<sup>10</sup> Note that this cohort will not take the exam; the first class to participate will be the Class of 2004, which is now entering the 9<sup>th</sup> grade. Because the first participating group is not yet in high school, principals were asked to assess current 10<sup>th</sup> graders (Class of 2002) as a proxy for the Class of 2004.

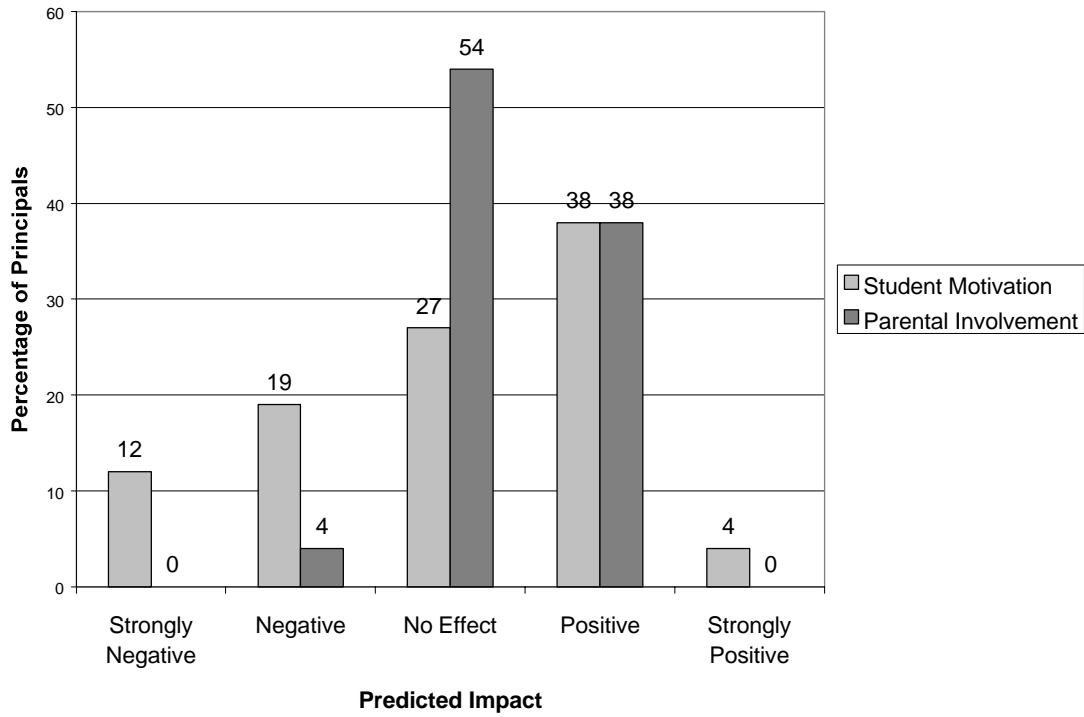


Figure 5.3a Principals' predicted impact of the HSEE on student motivation and parental involvement prior to taking the exam for the first time.

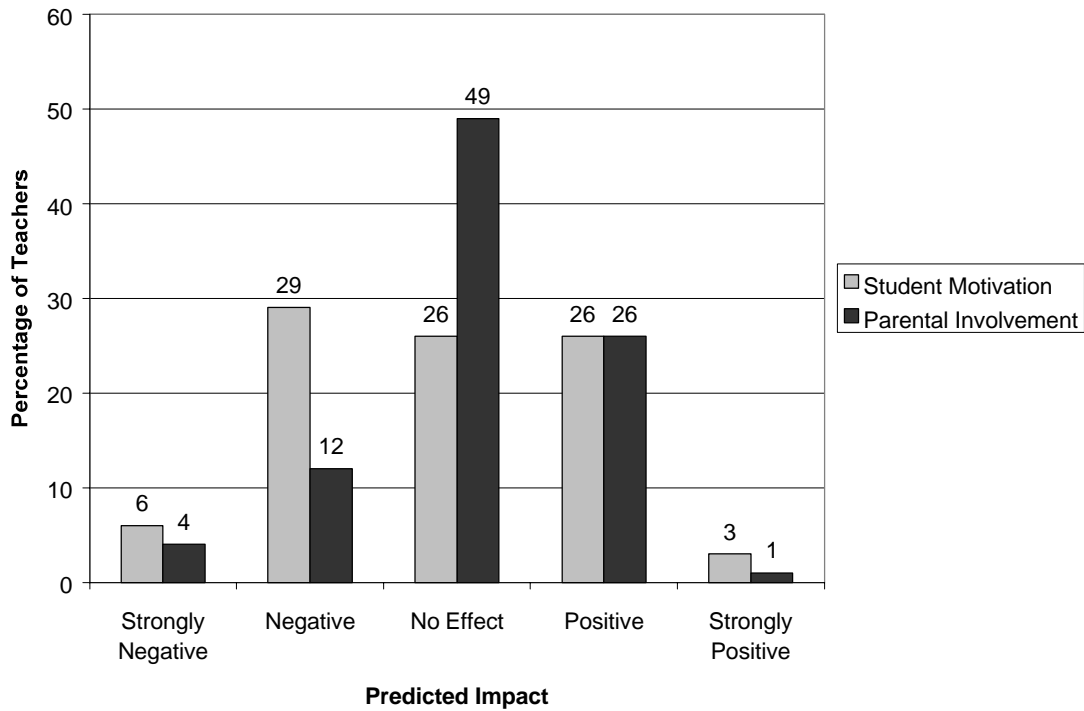


Figure 5.3b Teachers' predicted impact of the HSEE on student motivation and parental involvement prior to taking the exam for the first time.

Principals and teachers were asked to predict the same two concepts—student motivation and parental involvement—for those students who pass the exam in the first administration. The predictions for this group were more positive. Only 4% of principals expected that student motivation would drop after students cleared the hurdle of the HSEE. Thirty-eight percent of principals predicted that student motivation would be unaffected by passing the exam; 58% predicted a positive or strongly positive effect. Half of principals expected no impact on parental involvement; 35% predicted a positive effect and 15% a strongly positive impact on parental involvement for those students who pass the exam early in their high school careers. Here again, principals were more optimistic than teachers. Three percent of teachers expected a strongly negative impact and another 12% anticipated a negative impact on student motivation after passing the exam on the first attempt. Thirty-five percent of teachers predicted that student motivation would be unaffected by passing the exam; 42% predicted a positive or strongly positive effect. Half of teachers (49%) expected no impact on parental involvement; 9% expected a negative or strongly negative effect; 29% predicted a positive effect and 5% a strongly positive impact on parental involvement for those students who pass the exam early in their high school careers. Eight percent of teachers declined to estimate the impact of student motivation or parental involvement.

For those students who fail the exam on the first try, the principals' and teachers' predictions were quite different. Principals were split on whether the impact of failing the exam would have a negative effect on student motivation; 4% predicted a strongly negative effect; 31%, negative; 19%, no effect, and 42%, positive. No principals predicted that an early failure would have a strongly positive effect on student motivation. Predictions for parental involvement were very similar to those of student motivation: 4% predicted a strongly negative effect; 31%, negative; 15%, no effect; 46%, positive; and 4%, strongly positive. There was a similar pattern for teacher responses, albeit slightly more negative overall: regarding student motivation, 8% predicted a strongly negative effect; 27%, negative; 14%, no effect, 38%, positive; and 4%, strongly positive. As for parental involvement, 7% of teachers predicted a strongly negative effect; 19%, negative; 27%, no effect; 36%, positive; and 1%, strongly positive.

Principals and teachers were also asked to predict the impact of the HSEE on student retention and dropout rates. Responses were somewhat negative overall. Predictions followed a similar pattern on both questions. Fifty percent of principals anticipated a strongly negative or negative impact on student retention rates; 57% predicted a strongly negative or negative impact on student dropout rates. Thirty-one percent predicted no effect on student retention and 27% predicted no effect on student dropouts. Nineteen percent anticipated a positive or strongly positive effect on student retention rate and 16% expected a positive or strongly positive effect on student dropout rate. Teachers responded very similarly to principals, although as in previous questions, their answers were slightly more negative.

Principals were asked to predict, based on what they knew about their schools, the influence of the HSEE on instructional practices. Responses ranged from moderately optimistic to neutral: 79% responded that practices would be improved and 9% predicted no effect. No respondents chose the options of strongly improved, weakened, or extremely weakened.

Teachers were asked the same question about the influence of the HSEE on instructional practices, but they were asked to provide separate estimates for three school years. Figure 5.4 provides the responses for all three years. The pattern of responses indicates that teachers expect the HSEE to have a positive impact on instruction, and they expect that impact to grow increasingly positive over time.

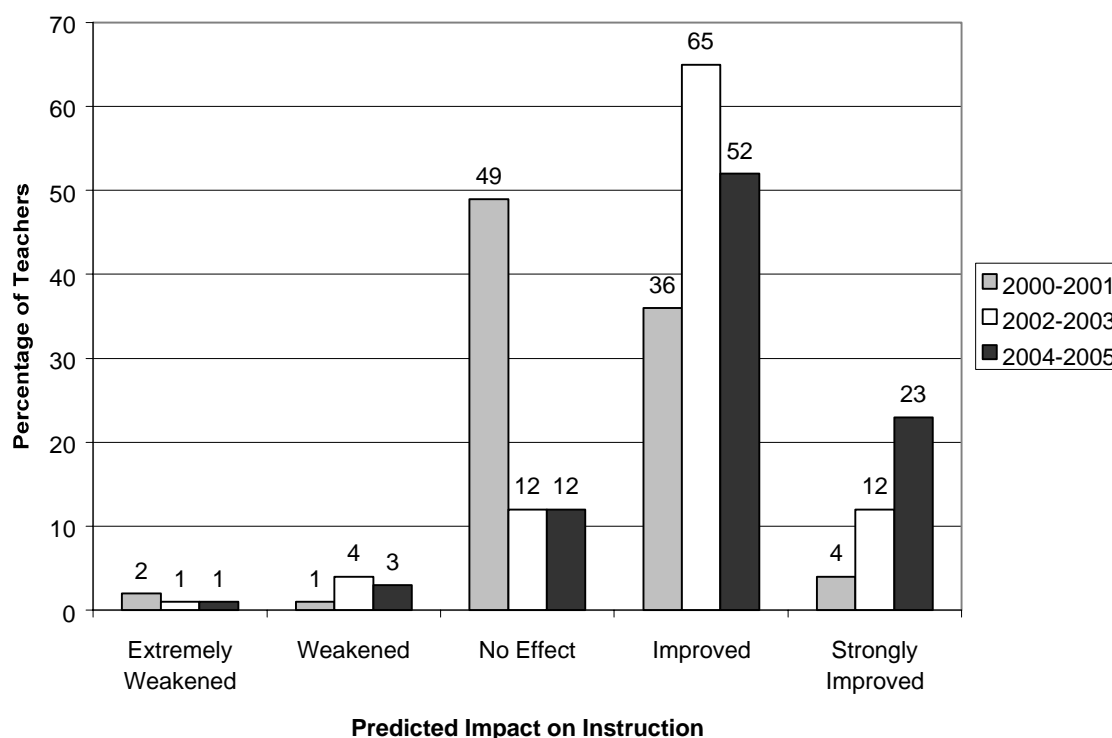


Figure 5.4 Teachers' prediction of influence of the HSEE on instructional practices over time.

One of the concerns when implementing a new exam is whether there is a differential impact on various subgroup populations. We asked principals and teachers to predict the opportunity to learn the material covered by the exam for the total student population, as well as for specific subgroups. Three percent of principals indicated that were not sure how they expected all students to have an opportunity to learn. Eighteen percent reported an excellent opportunity to learn; 24% selected good; 36%, adequate; and 12%, poor. No principals reported "no opportunity" to learn.

These same questions were asked about four other groups: students with disabilities, English-language learners, English-language learners in targeted subject areas, and economically disadvantaged students. The predictions were slightly more negative for the targeted groups; the predictions of poor opportunity to learn increased from 12% for all students, to 30% for students with disabilities, 30% for English-language learners, 24% for English-language learners in targeted subjects areas, and 18% for economically disadvantaged students. Comparison of principal responses and teacher responses revealed similar patterns.

We asked principals and teachers a similar set of questions regarding students' opportunity to demonstrate their knowledge and skills on the exam. For the full student population, 3% of principals expressed that they were unsure; 9%, excellent; 30%, good; 39%, adequate; and 9% poor. No principals selected a response of "none." For the various student subgroups, responses were less optimistic; a poor opportunity to demonstrate knowledge and skills was anticipated for students with disabilities by 27% of principals; English-language learners, 27%; English-language learners in targeted subject areas, 24%; and economically disadvantaged students, 18%. Teachers provided similar responses, although teachers estimated the proportion of each group having a poor opportunity as about 2–3 percentage points higher, across the board, than did principals. The sole exception was for the category of students with disabilities; teachers were more optimistic than principals, predicting that 24% would have a poor opportunity.

## **Challenges**

When principals were asked to "describe challenges facing the school and students in successfully meeting the exam requirements," 30% of the 30 open-ended responses commented on existing low levels of student competency and skills of incoming high schoolers—especially for Continuation and Community Day schools. Also 30% of the comments described alignment issues, and 13% referred to meeting algebra and English language arts proficiencies—especially for English-language learners. Of the 20% who cited time requirements and the burden of testing, two comments particularly captured this challenge and underscore the lack of knowledge about the purpose of the test:

"We test too much behavior SAT9, SAT, ACT, Golden State, exit exam, end of course exams, A.P. When do we teach? It will take up almost the whole month of May—can we combine any of these tests?"

"We will offer a summer remedial program for 9<sup>th</sup> graders. We will visit the homes of the incoming 9<sup>th</sup> graders; [and we] will provide tutoring, [but] I think the testing system is too fragmented—too thinly spread out to be successful."

In describing "benefits to the school and students associated with the exit exam" two-thirds of the 19 comments cited having students meet a standard of basic skills in English and mathematics before leaving high school. The remaining responses were split between those placing a focus on curriculum and those who said there were no benefits or they were unsure about any benefits.

## **Other**

Principals were asked to add any comments about specific factors at their schools that they felt would influence the exit examination. Of these 17 rather extensive entries, half described schools operating with students at the poverty level, with low academic preparation, and with disengaged parents. They also expressed concern that the exit exam will result in increased dropout rates. Two comments reiterated concern about the burden of adding one more test to an already challenging schedule. Two comments focused on the pilot test items. One of these stated that the items are very White, middle class and not representative of a diverse student population. The other objected to the group proportions

used in the pilot testing as over-representing special education and minority students and under-representing Caucasian students. They feared that the test results will not be a true reflection of their predominantly Caucasian school.

### **Summary**

This preliminary analysis was conducted on a subset of the survey sample (i.e., surveys received by 06/19/00), therefore results should be interpreted with caution. Several points, however, stand out. Unsurprisingly, principals and teachers agree that they are more familiar with state content standards than with the HSEE. Principals rated themselves as more familiar than teachers rated themselves. These teachers, in turn, rated themselves as more familiar than their peers. This latter point may indicate that the sample of teachers who responded to the survey may be more knowledgeable about the HSEE than the typical teacher; this should be kept in mind when generalizing from these responses.

Some principals and teachers reported that they had no source of information on the HSEE. Most relied primarily upon official channels such as state sources and district sources; teachers reported a greater reliance upon newspaper accounts than did principals. Principals believed that students and parents are largely unfamiliar with the HSEE at this time.

Some preparatory activities have already begun. For example, many districts have made an effort to align their content standards with those of the state. The vast majority of principals indicated that their district content standards encompass all state content standards. Principals reported more preparatory activities than teachers did; a third of teachers were unaware of any preparatory activities thus far.

In addition to adopting the state content standards in preparation for the HSEE, most principals reported the importance of plans for preparing staff such as curriculum workshops, and inservice training. Most principals also reported initiating some type of activity to prepare students for the first administration of the HSEE such as altered curriculum and increased summer school courses. A third of the teachers, however, reported having no activity underway at the present specifically to help students prepare for the test.

Student preparedness estimates were mildly pessimistic; in general, principals provided slightly more optimistic predictions than did teachers. Some concern was expressed that students arrive at high school unprepared, and that elementary and middle schools must become involved in the process of preparing students for the HSEE.

Teachers and principals were in basic agreement about the impacts of the test in various situations. For example, predictions of the impact of the HSEE on student motivation and parental involvement, prior to the first administration, were neutral-to-mildly positive. For those students who pass the exam on the first attempt, school personnel expect that the effects on both student motivation and parental involvement will be positive or neutral; this expectation runs counter to the concern that students may lose motivation if they clear the exam hurdle too soon in their high school careers. For those students who fail on the first attempt, however, expectations are different. Relatively few respondents predicted that



failure would have a neutral effect on student motivation, but two camps emerged: nearly the same number of respondents expected a negative or strongly negative impact as predicted a positive impact. Principals and teachers were very consistent in their prediction that the effects of the HSEE upon student retention rates and students dropout rates will be negative.

Despite these concerns about the effects on student motivation and parental involvement, principals and teachers expected that the impact of the HSEE on instructional practices will be positive. Further, teachers were asked to estimate effects next year and in 3 and 5 years; they predicted greater improvement with time.

Respondents expect differential impacts for certain student subgroups. They anticipate that opportunity to learn will be lower for English-language learners and students with disabilities than for the student population as a whole. Fewer respondents believed that these differences will be seen with economically disadvantaged students.

In short, the preliminary analysis of surveys received to date indicate:

- A need for more information on the exam and staff development to support its implementation;
- Concerns about student preparedness;
- Mixed predictions about the impact of the exam on student motivation;
- Concerns about the impact of the exam on retention rates and dropout rates;
- Concerns about the success of disadvantaged groups, especially English-language learners and students with disabilities; and
- Positive expectations of the impact of the HSEE on instructional practices.

## CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS

In Chapters 1 through 5, we described the main activities conducted in the first year of the HSEE evaluation and summarized the results of these activities to date. In this chapter, we present the main conclusions that we draw from these results and list recommendations for further HSEE development based on these conclusions.

### General Conclusions

*Conclusion 1. A great deal has been accomplished since the HSEE development was launched.*

As specified in the enabling legislation, an *HSEE Standards Panel* was appointed by the Superintendent with review and approval by the Board. The Panel has met monthly and has reviewed and discussed a wide range of issues from test specifications to administration (particularly required accommodations) and reporting issues. Along with a technical committee of content and measurement experts, the Panel recommended test specifications incorporating additional information on test blueprints proposed by the development contractor.

*Contracts were let* for both the development and the evaluation of the HSEE. This was not a small accomplishment, as contractors were initially hesitant to bid on the development effort due to its ambitious demands and tight timelines. In the interim, Department staff worked hard to support the HSEE Panel and identify appropriate items to keep the project on schedule. Since the contracts were let in January and February of this year, Department staff have provided oversight and guidance to each effort and have continued to be heavily engaged in coordinating the efforts of the contractors, the Panel, and others.

A large number of *test items have been developed* and shepherded through an extensive review process. The process included one or more rounds of review for technical, editorial, bias and sensitivity, and policy issues, including a review by the HSEE Panel. A total of 362 English Language Arts items and 396 Mathematics items survived this process and were included in field test booklets.

The *field test was conducted* as scheduled. The development contractor identified a broadly representative set of schools, recruited participation, and ended up testing more than 7,500 10<sup>th</sup> grade students from these schools. Each student completed a lengthy test booklet containing 99 or 102 test items and nearly all of the students answered all or nearly all of the items.

*Conclusion 2. Development efforts to date have been highly successful.*

Agreement was reached by the Panel, the developers, and the department concerning the detailed *contents of the exam*. The contents that will be proposed for Board adoption reflect a hard-won consensus among the different perspectives and positions reflected on the Panel and across other stakeholder groups.

The quantity and, most importantly, quality of the items developed to date are high. The HSEE Panel and its Technical Committee reviewed these items and agreed that they were each acceptable measures of the targeted standards. We conducted an independent review of all of the items by panels of teachers and curriculum specialists and, with few exceptions, they agreed that the items were reasonably aligned to the proposed test standards.

Relatively few items were flagged for statistical problems in the field test. A high proportion of the items were within an acceptable range of difficulty, differentiated between high and low performing students as measured by the other items, and did not function differently for minority students. The positive results from the field test validate the development and review process conducted by the contractor and the Department.

Preliminary evidence suggests that scores based on the items developed to date would probably lead to accurate pass/fail decisions across a range of plausible minimum scores. (See the analyses reported in Chapter 4.) Ultimately, however, decisions about the passing scores and the required level of accuracy in classifying students with respect to these scores require policy judgments.

*Conclusion 3. Much more needs to be done before operational administration can begin.*

Several important questions about the exam and its use remain to be answered. Some of the possible answers would appear to make operational administration in winter and spring 2001 infeasible. Some of the more critical questions are:

1. *Will the test blueprints, specifying the number of items to be included for each of the targeted standards, be approved as currently recommended?* If significant changes are made, it may be too late to develop and field test items in time for inclusion in the 2001 test forms.
2. *How many different test forms will be needed for the 2001 administrations?* Separate March and May administrations are planned, in order to accommodate school schedules and year-round schools, in particular. Thus a minimum of two test forms will be needed, and one or more backup forms may also be needed in case of security breaches. Based on the items developed to date, it may be possible to develop two independent forms, but if more forms are needed, it is not clear that current deadlines can be met.
3. *How will the minimum passing score be determined?* It seems unlikely that this question can be answered until one or more operational test forms have been assembled. The procedures used with the National Assessment of Educational Progress and many state assessments where standards-based reporting is employed require some normative information. At some point in the process prior to final approval, information on consequences, meaning the proportion of students who will pass, is examined. Unless reporting of the 2001 results is significantly delayed (and the legislation requires reporting within 8 weeks), the passing standards will have to be set before the operational test form(s) are administered to any students, let alone a

representative (norming) sample. It may be possible to “simulate” passing rates based on individual item data from the field tests. Such simulated passing rates require a number of statistical assumptions (more complex than the assumptions underlying sampling with the U.S. Census, which were rejected by Congress). There are no suggestions on the table for collecting normative data for use in setting passing scores and still meeting the schedule for the 2001 administrations. We have no suggestions either.

4. *What information, if any, will be reported to students, parents, and teachers in addition to the overall pass or fail outcome for English/Language Arts and for Mathematics?* Time will be required to develop and test procedures for printing score reports, so final decisions about the content of these reports will likely be needed two or more months before the first administration.
5. *How will scores from different test forms be put on a comparable scale?* A minimum of two forms will be needed (i.e., one for March 2001 and another for May 2001, and possibly backup forms will be required in case of security breaches. Due to inevitable minor differences in item difficulty, it is not reasonable to assume that a given number correct score represents the same level of proficiency for different test forms. Equating based on common items and/or administration of the different test forms to strictly comparable samples of students is required to equate the scores from these forms. Equating plans must be developed, reviewed, and approved before additional test forms can be assembled, meaning well in advance of operational use of these forms.
6. *What specific accommodations will be provided on each of the two tests for special needs students and for English-language learners?* Final approval of proposed accommodations (as specified in EC 60850g) is needed along with evidence that scores for students tested with these accommodations are comparable to scores for students with similar proficiencies who do not require such accommodations.

Even if the answers to these questions do not render current plans for the first operational administration infeasible, the schedule is daunting. A second field test scheduled for this fall is needed to fill in some gaps in coverage of the test standards and to support the construction of additional test forms. Given the need to complete the fall field test, it seems unlikely that the first test forms can be assembled before December, and final copy for these forms will probably need to go to the printers in January. This leaves little time for review or any other possible delays.

*Conclusion 4. Available evidence suggests reason to be concerned about how well students are currently prepared to meet the standards of the assessment.*

Feedback from principals and teachers, ratings by our panels of educators, and the percent of students who answered each item correctly in the field test all suggest that half or more of today's 10<sup>th</sup> graders would fail one or both of the HSEE tests, particularly the mathematics test. On the one hand, this evidence underscores the need for the HSEE or some other interventions to raise levels of achievement relative to the targeted standards. On the

other hand, if a significant number of high school students are denied a diploma on the basis of this exam, there are likely to be many negative, as well as positive consequences.

Results from the field test were consistent with 1999 STAR results and results from other examination in suggesting that minority students, special needs students, and English-language learners are particularly at risk of failing the HSEE.

### **Recommendations**

In addition to reporting on consequences of the new HSEE requirements, the Education Code (60855c) requires the evaluators to “include recommendations to improve the quality, fairness, validity, and reliability of the examination.” Based on the results of the evaluation to date, as summarized in the general conclusions above and the more detailed findings in the body of this report, we offer one general and three more specific recommendations at this time. In making these recommendations, we are mindful of the need to maintain the independence of this evaluation. To the extent that evaluators become proponents for specific plans, they lose at least the appearance of independence when evaluating these same plans. Consequently, we will avoid proposing our own plans for processes where the Department, the Panel, and the development contractor have not yet finalized their own plans (e.g., setting minimum passing scores). We shall have more to say about those processes in subsequent reports, after initial plans are proposed by others.

#### **General Recommendation**

*The State Board of Education, Legislature, and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.*

There are two reasons why it might be important to delay full implementation of the exit examination, as was done recently in Maryland. First, as suggested by the list of decisions yet to be made in Conclusion 3 above, it may take more time to develop an assessment that fully meets professional and legal standards. For example, the timeline for reaching a decision on the minimum score needed to pass the test seems particularly problematic. A hasty or ill-informed decision about minimum passing scores will be hard to correct once students take the exam in 2001. The need for careful consideration and analyses of accommodations offered to special needs students is another point at which additional time prior to full operational implementation may be needed. The current schedule offers little time to develop the accommodations and no time for additional research on the comparability of scores across different possible testing accommodations. We note that established testing programs with which we are familiar (e.g., the Medical College Admissions Test and the Armed Services Vocational Aptitude Battery) spend three to four years on the development of each new test form. To complete the development of operational HSEE test forms in 18 months, with most of the first 6 months spent in developing test specifications (which are even yet not final) is ambitious.

The second, and we think more important, reason for considering a delay is that schools will need more time to prepare students to meet the standards assessed by the HSEE. As

described in Chapter 2, the key legal issue in prior challenges to high stakes test is whether students have been provided adequate instruction in the material covered by the test. Current plans call for students and schools to be fully notified about the exam and its requirement this fall, as the first affected class (the Class of 2004) enters 9<sup>th</sup> grade. This will be too late to allow very significant changes in the 9<sup>th</sup> grade curriculum for these students. In addition, the 7<sup>th</sup> and 8<sup>th</sup> grade curriculum must be aligned to prepare students for new curriculum changes at the high school level and because some of the standards being assessed are taken from the 8<sup>th</sup> grade standards. The need for more time to prepare seems particularly acute for special needs students, as updating IEPs to cover the contents of the exam will take some time.

Evidence that a significant proportion of today's students would not pass the test summarized under Conclusion 4, suggests that schools may need more time to provide adequate preparation for all students and to develop and implement remediation programs for students who do not initially pass.

With a delay of at least a year, a fully developed form could still be administered to 9<sup>th</sup> and 10<sup>th</sup> grade students in a representative sample of schools in 2001. The result of this would be important information for setting passing rates and, after the rates are set, important feedback to participating schools about how well their students are prepared for the exam<sup>11</sup>. This would provide schools both information and time to implement further improvements to the curriculum before holding the students, themselves, fully responsible.

### **More Specific Recommendations**

The process of developing and reviewing items appears to be working well and we have no specific recommendations for improving these processes at this time. Plans for assembling and equating different test forms and for setting minimum passing scores are still being developed and so we will reserve comment until a later report. At this time, we offer specific recommendations in three areas: (1) clarifying the relationship of the HSEE to other statewide testing programs, (2) communicating with district and school personnel about the HSEE, and (3) gathering information specific to the HSEE on the appropriateness and effectiveness of different testing accommodations.

*Specific Recommendation 1. The Department and the Board need to work together to clarify the relationships and differences among the different high school testing programs, most notably the HSEE, the standards-based STAR assessment, and the Golden State Examinations.*

Responses to our surveys of principals and teachers indicate that the fact that the HSEE and the standards-based STAR assessment are both designed to test student achievement relative to the same content standards is confusing, at best. Concerns about coordination of these different testing programs were also expressed at the June meeting of the State Board. One important difference is that the HSEE is targeted at levels of proficiency considered to be the minimum for high school graduation while the standards-based STAR assessment

---

<sup>11</sup> Schools will get important feedback about the performance of their students relative to the broader set of state curriculum standards from the STAR standards-based assessments. The STAR results do not, however, include a passing score, so the impact of the HSEE requirements will not be clear.

covers a wide range of proficiency levels from very basic through advanced. One area where clarification would be helpful to districts as well as to state-level policy-makers is whether it would or would not be reasonable to relate the reporting scales for these two programs. For example, can parents and teachers use a student's scores on the STAR exam to anticipate his or her results on the HSEE exam?

In addition to clarifying policy, design, administration, and reporting differences among the different high school testing programs, the Department should consider the feasibility of assembling a database that would allow for comparisons of the scores of individual students on these different examinations. Such a database would provide important empirical information on the relationships among the examinations to complement the policy and design distinctions.

*Specific Recommendation 2. The Department and Board should establish, expand, or accelerate processes for communicating with local districts about the HSEE and supporting their preparation for its implementation.*

Information from our surveys of principals and teachers and discussions with the panels of educators who participated in our item review workshops indicated a strong desire for information about the HSEE at the district and school level and some confusion about how the new program will be implemented. We believe that it will be important for the state to establish a consistent means of communication with districts and schools that reduces the sense of “insiders” and “outsiders” with regard to the flow of information about the exit examination. By consistent we mean coordination at the state level so different messages from the Department, the State Board, and others do not trickle down to the districts. In designing a communication program, state personnel should seek to anticipate concerns or misunderstandings and provide the district and school leadership with information that can be used to address confusion or turmoil. While there may not always be a definitive answer or resolution that can be offered, an open, straightforward response acknowledges the importance of the educators' stake in the exit examination.

A consistent communication program would also be useful for facilitating communication among the districts, as well as between state and district levels. CDE might work to identify promising practices of “early planners” to provide models for other districts and schools. Examples of district-level practices that might be included are: (a) developing community outreach programs and parent communication systems, (b) implementing effective staff development strategies, and (c) initiating academic supports to students. In addition, information from the surveys indicated a desire for concrete support from CDE with staff development and plans to prepare students, parents, and the community for the initial exam administration.

*Specific Recommendation 3. The Department and development contractor need to gather, review, and discuss more information on the appropriateness and effectiveness of testing accommodations for special needs students and English-language learners.*

The HSEE Panel has heard presentations on a number of issues relating to testing accommodations. It is likely that the effectiveness of different accommodations in allowing for the inclusion of all students and the appropriateness of the accommodations for generating comparable scores will vary greatly across testing programs. Information on the appropriateness and effectiveness of different accommodations on the HSEE itself is needed. A limited number of special needs students and English-language learners were included in the Spring 2000 field test, but we are not aware of any plans to use field test data to analyze the impact of testing accommodations. If the current schedule is maintained, the fall 2000 field test will be the only opportunity to collect information on testing accommodations prior to the first operational administration. If the schedule is relaxed, there would be more significant opportunities in a spring 2001 test administration.

In collecting information on the effects of different accommodations, it will be important to collect information on the students who are not tested as well as those who are. How do students who cannot be accommodated with available alternatives function in the classroom? What sort of accommodations might be considered in developing different forms of assessment for these students? In addition, it would be highly desirable to include planned variation in the accommodations offered to different students so that the impact of these accommodations on the scores of students without special needs can be investigated. We recognize, however, that this type of study can only be conducted in a research and development mode and not in conjunction with operational testing where administration conditions must be the same for all students.

The appropriateness of testing accommodations for English-language learners is particularly critical, given the significance of this population in California. The use of bilingual dictionaries, for example, may be common in daily instruction for many students and might be a common accommodation in employment or other settings after high school. Nonetheless, the use of such dictionaries may not be judged to be an appropriate accommodation for the HSEE, at least on the Language Arts examination. Clarification of the knowledge and skills to be assessed may be required to explain why this and many other accommodations are or are not allowed.



## REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Argetsinger, A. (2000, May 25). Maryland delays requiring tests for diploma. *The Washington Post*, p. B1.
- Bempechat, J. (1998). *Against the odds: How at-risk children exceed expectations*. San Francisco: Jossey-Bass Publishers.
- Bond, L.A., & King, D. (1995). *State high school graduation testing: Status and recommendations*. Oak Brook, IL: North Central Regional Educational Laboratory.
- California Department of Education. (1999a). *Guiding principles for the high school exit examination standards panel*. Sacramento, CA: Author. [Online], Available: [www.cde.ca.gov/](http://www.cde.ca.gov/)
- California Department of Education. (1999b). *Selected statewide data for the year 1998-99: Number and percent of English learners in California public schools*. Sacramento, CA: Author. [Online], Available: <http://dq.cde.ca.gov/dataquest/dataquest.asp>
- California Department of Education. (1999c). *Statewide enrollment in California public schools by ethnic group for the year 1999-00*. Sacramento, CA: Author. [Online], Available: <http://dq.cde.ca.gov/dataquest/dataquest.asp>
- California Department of Education (1999d). *Mathematics Framework for California Public Schools*. Sacramento, CA: Author.
- California Department of Education (1999e). *Reading/Language Arts Framework for California Public Schools*. Sacramento, CA: Author.
- The case against TAAS. (1999, Fall). *FairTest Examiner*. [Online], Available: [http://www.fairtest.org/examarts/fall99/The\\_Case\\_Against\\_TAAS.html](http://www.fairtest.org/examarts/fall99/The_Case_Against_TAAS.html). [Note, the preceding Web address is no longer valid.]
- Chudowsky, N. , & Behuniak, P. (Winter 1998). Using focus groups to examine the consequential aspect of validity. *Educational Measurement: Issues and Practice*, 17(4), 28-38.
- Court rules for high-stakes testing. (2000, Winter). *FairTest Examiner*. [On-line], Available: [http://www.fairtest.org/examarts/winter00/Court\\_Rules\\_For\\_High-Stakes\\_Testing.html](http://www.fairtest.org/examarts/winter00/Court_Rules_For_High-Stakes_Testing.html). [Note, the preceding Web address is no longer valid.]
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan*, 73(3): 220-25.
- Debra P. v. Turlington*, 474 F. Supp. 244 (M.D. Fla. 1979); *aff'd in part and rev'd in part*, 644 F2d 397 (5<sup>th</sup> Cir. 1981); *rem'd* 564 F. Supp. 177 (M.D. Fla. 1983); *aff'd*, 730 F2d 1405 (11<sup>th</sup> Cir. 1984).

- Debra P. v. Turlington*, 644 F.2d 397 (5<sup>th</sup> Cir. 1981).
- Decision in *G.I. Forum, et al., v. Texas Education Agency, et al.*, Civil Action No. SA-97-CA-1278EP, (2000). [On-line], Available: <http://www.txwd.uscourts.gov>
- Dorans, N.J., & Holland, P.W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- G.I. Forum and Image de Tejas v. Texas Education Agency*, Civil Action No. SA-97-CA-1278EP, (1997).
- Goals 2000: Educate America Act*. (1994; 1996). Public Law 103-227.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.
- Hauser, R.M. (1997). Indicators of high school completion and dropout. In R.M. Hauser, B.V. Brown, & W.R. Prosser (Eds.), *Indicators of children's well-being*. New York: Russell Sage Foundation.
- Herrick, T., & Walt, K. (1997, Oct. 14). Hispanic group files suit against use of high school exit exams. *Houston Chronicle*. [Online], Available: <http://www.chron.com/content/chronicle/page1/97/10/15/taas.html> Note: the preceding Web address is no longer valid.]
- Hess, F. M., & Brigham, F. (January 2000). None of the above: The promise and peril of high-stakes testing. *American School Board Journal*, 187(1), 26-29.
- Heubert, J.P. (2000). High stakes testing and civil rights: Standards of appropriate test use and a strategy for enforcing them.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F.M., & Novick, M.R. (1968) *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- McLaughlin, M.W., Shepard, L.A., & O'Day, J.A. (1995). *Improving education through standards-based reform report*. The National Academy of Education Panel on Standards-Based Education Reform. New York: National Academy of Education.
- Messick, S. (1989). Validity. *Educational Measurement*, 3<sup>rd</sup> Ed., R.L. Linn, ed. New York: MacMillan.
- The National Education Goals Panel. (1998, Feb.). *Principles and recommendations for early childhood assessments*. [Online], Available: <http://www.negp.gov/Reports/prinrec.pdf> Note: the preceding Web address is no longer valid.]
- National Research Council. (1999). *High stakes: Testing for tracking, promotion, graduation*. J. Heubert & R. Hauser, eds. Board on Testing and Assessment. Washington, DC: National Academy Press.
- National Research Council. (1997). *Educating one and all: Students with disabilities and standards-based reform*. L. McDonnell, J. McLaughlin, & P. Morison, eds. Board on Testing and Assessment. Washington, DC: National Academy Press.

- National Research Council and Institute of Medicine. (1997). *Improving schooling for language-minority children*. D. August & K. Hakuta, eds. Board on Children, Youth and Families. Washington, DC: National Academy Press.
- Natriello, G., & Pallas, A.M. (1999, Nov. revised). *The development and impact of high stakes testing*. Paper presented at the High Stakes K-12 Testing Conference, Dec. 4, 1998; Sponsored by the Civil Rights Project, Harvard University, Teachers College, Columbia University, and Columbia Law School, New York. [On-line], Available: <http://www.law.harvard.edu/groups/civilrights/conferences/testing98/drafts/natriello99.html> [Link is no longer valid.]
- N.C. lawsuit charges test misuse. (1997, Fall). *FairTest Examiner*. [Online], Available: <http://www.fairtest.org/examarts/fall97/k-nc.html>. [Link is no longer valid.]
- NGA Reports Online. (1998, Sept. 1). High school exit exams: Setting high expectations. [Online], Available: <http://www.nga.org/Pubs/IssueBriefs/1998/980901ExitExams.asp>. [Link is no longer valid.]
- Olson, J.F., Bond, L., & Andrews, C. (1999). *Summary Report of State Student Assessment Programs*. Council of Chief State School Officers.
- Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress*. NCES 97-482. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Racial bias built into tests. (2000, Winter). *FairTest Examiner*. [Online], Available: [http://www.fairtest.org/examarts/winter00/Racial\\_Bias\\_Built\\_into\\_Tests.html](http://www.fairtest.org/examarts/winter00/Racial_Bias_Built_into_Tests.html). [Link is no longer valid.]
- Riley, R.W. (2000, Feb. 22). *Setting new expectations*. The Seventh Annual State of American Education Address. Southern High School, Durham, NC. [Online], Available: <http://www.ed.gov/Speeches/02-2000/000222.html> [Link is no longer valid.]
- S.F. files suit over testing of LEP students. (1998, Apr. 8). *Education Week*.
- SB 2X high school exit exam: Highlights. (1999). California Department of Education, Standards, Curriculum, and Assessment Division. [Online], Available: <http://www.cde.ca.gov/>
- Shannon, K. (1997, Oct. 27). Judge hears argument in trial over 10<sup>th</sup> grade test. Texas News in *Abilene Reporter-News*. [Online], Available: <http://www.reporternews.com/1999/texas/taas1027.html>. [Link is no longer valid.]
- Shepard, L.A. (1991). Negative policies for dealing with diversity: When does assessment and diagnosis turn into sorting and segregation? *Literacy for a Diverse Society: Perspectives, Practices and Policies*, E. Hiebert, ed. New York: Teachers College Press.
- Smith, M.L., Heinecke, W., & Noble, A.J. (1999a, Nov. 9). State assessment becomes political spectacle: Part I: Introduction to policy stories and policy studies. *Teachers College Record*. [On-line], Available: <http://www.tcrecord.org/printertext.asp?@IDNumber=10454> [Link is no longer valid.]
- Smith, M.L., Heinecke, W., & Noble, A.J. (1999b, Nov. 24). State assessment becomes political spectacle: Part II: Theoretical and empirical foundations of the research. *Teachers College Record*. [On-line], Available: [www.tcrecord.org/PrinterText.asp?@IdNumber=10457](http://www.tcrecord.org/PrinterText.asp?@IdNumber=10457)
- Smith, M.L., Heinecke, W., & Noble, A.J. (1999c, Dec. 2). State assessment becomes political spectacle: Part III: Defining ASAP: Scripts and readings. *Teachers College Record*. [Online], Available: <http://www.tcrecord.org/PrinterText.asp?@IdNumber=10459> [Link is no longer valid.]

- Smith, M.L., Heinecke, W., & Noble, A.J. (1999d, Dec. 11). State assessment becomes political spectacle: Part IV: First act in the history of Arizona assessment policy. *Teachers College Record*. [On-line], Available: <http://www.tcrecord.org/printertext.asp?@IDNumber=10474> [Link is no longer valid.]
- Smith, S.E. (1999, Oct. 28). Texas test patterns. *Black Issues in Higher Education*, 16(18), 16.
- Solida, M.M. (2000, April 12). Exit-exam reflects trend across country. *The Indianapolis Star*. [Online], Available: [http://www.starnews.com/news/citystate/0412st\\_istep.html](http://www.starnews.com/news/citystate/0412st_istep.html) [Link is no longer valid.]
- Stake, R. (1998, July). Some comments on assessment in U.S. education. *Education Policy Analysis Archives*. [Online serial] 6(14), Available: <http://epaa.asu.edu/epaa/v6n14.html> [Link is no longer valid.]
- Texas test sued in federal court. (1997, Fall). *FairTest Examiner*. [Online], Available: <http://www.fairtest.org/examarts/fall97/k-texas.html> [Link is no longer valid.]
- Thissen, D. (1991). *MULTILOG User's Guide*. Chicago, IL: Scientific Software Inc.
- Seyfarth, A. (1998). *Educational results for students with disabilities: What do the data tell us?* Minneapolis, MN: National Center on Educational Outcomes.
- Ysseldyke, J.E., Thurlow, M.L., Langenfeld, K.L., Nelson, J.R., Teelucksingh, E., & Wise, L. L., Hoffman, R. G., & Harris, C. D. (2000). *The California high school exit examination (HSEE): Evaluation plan*. Alexandria, VA: Human Resources Research Organization.

## **Appendix A**

### **Explanation of Item Response Theory Concepts Used in Analyzing Potential Test Score Accuracy**

## Explanation of Item Response Theory Concepts Used in Analyzing Potential Test Score Accuracy

The analyses of potential test score accuracy reported in Chapter 4 rely heavily on item response theory (IRT) models. Three constructs are important to understanding our analyses: item characteristic curves, test characteristic curves, and conditional standard errors. What follows is an attempt to describe these concepts without getting bogged down in the underlying assumptions and mathematical formulae.

**Item Characteristic Curves.** Item characteristic curves are functions that predict the probability of passing a given item for all students at a given level of ability. In IRT models, this probability depends on the *relative* difference between the item’s difficulty and the examinee’s ability. Initially, neither item difficulty nor examinee ability are known, so ability is placed on an arbitrary scale and item difficulties are “scaled” (estimated) relative to these abilities. Most commonly, the examinee ability is assumed to have a normal (“bell-shaped”) distribution with a mean (average score) of zero and a standard deviation of 1.0. Figure A.1 shows an example of this normal distribution, where the height of the normal curve indicates the relative frequency of students with that level of ability.

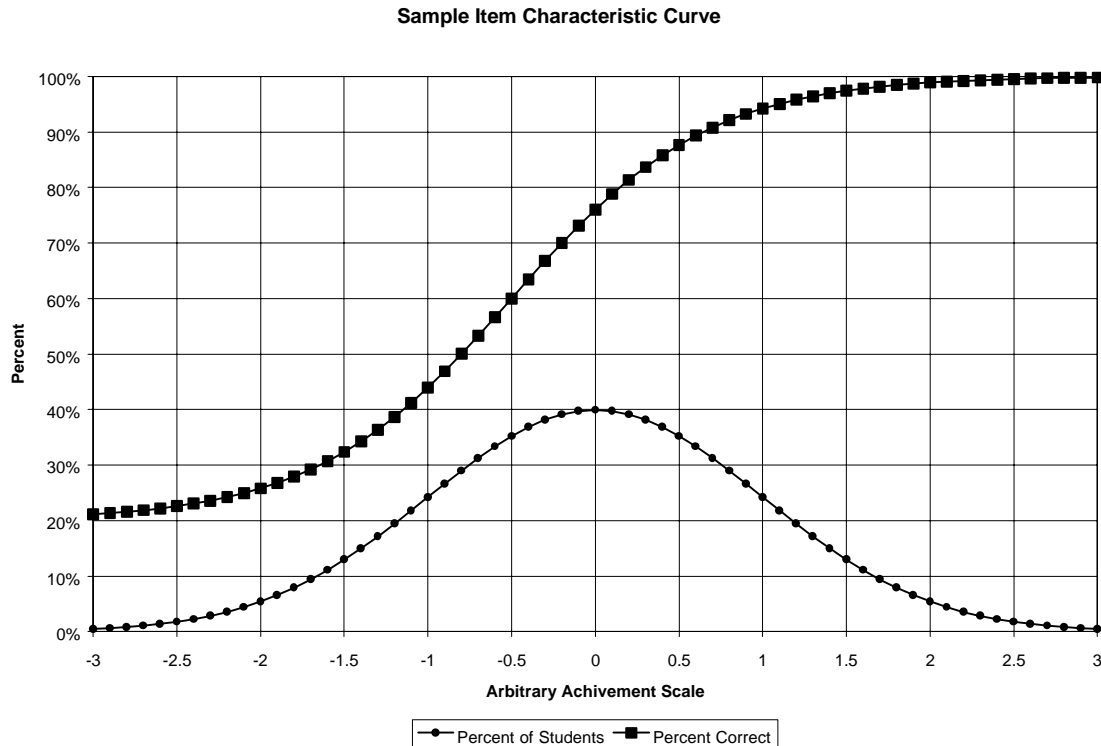


Figure A.1. Illustration of IRT concepts of examinee ability distribution and the probability of passing an item.

Figure A.1 also shows an example of the most common form (the three-parameter-logistic or 3PL model) of an item characteristic curve (ICC). This function has three parameters that capture important differences in the relationship of ability to the probability of correct responses for different items. ICCs are generally S-shaped curves that vary

between a lower asymptote and 1.0. The lower asymptote (or “c” parameter) is the probability of answering correctly for students with no ability at all. For multiple choice items, the lower asymptote is generally close to one divided by the number of possible response options. This is the probability of getting the correct answer through random guessing. The c-parameter will vary also across items as a function of the allure or repugnancy of the different incorrect choices. The item in Figure A.1 has a c-parameter of .20.

The other two parameters that define the shape of different ICCs are the difficulty (“b”) and slope (“a”) parameters. The b-parameter is the point on the ability scale at which the probability of a correct response is half way between the lower asymptote and 1.0. As the b-parameter increases, the ICC moves to the right and the probability of a correct response at each ability level goes down, consistent with the general ideal that passing rates are lower for more difficult items. The item in Figure A.1 has a b-parameter of  $-0.5$ . The slope (“a”) parameter gives the slope of the curve at the ability given by the b-parameter. It is sometimes call a discrimination parameter because it indicates how well the item discriminates between high and low ability students. (This has nothing to do with discrimination among students based on other characteristics.) For items with high a-parameters, the probability of a correct response drops off (increases) sharply for students below (above) the ability level defined by the b-parameter.

**Test Characteristic Curves (TCC).** Test characteristic curves are similar to item characteristic curves except that what is predicted is the total number of items in a whole test that an examinee at a given ability level will answer correctly rather than the probability of answering an individual item correctly. The expected number correct is simply the sum of the probabilities of answering each of the individual items correctly, so TCCs are computed by summing the ICCs. For example, for a three-item test, a given examinee’s probability of answering the first item correctly was .25, the probability for the second item was .50 and the probability for the third item was again .25. This examinee’s expected or average score would be 1.0 ( $.25 + .5 + .25$ ). Note that the estimated scores are averages and do not have to be whole numbers.

**Conditional Standard Errors.** In the preceding example, an examinee of a given ability had an expected score of 1.0 on a three-item test. This does not mean that the examinee will get exactly one item correct every time. Some of the time (about 28%) he or she will miss all of the items and some of the time (about 3%) he or she will answer all of the items correctly. The “standard error of measurement” provides an indication of how often an examinee’s actual score on a particular test form will differ significantly from their expected score (across all parallel forms).

For normative tests, it is common to summarize measurement error by a single number. The *test reliability* of a test is defined for a specific population equivalently as the expected correlation of scores from parallel forms or the ratio of the variance of the true scores to the total (true and error) score variance. From this last definition, the reliability is said to give the proportion of total variance “accounted” for by the underlying ability. (Note that the proportion not accounted for is error.) Coefficient alpha (Cronbach, 1951) is a statistic used to estimate test reliability by summarizing the agreement among a set of items.

For tests such as the HSEE, which are used to make important decisions at particular levels of achievement, overall test reliability is not the main issue. The critical concern is accuracy at the specific point on the score scale where a pass/fail decision must be made. The idea of conditional standard errors, is that error of measurement can be estimated separately for (“conditioned on”) each level of ability.

IRT provides a basis for estimating conditional standard errors. Because of independence assumptions, it is possible to compute the probability of each possible number correct score for a student at a given achievement level (once the item parameters have been estimated). From this information, it is possible to determine how often observed scores will vary from the true score by any given amount. The conditional standard error is defined such that the absolute difference between the observed and true scores will be less than one standard error about two-thirds of the time. If errors are distributed normally, the difference will be less than two standard errors about 95% of the time and this criterion is used most often in constructing confidence bands.



## **Appendix B**

### **Principal and Teacher Surveys**

# California High School Exit Examination Evaluation

## Principal Baseline Survey

**DIRECTIONS:** *Please provide the following information by marking in the circle of the appropriate response or by writing an appropriate response.*

### About You and Your School

<p>1. What is your highest level of education?</p> <p><input type="radio"/> Bachelor's (4-year) degree</p> <p><input type="radio"/> Some graduate school</p> <p><input type="radio"/> Master's Degree</p> <p><input type="radio"/> Doctorate Degree</p> <p><input type="radio"/> Other (please specify) _____</p>	<p>3. What is your primary ethnic background?</p> <p><input type="radio"/> American Indian/Alaskan Native</p> <p><input type="radio"/> Asian or Pacific Islander</p> <p><input type="radio"/> Black or African American, not Hispanic origin</p> <p><input type="radio"/> Caucasian, not Hispanic origin</p> <p><input type="radio"/> Hispanic/Latino</p> <p><input type="radio"/> Other (specify) _____</p>
<p>2. What is your gender?</p> <p><input type="radio"/> Female</p> <p><input type="radio"/> Male</p>	<p>4. When you were a teacher, what was the primary subject area that you taught?</p> <p>_____</p>

5. Including the 1999–2000 school year, how many years...

- a. ...have you been a principal (or school-level administrator)? \_\_\_\_\_
- b. ...were you a teacher? \_\_\_\_\_
- c. ...have you worked in your present school? \_\_\_\_\_
- d. ...have you worked in public schools? \_\_\_\_\_

6. For the 1999–2000 school year:

- a. How many teachers are on the staff? \_\_\_\_\_
- b. What percentage of your teachers has advanced degrees (i.e., beyond BA/BS)? \_\_\_\_\_

7. Have there been any major staff or faculty changes over the past three years? If so, please describe. \_\_\_\_\_

8. What is your school's counselor-student ratio? \_\_\_\_\_

9. Does your school have a testing coordinator? ☐ Yes ☐ No ☐ Will have by \_\_\_\_\_  
date

<p>10. How is your school year configured?</p> <p><input type="radio"/> Semesters</p> <p><input type="radio"/> Trimesters</p> <p><input type="radio"/> Quarters</p> <p><input type="radio"/> Year-Round School</p> <p><input type="radio"/> Other (please specify) _____</p>	<p>12. What grades are taught at your school?</p> <p><input type="radio"/> 9<sup>th</sup>, 10<sup>th</sup>, 11<sup>th</sup>, 12<sup>th</sup></p> <p><input type="radio"/> 10<sup>th</sup>, 11<sup>th</sup>, 12<sup>th</sup></p> <p><input type="radio"/> 7<sup>th</sup>, 8<sup>th</sup>, 9<sup>th</sup></p> <p><input type="radio"/> Other (please specify) _____</p>
<p>11. How many academic class periods are in your school day? _____</p>	<p>13. How long is each academic class period (in minutes)? _____</p>

14. If you use any block scheduling, please describe. \_\_\_\_\_

15. Indicate the various specialty education programs offered by your school. (Mark all that apply; estimate % of students who participate in each; and comment.)

☐ Remedial Courses (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Magnet Program (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Special Education (specify category(ies)) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ English-Language Learners (specify language(es)) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Multicultural/Diversity-Based (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Advanced Placement (specify subject(s)) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ International Baccalaureate (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ School/Community/Business Partnerships (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Targeted Tutoring (briefly describe) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

☐ Other (specify) \_\_\_\_\_ % \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

16. Consider your students, overall, and within each of the following racial/ethnic group. What is your current graduation rate? What is the mobility rate in a typical school year?

	Seniors Overall	American Indian/ Alaskan Native	Asian or Pacific Islander	Black or African American, not Hispanic origin	Caucasian not Hispanic origin	Hispanic /Latino	Other (specify) _____
Current graduation rate (% entering 9 <sup>th</sup> graders who graduate within 4–5 years)							
Typical mobility rate (% students who transfer in and/or out of your school within a school year)							

17. Based on your own most recent school data (e.g., CBEDS), what percentage of your seniors overall, indicated each main activity as their choice for year after they graduate from high school? Then, consider seniors within each racial/ethnic group. Again, what percentage of each group indicated each main activity as their choice for the year after they graduate from high school?

Indicate Source _____ 1997–98 school data _____ 1998–99 school data _____ other _____	Seniors Overall	American Indian/ Alaskan Native	Asian or Pacific Islander	Black or African American, not Hispanic origin	Caucasian not Hispanic origin	Hispanic /Latino	Other (specify) _____
Working full time							
Attending a vocational, technical, or business school							
Attending a 2-year college							
Attending a 4-year college, service academy, university							
Serving in the regular military service							
Other _____							
TOTAL	100%	100%	100%	100%	100%	100%	100%

18. Have there been any changes in the student demographics or academic goals over the past three years (e.g., push for new programs—advanced or remedial, graduation or dropout rate, interest in college, school boundaries)? If so, please describe. \_\_\_\_\_

19. How would you describe the academic atmosphere of your school (e.g., rigor of the curriculum, staff's satisfaction with the curriculum, student motivation and effort, parental involvement, etc.)? \_\_\_\_\_

20. How would you describe the education level of the majority of your students' parents? Estimate the overall average percent of parents in each of the following categories:

Less than high school graduation _____	College graduate (4- or 5-year degree) _____
High school diploma or GED _____	Some graduate school or graduate degree _____
Vocational, technical, or business training _____	Other (specify) _____
Associate, 2-year degree _____	Total = 100%

### About the California New High School Exit Examination

21. How familiar are you with the:

California High School Exit Examination?

- a. ☐ I am not at all familiar with the exit exam.  
☐ I have only general information about the exam.  
☐ I am very familiar with the new exam.

State Content Standards?

- b. ☐ I am not at all familiar with the state content standards.  
☐ I have only general information about the content standards.  
☐ I am very familiar with the content standards.

22. What have been your sources of information about the High School Exit Examination? (Mark all that apply.)

- |   |   |
|---|---|
| <input type="radio"/> None                          | <input type="radio"/> Education organization (e.g., publication, meeting, etc.)   |
| <input type="radio"/> State-provided information    | <input type="radio"/> Professional association (e.g., publication, meeting, etc.) |
| <input type="radio"/> District-provided information | <input type="radio"/> Computer-based source (e.g., listserv, newsgroup, etc.)     |
| <input type="radio"/> Newspaper                     | <input type="radio"/> Other (specify) _____                                       |

23. How familiar do you think your students and their parents are with the exit exam?

①	②	③	④	⑤
Not At All Familiar	Not Very Familiar	Somewhat Familiar	Familiar	Very Familiar

Comments \_\_\_\_\_  
 \_\_\_\_\_

24. Consider the Content Standards described by the *Content Standards for California Public School* and the *Framework for California Public Schools* and mark all the following statements that apply.

- ☐ My school and/or district encourages use of the content standards to organize instruction.
- ☐ Our textbooks do not align well with the content standards.
- ☐ We can cover all of the content standards with a mix of textbooks and supplemental material.
- ☐ Our school and/or district is in the process of aligning its curriculum to the standards.
- ☐ Our school and/or district has a plan which ensures that all students receive instruction in each of the content standards.

25. The relationship between our district standards and the California state content standards can best be described as: (Mark only one.)

- ☐ Our district has adopted the state content standards.
- ☐ The state content standards include more than our district content standards.
- ☐ Our district content standards include more than the state content standards.
- ☐ The two sets of content standards are different.
- ☐ I can not judge the relationship between our district standards and the state standards.
- ☐ Our district does not have an official set of content standards.

26. What are you planning to do to prepare your faculty/staff, parents, and community for the first administration of the exit exam? \_\_\_\_\_

\_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

27. During this school year (1999–2000), how much time, in total, do you estimate you have spent in activities related to the pending implementation of the California High School Exit Examination (e.g., meetings, discussions, curriculum review, your professional development, your staff's development, etc.)?

- |   |  |
|---|--|
| <input type="radio"/> None              | <input type="radio"/> 16–35 hours        |
| <input type="radio"/> Less than 6 hours | <input type="radio"/> More than 35 hours |
| <input type="radio"/> 6–15 hours        |  |

28. What activities have you and your faculty/staff undertaken to prepare your students for the first administration of the California High School Exit Examination? (Mark all that apply.)

- |   |  |
|---|--|
| <input type="radio"/> None                              | <input type="radio"/> Use school test results to change instruction          |
| <input type="radio"/> Provide individual/group tutoring | <input type="radio"/> Use school test results to design remedial instruction |
| <input type="radio"/> Add homework                      | <input type="radio"/> Increase summer school offerings                       |
| <input type="radio"/> Administer "early warning" tests  | <input type="radio"/> Eliminate electives in favor of remedial classes       |
| <input type="radio"/> Adopt state content standards     | <input type="radio"/> Develop parent support program                         |
| <input type="radio"/> Alter curriculum                  | <input type="radio"/> Other (specify) _____                                  |

29. After the first administration of the exit examination, there likely will be students who did not pass. Describe any procedures, personnel assignments or accountability, etc. your school is planning, to assume responsibility for getting students through successful completion of the exam.

---



---



---



---

30. What plans have you and your faculty/staff made to prepare for remediation of students who do not pass the exit exam or who do not seem prepared to take it? (Mark all that apply.)

- |   |  |
|---|--|
| <input type="radio"/> None                              | <input type="radio"/> Use school test results to change instruction          |
| <input type="radio"/> Provide individual/group tutoring | <input type="radio"/> Use school test results to design remedial instruction |
| <input type="radio"/> Add homework                      | <input type="radio"/> Increase summer school offerings                       |
| <input type="radio"/> Administer "early warning" tests  | <input type="radio"/> Eliminate electives in favor of remedial classes       |
| <input type="radio"/> Adopt state content standards     | <input type="radio"/> Develop parent support program                         |
| <input type="radio"/> Alter curriculum                  | <input type="radio"/> Other (specify) _____                                  |

31. Based on what you know about your school and the requirements of algebra 1 content and English-Language Arts 9<sup>th</sup>-10<sup>th</sup> content, what percentage of your current 10<sup>th</sup> grade students do you think deserve a passing score on the High School Exit Examination?

- |                                      |  |
|--------------------------------------|--|
| <input type="radio"/> Fewer than 50% | <input type="radio"/> 75–95%           |
| <input type="radio"/> 50–74%         | <input type="radio"/> Greater than 95% |

32. Based on what you know about your school, what do you predict the impact of the High School Exit Examination, will be on...?

	Strongly Negative	Negative	No Effect	Positive	Strongly Positive
a. ...student motivation prior to taking the exam for the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...motivation to excel for students who pass the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ... motivation to excel for students who fail the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...parental involvement prior to the first required administration of the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...parental involvement for students who pass the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. ...parental involvement for students who fail the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. ...student retention rates?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. ...student dropout rates?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

33. Based on what you know about your school, what do you predict the influence of the High School Exit Examination will be on instructional practices?

①	②	③	④	⑤
Strongly Improved	Improved	No Effect	Weakened	Extremely Weakened

Comments \_\_\_\_\_

34. Please rate each of the following topics related to the initial administration of the California High School Exit Examination.

*Opportunity to learn the material covered by the exam (content standards) for...*

	None	Poor	Adequate	Good	Excellent	Not Sure
a. ...all your school's students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...students with disabilities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ...English-language learners—acquiring necessary English proficiency.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...English-language learners—in target subject area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...economically disadvantaged students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Comments \_\_\_\_\_

*Opportunity to demonstrate their knowledge and skills on the exam for...*

	None	Poor	Adequate	Good	Excellent	Not Sure
a. ...all your school's students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...students with disabilities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ...English-language learners—acquiring necessary English proficiency.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...English-language learners—in target subject area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...economically disadvantaged students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Comments \_\_\_\_\_

35. What plans or strategies do you and your faculty/staff have to prepare for IEP changes that will address participation of a student with a disability in the exit exam?

---



---



---



---



---



---

36. Please describe any specific challenges you feel your school and students face in successfully meeting the requirements of the California High School Exit Examination.

---

---

---

---

---

---

37. Please describe any specific benefits for your school and students that you feel are associated with the requirements of the California High School Exit Examination.

---

---

---

---

---

---

38. Please write any comments about other factors specific to your school that are influencing the exit examination (e.g., community conditions, economic changes, parental views, etc.)

[illegible]

*Thank you for your cooperation.*



# California High School Exit Examination Evaluation

## Teacher Baseline Survey

**DIRECTIONS:** Please provide the following information by marking in the circle of the appropriate response or by writing an appropriate response.

### About You

1. What is your highest level of education? <input type="radio"/> Bachelor's (4-year) degree <input type="radio"/> Some graduate school <input type="radio"/> Master's Degree <input type="radio"/> Doctorate Degree <input type="radio"/> Other (specify) _____	4. What is your primary ethnic background? <input type="radio"/> American Indian/Alaskan Native <input type="radio"/> Asian or Pacific Islander <input type="radio"/> Black or African American, not Hispanic origin <input type="radio"/> Caucasian, not Hispanic origin <input type="radio"/> Hispanic/Latino <input type="radio"/> Other (specify) _____
2. What is your primary subject area? <input type="radio"/> English/Language Arts (E/LA) <input type="radio"/> Mathematics (Math) <input type="radio"/> Other (specify) _____	5. What is your gender? <input type="radio"/> Female <input type="radio"/> Male
3. Was your college training in your primary subject area? <input type="radio"/> Yes <input type="radio"/> No (specify other area) _____	6. Including the 1999–2000 school year, how many years have you... a. ...been a teacher? _____ b. ...been a teacher in your primary subject area? _____ c. ...taught in your present school? _____

### About You and Your Classes

7. What is your average enrollment per class period this year? _____	10. Do you require students to maintain a portfolio? <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Other (describe) _____
8. Do you create groups within your classes for instruction? <input type="radio"/> Yes <input type="radio"/> No If YES, on what basis are the groups formed? <input type="radio"/> Random assignment <input type="radio"/> Math or English-Language Arts ability <input type="radio"/> Student choice <input type="radio"/> Other (describe) _____	11. Think about the level of preparation students in your classes have in your subject area—math or English-Language Arts (E-LA). If you are a <u>mathematics teacher</u> , estimate the overall average percentage of students in each of the following categories: Excellent math preparation _____ Good math preparation _____ Fair math preparation _____ Poor math preparation _____ Total = 100%  If you are an <u>English-Language Arts teacher</u> , estimate the overall average percentage of students in each of the following categories: Excellent E-LA preparation _____ Good E-LA preparation _____ Fair E-LA preparation _____ Poor E-LA preparation _____ Total = 100%
9. What is the average percentage of the students in your classes who speak English fluently? <input type="radio"/> 100% <input type="radio"/> 50%–74% <input type="radio"/> 90%–99% <input type="radio"/> Less than 50% <input type="radio"/> 75%–89%	

12. In your teaching practices, how important do you consider each of the following to be?

(Mark the appropriate circle for each.)	Very Important	Some-what Important	Not Very Important	Not Important
a. Using direct instruction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Using problem solving both as a goal of instruction and as a means of investigating important concepts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Using questioning techniques that promote student interaction and discussion	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Covering as many content topics as possible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Developing the students' ability to make connections among content topics	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Developing the students' ability to make connections across subject areas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Using drill and practice to reinforce knowledge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. Using the results of classroom assessments to inform instructional decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. Using collaborative learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Using peer teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<p>13. On average, how much class time do students in your classes spend each week working with a partner or in a small group?</p> <p><input type="radio"/> None</p> <p><input type="radio"/> Less than ½ hour</p> <p><input type="radio"/> ½–1 hour</p> <p><input type="radio"/> More than 1 hour</p>	<p>14. On average, how much time do you believe students in your classes spend each week on your assignments <i>outside</i> of the classroom?</p> <p><input type="radio"/> None</p> <p><input type="radio"/> Less than 1 hour</p> <p><input type="radio"/> 1–3 hours</p> <p><input type="radio"/> More than 3 hours</p>
--	---

15. In general, how often do you plan for students in your classes to: ...?

(Please mark the appropriate circle for each of the following.)

	Almost Every Day	Once or Twice a Week	Once or Twice a Month	Once a Grading Period	Never or Hardly Ever
a. Do work from their textbooks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. Do work from supplemental materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. Work with hands-on materials, physical models or manipulatives	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Work in pairs or small groups	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Take quizzes or tests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Be asked to apply subject area knowledge to real-world situations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. Write a few sentences about a topic or its consequences (or math problem or its solution)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. Write reports or complete projects	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. Conduct research on issues or ideas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. Present their work to the class	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### About You and Your School

16. Have there been any major staff or faculty changes at your school over the past three years? If so, please describe.

---

17. During the current school year (1999–2000), how much time, in total, did you spend in professional development workshops, inservice, or seminars in your primary subject area or related to interpreting data from the state? Include attendance at district-sponsored training and external training.

- ☐ None  
☐ Less than 6 hours  
☐ 6–15 hours  
☐ 16–35 hours  
☐ More than 35 hours

18. Indicate your involvement in specialty education programs at your school. (Mark all that apply and comment.)

- ☐ Remedial Courses (briefly describe) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ Special Education (specify category(ies)) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ English-Language Learners (specify language(es)) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ Multicultural/Diversity-Based (briefly describe) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ Advanced Placement (specify subject(s)) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ International Baccalaureate (briefly describe) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ School/Community/Business Partnerships (briefly describe) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ Targeted Tutoring (briefly describe) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_  
☐ Other (specify) \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

### About the New High School Exit Examination

19. How familiar are you with the:

California High School Exit Examination?

- a. ☐ I am not at all familiar with the exit exam.  
☐ I have only general information about the exam.  
☐ I am very familiar with the new exam.

State Content Standards?

- b. ☐ I am not at all familiar with the state content standards.  
☐ I have only general information about the content standards.  
☐ I am very familiar with the content standards.

20. Did you participate in one of the High School Exit Exam Educator Panel Item Rating Workshops?

- ☐ Yes–Saturday, May 6<sup>th</sup> in Sacramento  
☐ Yes–Saturday, May 13<sup>th</sup> in Orange County  
☐ No

21. What have been your sources of information about the California High School Exit Examination?

(Mark all that apply.)

- ☐ None  
☐ School-provided information  
☐ State-provided information  
☐ District-provided information  
☐ Newspaper
- ☐ Education organization (e.g., publication, meeting, etc.)  
☐ Professional association (e.g., publication, meeting, etc.)  
☐ Computer-based source (e.g., listserv, newsgroup, etc.)  
☐ Other (specify) \_\_\_\_\_

22. How familiar do you think other teachers at your school are with the:

## California High School Exit Examination?

- a. ☐ Most are not at all familiar with the exit exam.
- ☐ Most have only general information about the exam.
- ☐ Most are very familiar with the new exam.

## State Content Standards?

- b. ☐ Most are not at all familiar with the state content standards.
- ☐ Most have only general information about the content standards.
- ☐ Most are very familiar with the content standards.

23. Based on what you know about your feeder schools, how well prepared do you feel the students will be to pass the High School Exit Examination...

	Very Well Prepared	Well Prepared	Prepared	Not Well Prepared	Not At All Prepared
a. ...when they are in 9 <sup>th</sup> grade?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...when they are in 10 <sup>th</sup> grade?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

24. During this school year (1999–2000), how much time, in total, do you estimate you have spent in activities related to the pending implementation of the High School Exit Examination (e.g., faculty and department meetings, discussions, staff development, etc.)?

- ☐ None
- ☐ Less than 6 hours
- ☐ 6–15 hours
- ☐ 16–35 hours
- ☐ More than 35 hours

25. During this school year (1999–2000), how much time, in total, do you estimate you have spent on classroom instruction activities related to the pending implementation of the High School Exit Examination (e.g., department planning, student preparation, curriculum review, etc.)?

- ☐ None
- ☐ Less than 6 hours
- ☐ 6–15 hours
- ☐ 16–35 hours
- ☐ More than 35 hours

26. What activities have you, personally, undertaken to start preparing for your students' involvement in the High School Exit Examination? (Mark all that apply.)

- ☐ None
- ☐ Provide individual/group tutoring
- ☐ Add homework
- ☐ Administer “early warning” tests
- ☐ Alter my curriculum
- ☐ Talk with my students
- ☐ Use class test results to change instruction
- ☐ Use class test results to design remedial instruction
- ☐ Encourage summer school attendance
- ☐ Suggest remedial classes rather than electives
- ☐ Talk or work with parents
- ☐ Other (specify)

27. Please describe any specific changes you made prior to May 1, 2000 to the subject area curriculum you are teaching or to your classroom instructional practices based on influences you anticipate from the exit exam.

---

---

---

28. Please describe any specific changes you plan to make in the future to the subject area curriculum you teach (or intend to teach) or to your classroom instructional practices based on influences you anticipate from the exit exam.

---



---



---



---

29. What activities has your school undertaken to start preparing for the first administration of the High School Exit Examination? (Mark all that apply.)

- |   |  |
|---|--|
| <input type="radio"/> None                              | <input type="radio"/> Use school test results to change instruction          |
| <input type="radio"/> Provide individual/group tutoring | <input type="radio"/> Use school test results to design remedial instruction |
| <input type="radio"/> Add homework                      | <input type="radio"/> Increase summer school offerings                       |
| <input type="radio"/> Administer "early warning" tests  | <input type="radio"/> Eliminate electives in favor of remedial classes       |
| <input type="radio"/> Adopt state content standards     | <input type="radio"/> Develop parent support program                         |
| <input type="radio"/> Alter curriculum                  | <input type="radio"/> Other (specify) _____                                  |

30. After the first administration of the exit examination, there likely will be students who did not pass. Please describe the responsibilities you are likely to be assigned for getting students through successful completion of the exam.

---



---



---



---

31. What plans has your school made to prepare for remediation of students who do not pass the High School Exit Examination or who do not seem prepared to take it? (Mark all that apply.)

- |   |  |
|---|--|
| <input type="radio"/> None                              | <input type="radio"/> Use school test results to change instruction          |
| <input type="radio"/> Provide individual/group tutoring | <input type="radio"/> Use school test results to design remedial instruction |
| <input type="radio"/> Add homework                      | <input type="radio"/> Increase summer school offerings                       |
| <input type="radio"/> Administer "early warning" tests  | <input type="radio"/> Eliminate electives in favor of remedial classes       |
| <input type="radio"/> Adopt state content standards     | <input type="radio"/> Develop parent support program                         |
| <input type="radio"/> Alter curriculum                  | <input type="radio"/> Other (specify) _____                                  |

32. Based on what you know about your school, what do you predict the impact of the High School Exit Examination, will be on...

	Strongly Negative	Negative	No Effect	Positive	Strongly Positive
a. ...student motivation prior to taking the exam for the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...motivation to excel for students who pass the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ... motivation to excel for students who fail the first time?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...parental involvement prior to the first required administration of the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...parental involvement for students who pass the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. ...parental involvement for students who fail the exam?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. ...student retention rates?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. ...student dropout rates?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

33. Based on what you know about your school, what do you predict the influence of the High School Exit Examination will be on instructional practices...

	Strongly Improved	Improved	No Effect	Weakened	Extremely Weakened
a. ...next year (2000–2001)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...in 3 years (2002–2003)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ...in 5 years (2004–2005)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

34. Please describe any specific challenges you feel your school and students face in successfully meeting the requirements of the California High School Exit Examination. \_\_\_\_\_

---



---

35. Please describe any specific benefits for your school and students that you feel are associated with the requirements of the California High School Exit Examination. \_\_\_\_\_

---



---

36. Please rate each of the following topics related to the initial administration of the California High School Exit Examination.

*Opportunity to learn the material covered by the exam (content standards) for...*

	None	Poor	Adequate	Good	Excellent	Not Sure
a. ...all your school's students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...students with disabilities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ...English-language learners—acquiring necessary English proficiency.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...English-language learners—in target subject area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...economically disadvantaged students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*Opportunity to demonstrate their knowledge and skills on the exam for...*

	None	Poor	Adequate	Good	Excellent	Not Sure
a. ...all your school's students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ...students with disabilities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ...English-language learners—acquiring necessary English proficiency.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. ...English-language learners—in target subject area.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. ...economically disadvantaged students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

37. Please write any comments about other factors specific to you, your classes, or your school that are influencing the exit examination (e.g., community conditions, economic changes, parental views, etc.)

---



---



---



---

***Thank you for your cooperation.***