HumRRO

Supplemental Evaluation Report

IR–00–37

August 25, 2000

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

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

Prepared for:
California Department of Education
Sacramento, CA

Contract Number: 9234

Human Resources Research Organization
66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314-1591

High School Exit Examination (HSEE):
Supplemental 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 mathematics and English /language arts 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 9th graders on a voluntary basis in March and May of 2001 with mandatory testing of all 10th graders (except those passing the exam as 9th 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 tryouts of the 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.

The first evaluation report was issued on June 30, 2000. That report covered a review of the implementation of exit examinations in other states, analyses of data from the Spring 2000 field test (tryout of questions), workshops conducted to determine how well draft questions were aligned to the targeted test content standards and to current classroom instruction, and surveys of teachers and principals to establish baseline data for determining the impact of the new requirement. The report concluded that a remarkable amount of progress had been made in developing the HSEE and that results to date were quite positive in that a large number of high quality test questions had been developed. The report cautioned, however, that a great deal remained to be done prior to implementation of the HSEE. In addition, the low proportion of correct answers in the field test, relatively low ratings of the alignment of the current curriculum to the test questions, and low estimated passing rates in the principal survey all suggested that students were currently not well-prepared for the HSEE. The overarching recommendation was that consideration should be given to delaying implementation of the new graduation requirement to allow more time to prepare the test for students and, more importantly, more time to prepare students for the test.

The present report describes supplemental analyses of the field test data, including information on the essay questions (Chapter 2). Because of the time required to score the responses to these questions, data were not available on these questions in time for inclusion in our June 30 report. This supplemental report also provides more detail on

curriculum-alignment ratings and passing rates for individual content standards (Chapter 3) and updated results from the teacher and principal surveys (Chapter 4). The report concludes with updates to our initial conclusions and new recommendations for improving the HSEE.

Key Findings and Recommendations

Our main conclusions are unchanged from those stated in the June 30 report. Further progress has been made in developing plans for remaining implementation activities. Analyses of the supplemental field test data indicate that a significant number of high-quality essay questions have been developed. Scores on the responses to these questions were relatively low, however, reinforcing concerns that students are not yet well prepared to pass this examination. Analyses of passing rates and ratings for individual content standards suggest that algebra standards and English/language arts standards involving integration and analysis are most problematic for students today, although performance and curriculum-alignment ratings were generally low for all of the standards.

Specific recommendations, discussed in detail in Chapter 5 of this report, are described briefly here. We continue to believe that a delay in implementing the HSEE should be considered to avoid negative consequences for students who are not well prepared and to reduce the possibility of flaws that might lead to termination of the program before it can achieve its goals. Consideration should be given to first holding schools accountable for teaching to state content standards before implementing consequences for students. Our first recommendation is:

Recommendation 1. The Legislature and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.

The trade-off between risks with the current implementation schedule and risks to students associated with delaying implementation is a policy-judgment that has already been made by the legislature in enacting provisions for the HSEE. The Superintendent of Public Instruction and the State Board of Education have no choice but to proceed with implementation of the HSEE on the current schedule. CDE has made good progress in developing plans for remaining implementation activities. We think it would now be useful to lay out a more detailed timeline for public review so that issues can be identified and addressed as early as possible. We also believe that rapid implementation must be supported by significant funding for state and district activities. Our specific recommendations in support of meeting the current development schedule are:

Recommendation 2. CDE should develop and seek comment on a more detailed timeline for HSEE implementation activities. This timeline should show responsibility for each required task and responsibility for oversight of the performance of each task. The plan should show key points at which decisions by the Board or others are required along with separate paths for alternative decisions that may be made at each of these points.

Recommendation 3. CDE and the Board should work with districts to identify resource requirements associated with HSEE implementation. The Legislature must

be ready to continue to fund activities to support the preparation of students to meet the ambitious challenges embodied in the HSEE.

The State Board is about to adopt specifications for the content to be covered by the examination (content standards) and will subsequently be asked to approve minimum passing scores (performance standards) for mathematics and for English/language arts. We are concerned that there could be some confusion between expectations for average student performance, which is commonly the focus of school-level accountability discussions, and expectations for the minimum performance that should be required for graduation. Specific recommendations with respect to the content and performance standards are:

Recommendation 4. The Board should adopt a clear statement of its intentions in setting HSEE content and performance standards. This statement should describe the extent to which these standards are targeted to ensure minimum achievement relative to current levels or to significantly advance overall expectations for student achievement.

Recommendation 5. The Board should exhibit moderation in selecting content standards and setting performance standards for the initial implementation of HSEE. Standards should be subsequently expanded or increased based on evidence of improved instruction.

Recommendation 6. Members of the HSEE Panel and its Technical Advisory Committee should participate in developing recommendations for minimum performance standards.

We also discuss a number of technical issues in Chapter 5, based on our review of plans developed by CDE and the development contractor. Given tight time constraints, we suggest that there needs to be a process for timely independent review of technical issues to ensure the feasibility and defensibility of the approaches taken. CDE has already made plans to engage a technical consultant for this purpose. Our general recommendation, which is intended to suggest that two or three such consultants might be useful, is:

Recommendation 7. CDE should move swiftly to establish an independent Technical Issues Committee (TIC) to recommend approval or changes to the HSEE development contractor's plans for item screening, form assembly, form equating, and scoring and reporting.

We also offer specific suggestions to strengthen the equating of scores from different test forms and suggest the feasibility and desirability of providing feedback to schools participating in the field tests.

Our report concludes with an update on specific recommendations offered in the June 30 report. These include the need to clarify the relationship between different high school testing programs, the need to provide information and support to districts as they prepare for the HSEE, and the need for more planning and research on test accommodations for special needs students and English language learners.

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

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

The California High School Exit Examination

California has embarked on a bold 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, the American Institutes for Research (AIR), to develop and try out test questions for use in the HSEE. The current schedule calls for testing 9th graders on a voluntary basis in March and May of 2001 with mandatory testing of all 10th graders (except those passing the exam as 9th graders) in 2002. Several additional testing opportunities will follow 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. A report describing evaluation activities, findings, and recommendations (Wise, et al., 2000) was submitted on June 30, 2000 in compliance with the legislative mandate for a preliminary evaluation report by that time. This supplemental report covers analyses of data that were not available at the time the first report was submitted and also addresses questions and suggestions received in response to the June 30 report.

Key Findings and Recommendations from the June 30 Evaluation Report

The main conclusions of the Year 1 Evaluation Report, hereafter referred to as the June 30 report, were that a great deal of progress has been made in developing the HSEE and that results to date were quite positive, as indicated by several measures of the quality of the test questions. Nonetheless, a great deal remained to be done before the HSEE can be administered operationally. Further, educators surveyed were concerned that students are currently not well prepared to pass the exam.

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

> 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 were included in the June 30 report. First, from discussions with panels of educators, surveys of principals and teachers, and discussions with

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

> 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 our principal and teacher surveys indicated clearly that much needed 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 was:

> 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 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 is needed to reach informed decisions. Our final specific recommendation was:

> 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.

Comments and Clarifications based on Responses to the June 30 Report

Since the June 30 report was issued, members of the evaluation team have met the Superintendent, the State Board, and the HSEE Panel. Extensive feedback during a telephone conference with the co-chairs of the HSEE Panel was particularly helpful. A brief summary of key issues and suggestions is provided here. Insofar as feasible, we have sought to provide an expanded discussion of these issues within the text of this supplemental report.

One particularly surprising reaction was the fear that schools familiar with the June 30 report recommendations would, based on belief that HSEE implementation was going to be delayed, relax their efforts to ensure that all students meet the California content standards. The message that we intended was, in fact, just the opposite. The most critical concern was that schools required time to develop and implement programs to help students pass the HSEE requirements. Having said that our most critical concern was getting students ready to take the test, we still talked more extensively about activities needed to get the test ready for the students. We provide some more discussion about what schools need to be doing in Chapter 4 of this report.

The HSEE Panel chairs found our background discussion of experiences in other states helpful. They would welcome suggestions for a plan of action to address the issues that other programs have confronted. The most critical of those issues arose in conjunction with legal challenges to exit exams, most recently in Texas and previously in Florida. Three main concerns are: (1) providing adequate notice to all concerned, (2) demonstrating educational

necessity, and (3) demonstrating that all students have a reasonable opportunity to learn the material on the test. In California, this last issue will be more of a shared activity between state and district educators than in Texas where the state had clear responsibility. We are certainly not a best source for legal guidance, and one of our first suggestions is to continue efforts that CDE has already begun to provide a broad array of legal advice to prepare for and preferably avoid program-stopping litigation. We provide suggestions for key activities that each stakeholder group needs to undertake prior to HSEE implementation in Chapter 5 below.

The Board recognized the critical importance of ensuring that all students have an opportunity to learn the material covered by the HSEE. They requested that more extensive information on district plans and efforts be gathered and reported. Our original intent was to collect baseline information on the current status of instruction relative to the content standards. We did not emphasize future plans leading some to consider our report "retrospective." We will conduct a survey of all districts serving high school students early this fall covering future plans as well as current instruction. Preliminary results will be reported to the Board in November with a final report to be issued in December of this year. Members of the HSEE Panel also requested more detailed information as to which particular content standards appear to be most problematic with respect to ensuring all students the opportunity to learn them. We provide additional information on this topic in Chapter 3 of this report.

Organization and Contents of this Supplemental Year 1 Report

This supplemental report is organized into five chapters, including this introductory chapter. Chapter 2 describes additional analyses of data from the Spring 2000 field test of HSEE test questions with primary focus on results for the essay questions that constitute a section of the English Language Arts (ELA) test. Reponses to these questions had to be hand-scored and score information was not available in time for inclusion in the June 30 report.

Chapter 3 provides more detail on indicators of student preparation with respect to specific content standards covered by HSEE. We also provide an analysis of the representativeness of the sample of districts that participated in our item review workshops.

Chapter 4 presents final results from our surveys of teachers and principals. The June 30 report contained preliminary results from these surveys, but many additional responses were received after the cut-off for that report. Whereas the other chapters of this report are intended as supplements to the June 30 report, Chapter 4 is designed as a complete replacement for the corresponding chapter (Chapter 5) in the June 30 report.

The final chapter includes a discussion of conclusions and recommendations. Since the June 30 report, we have had an opportunity to review drafts of contractor plans and recommendations for a Fall Field Test, for determining the minimum passing score, and for the operational test administrations in March and May 2001. In Chapter 5, we raise some questions that must be addressed in carrying out these plans and offer some suggestions for their refinement.

CHAPTER 2: SUPPLEMENTAL FIELD TEST DATA ANALYSIS

Introduction

The June 30 Year 1 Evaluation Report included extensive analyses of the HSEE multiple choice (MC) questions for mathematics and English/language arts (ELA) in the Spring 2000 field test. At that time, hand-scoring of answers to the ELA essay questions[1] had not been completed. Scores for the essay questions were received on June 28. In this section, we describe the results of our analyses of these questions. First, we examined statistical indicators of the functioning and quality of the scores for the essay questions. Next, we analyzed the consistency of the scores of the responses provided by alternative raters. Then we turned to consideration of the effect of adding the essay questions on the accuracy of overall student scores. We conclude this section with additional analyses of the degree to which the field test samples are representative of California 10th grade students in general.

Quality of the Essay Questions

Booklet Design. After completing 100 multiple-choice questions, each student in the English Language Arts field test was presented two prompts, each requiring an essay response. Responses to the first prompt were scored twice, once as a reading measure based on the content of the response and a second time as a writing measure based on mechanics and style. Three versions of each of the four field test booklets were created, with a different pair of essay questions included in each version of each booklet. The Spring 2000 field test thus included 12 pairs of essay questions.

Scoring. Each response was read, independently, by two different scorers. Following the scoring guide developed for each question, the scorers assigned a score of 1 to 4, with 1 indicating no mastery and a 4 indicating complete mastery. Special codes were used to indicate responses that were: off topic (10), blank (11), or simply illegible (13). (Code 12, indicating a foreign language response, was not given to any of the responses.)

Table 1 shows the distribution of scores across students and across the six scores generated for each student. Two points are noteworthy. First, more than 10 percent of the papers were blank. We could not tell from available information whether a blank paper meant that the student did not have enough time to answer, was not motivated to answer, or simply did not know where to begin. We deleted blank, off topic, and illegible papers from our analyses, but in operational use such papers would be assigned a score of either zero or one. In any event, response rates for the essay questions were significantly different from those found earlier for the MC questions where nearly all students answered all or nearly all of the questions.

[1] Each ELA test booklet included two questions that required students to write an extended answer, usually several paragraphs. Question of this type are sometimes called open-ended, open-response, constructed-response, or extended constructed response questions. We refer to them as essay questions even though the responses are not always essays in the strictest sense.

The second noteworthy point about Table 1 is that relatively few students received full-credit for their responses. In fact, the number of responses scored as 4 was less than the number of blank responses. If blank papers are assigned a score of 0, then the average essay question score was just under 2.0. If blank papers are considered an indicator of low motivation and excluded, the average score for the remaining papers was 2.27.

Table 2.1
Distribution of Scores for the Essay Questions

Score	Number	Percent
1	4,668	20.7
2	6,780	30.0
3	6,298	27.9
4	1,806	8.0
Off Topic	668	3.0
Blank	2,377	10.5
Illegible	5	0.0

Difficulty. While minimum passing scores have not yet been established, it seems likely that students would have to score at least 3 to be considered as passing the standard measured by an essay question. If the purpose of the ELA portion of HSEE is to identify a relatively small proportion of students whose language arts or mathematics skills are below some minimally acceptable level of competency, then we would want relatively easy test questions. For essay questions, this would mean prompts and scoring guides where most students score in the 3 and 4 range, and only the very low-performing students would receive scores of 1 or 2. What we have is just the opposite with most students scoring 1 or 2 and only 8 percent receiving full credit. If, on the other hand, the intent is to require all students to perform above the current average (2.27 on the average essay question), then requiring a score of at least 3 for these essay questions just about right.

Regardless of the desired level of difficulty, it is essential that alternative questions have roughly the same level of difficulty. Each form will have a large number of multiple-choice questions and so parallel forms can be constructed by choosing a similar mix of easy and hard questions for each form. As currently envisioned, each form will have only two essay prompts, only one of which will be scored against the reading standards. There will not be much opportunity to balance easy and difficult questions, so the essay questions to be used in different forms must all have about the same difficulty.

The development contractor was highly successful in creating writing prompts and scoring guides of similar difficulty. Across the 24 prompts, the average scores ranged from 2.1 to 2.5, with most falling between 2.2 and 2.4. Scores were somewhat more variable across the 12 prompts scored for reading, ranging from 1.3 to 2.9. We would recommend dropping or revising 3 of the 12 reading prompts where half or more of all responses were assigned a score of 1. The minimum average for the remaining prompts was 1.9. If equal difficulty were the only goal, two prompts with average scores of 2.9 could also be dropped, leaving a range of score averages from 1.9 to 2.4. Overall, however, the prompts and scoring guides appear to be too difficult and so we would not recommend dropping the two easier prompts.

Item-Total Correlation. As with the MC questions, a second criterion in evaluating the quality of the essay questions is the extent to which scores on the question are consistent with information provided by all of the other questions. For MC questions, we looked for questions with an item-total score correlation[2] less than .20. Essay questions require a considerably greater investment of student time in responding and result in a maximum of 4 possible points compared to just 1 for multiple-choice questions. We expect more information from each essay score. Consequently, we chose to flag questions with item-total score correlations less than .40 as being inefficient.

Again, the writing scores were all highly efficient. Item-total correlations ranged between .53 and .73, all highly acceptable. The item-total correlation for one of the reading questions was .36, clearly below our cutoff. Item-total correlations for the remaining reading questions ranged from .42 to .60.

Disparate Impact. Table 2.2 shows average English/language arts MC and essay question scores for the different demographic groups typically included in disparate impact analyses. There were 100 MC questions, each scored 1 for correct responses or 0 for incorrect or omitted responses, so MC scores ranged from 0 to 100. Each student received three essay question scores, with each score ranging from 1 to 4, so the essay scores ranged from 3 to 12. Students who did not respond to both questions were excluded from these analyses. Table 2.2 also shows the standard deviation (SD) of the scores for each group. The standard deviation is a measure of how much the scores vary from the average. Roughly two-thirds of the scores will fall into the range running from one standard deviation below the average to one standard deviation above the average. The bottom half of Table 2.2 shows standardized differences. These are the difference between the average for a particular group and the overall average divided by the overall standard deviation. The purpose of this transformation is to provide comparisons across question formats that are adjusted for differences in the ranges of scores for these two formats.

The pattern of group differences for the essay questions is very similar to the pattern for the MC questions. More importantly, the differences among groups are not larger for the essay questions when converted to a common (standard deviation) metric.

[2] A correlation coefficient indicates the level of agreement between two measures. It ranges from -1.0 for perfect disagreement, where above-average scores on one measure are always accompanied by equally below-average scores on the second measure, to +1.0 for perfect agreement. The correlation coefficient will be 0.0 if there is no relationship between the two measures.

Table 2.2
Average Multiple Choice and Essay Scores by Demographic Group

Group	Multiple Choice (MC) (Percent Pass)			Essay Questions (Average Student Score)		
	Number	Average	SD	Number	Average	SD
ALL	3767	59.20	20.73	2997	6.94	2.13
Hispanic	1316	51.15	18.51	992	6.19	1.92
African American	197	50.51	20.22	138	6.33	2.06
Female	1840	63.26	19.14	1555	7.29	2.04
English language learners	430	40.73	14.60	265	5.62	1.78
Standardized Differences:						
Hispanic		-0.39	-0.11		-0.35	-0.10
Black		-0.42	-0.02		-0.29	-0.03
Female		0.20	-0.08		0.16	-0.04
English language learner		-0.89	-0.30		-0.62	-0.16

Table 2.3 provides very preliminary information on passing rates. To be sure, students will not pass or fail the essay questions separately; but scores on these questions will contribute to overall pass and fail decisions. It seems unlikely that performance on the essay questions will be considered satisfactory for students scoring 1 or 2 on the 4-point scale. We examined the effects of requiring an average score of 2.5 or a total score of 7.5 across the three essay scores for satisfactory performance. Overall, only 43 percent of the students would meet this criterion for the essay questions included in the Spring 2000 Field Test. The pass-rate for Hispanic and African American students would be less than 30 percent and the pass rates for students identified as English language learners would be less than 20 percent. Note also that the percentage of students not responding to one or both of the essay questions was significantly higher for the lower scoring groups. Overall, 80 percent of the students responded to both essay prompts. For African-American students, only 70 percent responded to both prompts and for English Language learners the figure was only about 60 percent.

Table 2.3
Percent with "Passing" Essay Question Scores by Demographic Group

Group	% Of Scores > 7.5 (If both essay questions answered)	% Missing One Essay	% Missing Both Essays
ALL	42.6	11.5	8.9
Hispanic	27.7	13.7	10.9
African American	29.7	15.7	14.2
Female	48.6	8.5	7.0
English language learners (ELL)	19.2	19.5	18.8

Differential Item Functioning. We used two relatively direct measures of differences across groups in the scores for each essay question. Other, more sophisticated indicators of group differences for multi-level scores have been identified (e.g., Zwick, Thayer & Mazzeo, 1997), but generally require larger sample sizes. Note that only about 300 students responded to each individual question. Across the 12 forms and subforms the number of students with valid responses to the essay questions (not blank, off-topic, or illegible) ranged

from 242 to 341. The number of females ranged from 109 to 190 and the number of Hispanic students in each of these samples ranged from 76 to 121. The numbers for other demographic groups were generally less then 30, far too small for useful analyses.

First, we looked at group differences in average scores for each question relative to the average of these group differences across all questions. Across all of the questions, the average essay question score for Hispanic students was .2 less than the average score for all students. We flagged one reading question with a significantly greater difference (.46). All of the other questions had average score differences of .4 or less. No large differences were found by gender.

We examined group differences in item-total correlations as a second indicator of differential item functioning. The same reading question that showed a large mean difference for Hispanic students also had a significantly lower item-total correlation for these students (.36 compared to .59 for all students). All of the other differences in item-total correlations were less than .2. Also, all of the item-total correlations for Hispanic students were well above zero, indicating that the essay questions did function effectively for these students.

Statistical Screening Summary. All of the writing-only prompts passed all of the item screens. Writing scores for the dual use prompts also passed all of the screens. Table 2.4 summarizes the number of reading essay questions flagged for different statistical reasons. Overall 5 of the 12 reading questions were flagged. Editorial review may suggest that some of these questions are perfectly valid, so this represents a worst-case scenario. Overall the survival rate (percent of questions not flagged) for the writing questions was exceptional and the survival rate for reading was above 50%. It is quite common to find significantly lower survival rates in other similar programs. Note, however, that the statistical criteria for screening these questions were limited by sample size. We could not, for example, examine differential item functioning for African-American students, students with disabilities, or English language learners.

Table 2.4

Summary of Item Screening Results: Essay Reading Questions

Statistical Screen	Number Flagged	Booklet Number(s)
Low Passing Rates	3	3.2, 4.2, 4.3
Low Item-Total Correlation	1	1.3
DIF: Passing Rates	1	1.2
DIF: Item-Total Correlations	1	1.2
Total Flagged	5	1.2, 1.3, 3.2, 4.2, 4.3

Rater Agreement in Scoring the Essay Questions

Each essay was scored by two independent raters. Table 2.5 indicates the level of agreement of the two raters for each response. Entries in Table 2.5 show the number of papers receiving each possible combination of scores from the two independent raters across all of the students and essay questions. Counts on the diagonal of this table indicated the number of times the two raters gave the same score. In most cases where the two raters gave

different scores, the scores were in adjacent categories, meaning that they differed by only one score point. The level of agreement is quite high as summarized in Table 2.6, which shows agreement level, by type of prompt and overall. There were, however, a small number of very dramatic differences where one rater assigned a score of 1 while the other assigned a score of 4. In an operational program, there is usually an "adjudication" process where disagreements of more than one score point are resolved by a third, typically more senior, rater.

Table 2.5
Counts of Essay Scores Assigned by Each Rater

Score assigned by the 1st rater	Score assigned by the 2nd rater							
	Valid Responses				Invalid Responses			Total
	1	2	3	4	Off-Topic	Blank	Illegible	
1	2045	285	11	6	1	0	0	2348
2	258	2704	371	16	0	2	0	3351
3	12	416	2605	104	0	0	0	3137
4	3	22	173	741	0	0	0	939
Off-Topic	2	2	0	0	318	8	0	330
Blank	0	0	1	0	16	1175	0	1192
Illegible	0	0	0	0	3	0	1	4
Total	2320	3429	3161	867	338	1185	1	11301

Table 2.6
Percent Agreement on Valid Responses by Question Type

Score Discrepancy	Type of Essay Question			
	Dual Score Prompts		Writing Only	All Prompts
	Reading	Writing		
Exact Match	84.9 %	81.0 %	82.5 %	82.8 %
1-Category	13.3 %	18.8 %	17.2 %	16.4 %
2+ Categories	1.8 %	0.1%	0.2 %	0.7 %
Total	100.0 %	100.0 %	100.0 %	100.0 %

We conducted a generalizability analysis (see Shavelson & Webb, 1991) as a final indication of the impact of discrepancies across scorers. In generalizability analyses, an estimate of true variation in student scores is compared to variation across different questions and scorers. The resulting information provides a basis for estimating the reliability of scores for different numbers of questions and scorers. We conducted a 3-question by 2-rater analysis of variance (see Scheffe, 1959) for each of the 12 form/subform combinations[3]. In these analyses, score variation by student is what we are trying to measure and so is labeled as "true" variance. The remaining sources of variation in scores are considered "error." Differences in the average score for each question is a source of error that will be eliminated through test form equating analyses. Interaction terms, such as student by question (S*Q), indicate the extent to which some students score higher on some questions while other

[3] We could not tell from available data exactly how many different scorers there were for each question or the extent to which the same or different scorers were used for different questions. We ran a variety of analyses with different assumptions about how scorers were nested within questions or students. Estimates for different sources of error varied slightly across these analyses, but the overall reliability estimates were essentially the same.

students score higher on other questions. Interactions between students and raters and between raters and questions are defined similarly. All are sources of error in the measurement of overall student achievement levels. Table 2.7 shows estimates for these different sources of score variation averaged across these 12 analyses.

Table 2.7
Sources of Variation in Scores for Essay Questions

Source	Type	Score Variance	% of Total
Student (S)	True	0.354	46.9 %
Question (Q)	Ignored[1]	0.102	N/A
Rater (R)	Error	0.000	0.0 %
S*Q	Error	0.303	40.2 %
S*R	Error	0.013	1.7%
Q*R	Error	0.000	0.0 %
S*Q*R	Error	0.083	11.0 %
TOTAL	True+Error	0.754	100 %

[1] Differences in question difficulties (main effects due to question) will be eliminated through equating.

Table 2.8 shows the "design" portion of the analyses, estimating the reliability for different numbers of questions and scorers. Reliability is a measure of score accuracy. It is equal to the ratio of "true" variation to the total variation in scores. When each rating of each response was considered separately, 47 percent of the total variation was "true" (between-student) variance so the reliability of a score from a single question would be .47. In the design analyses, statistical formulae are applied to estimate the reliability of scores that are averages across more than one question and/or more than one rater.

The overall reliability estimates are high considering that only three essay scores are included. Overall reliability, combining both MC and essay scores, is consistently in the range of .96. The proposed design of using two prompts to generate 3 scores increases the reliability of the essay question scores considerably in comparison to a single score from a single prompt. Adding a second rater does not increase the overall reliability very much. We do not, however, recommend using only a single scorer for each response. Because of the high-stakes nature of the individual student scores, a process for identifying and eliminating inconsistencies in scoring essay responses will be important. The few cases where one rater assigned a score of 1 while the other assigned a score of 4 illustrate, dramatically, the need for identifying (through multiple raters) and resolving (through a third reading) score discrepancies.

Table 2.8
Estimated Score Reliability by Number of Questions and Raters

	Number of Raters	
Number of Questions	1	2
1	.47	.50
2	.63	.66
3	72	.74

Revised Estimates of Test Accuracy

Overall, the accuracy of scores from the HSEE questions is likely to be quite high. The reliability estimate of .96 for ELA total scores means that the amount of measurement error is small (4 percent of the total score variation). This figure is quite good in comparison to most standardized tests. Reliabilities greater than .80 are considered acceptable for many purposes. For high-stakes uses, reliability estimates of .90 or higher are more commonly required.

Even with very high overall reliabilities, there will still be some inaccuracy in making pass-fail decisions based on a single score. Our June 30 report included extensive detail on analyses of the potential accuracy of HSEE total scores for ELA and mathematics when used to classify students as passing or failing. In this section we report the results of further analyses for ELA scores when essay scores are included in the total. We used an item response theory (IRT) model for multi-level scores, the Partial Credit Model (Muraki, 1992), to predict the distribution of scores for students resembling the field test participants. (See Wise, et al., June 30, for details on the procedures used.)

The new ELA scores include 100 MC questions plus 3 essay scores with 4 points each for a total of 112 possible scores. In these analyses, we assumed that blank and off-topic responses would be assigned a score of 0. Table 2.9 shows estimates of the percent of students who would score at different levels defined by plausible passing cutoffs. Again, we identified 50%, 60%, and 70% of the total possible score as plausible points for setting the minimum passing score. The addition of the essay scores leads to lower plausible passing rates in comparison to the prior analysis based on MC only. This difference should be interpreted cautiously, however, as higher omit rates for the essay questions may indicate lower effort on responses to these questions in this field-test setting.

Table 2.9
Number of Simulated Examinees at Different ELA Total Score Levels

Score Range	Minimum % Correct	Estimate % of Students	Estimated % Passing
0-55	0	38.0 %	
56-67	50	17.5%	62.0%
68-78	60	15.0 %	44.5%
79 – 112	70	29.5%	29.5%

There will, of course, always be some students whose true achievement level is right at the border between passing and failing. No test, no matter how reliable, can provide perfect classification for these students. To get an operational idea of what "near the border" might mean, we estimated the conditional standard errors (the standard error of measurement for students with a particular true number right score). Near the middle of the score range, these standard errors were 4.9 score points. To illustrate classification accuracy for students of different true achievement levels, we used the conditional standard error estimates to define a zone of uncertainty where student's true achievement was very near the pass-fail border. Table 2.10 shows the number of students expected to have true achievement scores more than 5 points below or above a minimum score of 56.0. For each true achievement level, we estimated the number of students whose observed score from a single testing would be above

or below 56. We classified these results as either correct or incorrect classifications, depending on whether the observed score level agreed with the examinee's true score. The results indicate that most students (more than 85%) would be clearly above or below the minimum and, for these students classification accuracy would be very high (98%–99%). For the 14.5% of students who are very near the minimum score level, about 70% (64% for those just below the passing level and 74 % for those just above) would be classified correctly.

Table 2.10
Estimated Percent Scoring Below/Above 56 Score Points by True Score Level

Subject	True (Expected) Number Correct	Percent of all Students	Percent of These Students Who Would Actually Score:	
			< 56	56+
ELA	00.00–51.99	30.5	97.8	2.2
	52.00-55.99	7.5	64.1	35.9
	56.00–59.99	7.0	26.6	73.4
	60.00+	55.0	0.8	99.2

Characteristics of the Field Test Samples

One important question is how well the students who participated in the field test represented the population of 10th grade public school students in California. AIR used 1999 STAR data to select representative samples of 100 schools each for the mathematics and English/language arts field tests. They then hoped to test 66 students from each school. The actual student participation rate varied considerably across schools and it is possible that more students participated from high (or low) performing schools than from low (or high) performing schools leading to a bias in estimates of student achievement levels. We conducted additional analyses to determine the extent to which this might be the case.

Table 2.11 shows a comparison of 10th grade STAR scores from spring 2000 for all schools in California and for schools participating in the HSEE field tests. Averages for all schools were weighted by the number of 10th graders in each school to generate averages for all students. This was the target against which results for the school and student samples were compared. Estimates for the schools in each of the field test samples were generated in two ways. First, the simple average of the school means was computed. This reflects the representativeness that would have resulted if the same number of students were tested from each school. Second, the means for each school were weighted by the number of field test participants from that school to provide an estimate of the effects of differential participation across schools. The results indicate a close correspondence with statewide averages (the first row in the table). There was a slight tendency to over-represent above-average schools in the ELA sample and a slight tendency to under-represent schools at very high and very low levels in both samples. Overall, however, these effects are slight.

Table 2.11

Comparison of FT Examinees to Statewide Averages: STAR 2000 Means and Standard Deviations

Population/Sample	Mathematics Average	SD	Reading Average	SD
Statewide – Weighted[1]	698	16.0	691	16.7
ELA Sample Schools	698	15.6	692	17.4
ELA Schools – Weighted[2]	699	14.6	694	16.2
Math Sample Schools	697	14.5	692	16.3
Math Schools – Weighted[2]	699	14.0	694	15.2

[1] Average STAR scores for each school in the state were weighted by the number of students in the school to compute the average score for all students. The standard deviation (SD) column shows the standard deviation of school averages when these weights are used.

[2] Average STAR scores for each participating school were weighted by the number of participants from that school to estimate average STAR scores for all of the students in the field test sample.

Tables 2.12 and 2.13 show similar comparisons using the 1999 STAR data, including demographic information that was not yet available for the 2000 STAR data. The field test samples were quite similar to the statewide averages for STAR reading and math scores. The demographic comparisons, however, show some under-representation of schools with higher proportions of Hispanic students, particularly for the math sample. For the demographic variables, we also have the responses from each of the students participating in the field test. Estimates of the percentage of Hispanic students based on these responses agree closely with percentages estimated from the overall school percent. This suggests that the students tested in each school were representative of the school as a whole, at least in this one respect. The percentage of English language learner (ELL) students tested in each school were slightly lower than percentages estimated from overall school percents, suggesting that ELL students were slightly underrepresented in the students tested from each school.

Table 2.12

Comparison of FT Examinees to Statewide Averages: STAR 1999 Means and Standard Deviations

Population/Sample	Mathematics Average	SD	Reading Average	SD
Statewide – Weighted[1]	697	16.3	690	16.6
ELA Sample Schools	696	15.2	691	16.4
ELA Schools – Weighted[2]	688	13.9	692	15.8
Math Sample Schools	696	14.4	691	16.5
Math Schools – Weighted[2]	698	13.5	693	15.4

See footnotes for table 2.11.

Table 2.13
Comparison of FT Examinees to Statewide Averages: Key 1999 10th Grade Demographics

Population/Sample	% Hispanic	% English language learners
Statewide – Weighted[1]	39	16
ELA Sample Schools	34	15
ELA Schools – Weighted[2]	35	16
ELA - Students Tested	36	14
Math Sample Schools	30	12
Math Schools – Weighted[2]	29	12
Math - Students Tested	29	10

See footnotes for table 2.11.

Summary

Scores for the ELA essay questions were analyzed to determine the quality of these questions and their scores. Five of the 12 reading questions were flagged for one or more potential statistical problems. None of the writing questions were flagged. Note however, that responses are available for only about 300 students for each question. A consequence was that analyses of differential item functioning for different demographic groups were quite limited.

Scoring consistency was analyzed and found to be quite high. Psychometric results suggested that a single read of each response by scorers might provide sufficient accuracy since the essay scores constitute only a small portion of the total scores. A more elaborate process may still be called for, however, to minimize challenges to results for individual students who end up just below the passing level.

We estimated the accuracy of ELA test scores and found to it be quite similar to the estimate provided in our June 30 report. These estimates were based on simulations that involved a number of assumptions. After key decisions about scoring and reporting are made and an intact form is administered under operational conditions, estimates of score accuracy involving fewer assumptions can be computed.

The schools participating in the field test appeared to be closely representative of the state as a whole. Student participation rates did not seem to be related to school performance means in a way that would bias estimates from the field test sample. However, the impact of within-schools non-participation and also of student motivation could not be estimated from available data.

CHAPTER 3: OPPORTUNITY TO LEARN

Introduction

Two item review workshops were conducted during Year 1 of the evaluation effort. Educators from districts in our longitudinal study met in northern or southern California to review the test items (questions) that were included in the Spring 2000 HSEE Field Test. For each question, workshop participants were asked to rate the degree to which the question measured the targeted content standard and to estimate the proportion of students in their schools who received sufficient instruction to answer the question correctly. The specific procedures used and detailed information on the initial findings are described in the June 30 report (Wise, et al., 2000). Three main findings emerged from the item review workshops:

1. Consistent with review by the HSEE Panel and its Technical Committee, the questions were found to measure the target standards well.
2. For a significant proportion of the questions, our workshop participants estimated that more than 25% of all students had not received instruction that would ensure their ability to answer the question correctly.
3. For the questions reviewed, there was a significant relationship between the curriculum alignment (opportunity to learn) ratings collected in the item review workshops and the passing rates in the field test.

Two additional questions were addressed in our supplemental analyses. These questions arose in discussions with the Board and the HSEE Panel about our original Year 1 report. They were:

1. How representative were the 12 districts that participated in the item review workshops? Is there any reason to expect that students from these districts would be significantly more or less well prepared for the HSEE in comparison to students from the state as a whole?
2. Which specific standards had the most significant indications of opportunity-to-learn problems?

In addition to the supplemental analyses reported here, we have been asked to collect additional information on opportunities to learn the HSEE content standards from all districts serving California high school students. We are developing a new survey timed to reach districts shortly after the Board is expected to formally adopt specific HSEE content standards at its September meeting. The survey will go beyond collecting baseline information on the current status of curriculum and instruction, requesting information on planned changes in response to HSEE requirements as well. Preliminary results from this survey are expected to be available for the Board's November meeting.

Comparison of Participating Districts to the State as a Whole

The primary focus of our item rating panels was on judging the content match of the field test questions to test specifications; however, we also used the opportunity to ask about ongoing instruction related to those questions. For the item content judgments, we were

assessing test questions and did not need to be particularly concerned about how well our workshop participants represented the whole state. On the other hand, the curriculum-alignment (opportunity-to-learn) ratings are an assessment of districts. Therefore, how well the participants represent the state is an important concern. We did indicate in the previous report that, on a question-by-question basis, our workshop participants' ratings did predict students' field test performance. Questions rated with relatively low "curriculum alignment" ratings tended to be the questions on which students performed the worst in the HSEE field test.

The tables that follow describe the extent to which the districts of our workshop participants are representative of all districts in the state. "Target sample" in these tables refers to the 24 districts selected for our longitudinal study of the HSEE. Our entire target sample was invited to send representatives to the workshops. "Participating districts" indicate the districts that did attend. Table 3.1 shows a comparison of the target and participating districts in terms of the measures we used in selecting the target sample. Half of the districts in our target sample were classified as having above average number of English language learners (High ELL). Of the districts participating in the ELA panels, 57 percent were High ELL. Similarly, 48% of the districts participating in the mathematics panels were High ELL. While most of the comparisons in Table 3.1 indicate close agreement between the target and participating districts, there were a few significant differences. More of the districts participating in the mathematics panels were classified as "High Math" based on 1999 STAR scores (64% compared to 50%). Somewhat fewer of the participating districts were classified as "Small" (25% and 24% compared to 33%). Also fewer of the districts participating in the ELA panels were classified as large (24% compared to 33%) resulting in overrepresentation of middle-sized districts (62% compared to 33%).

Table 3.1
Comparison of Participating Districts to the Target Sample of Districts

District Statistics	Target Sample	Participating Districts[1]	
		ELA	Math
Percent High English Lang. Learners	50	57	48
Percent High STAR 1999 Math	50	52	64
Percent in Large Districts	33	24	32
Percent in Medium Districts	33	62	44
Percent in Small Districts	33	25	24
Number of Districts	24	12	12

Tables 3.2 and 3.3 show comparisons of achievement scores from the Standardized Testing and Reporting (STAR) Program for the participating and target districts and for all districts in the state. Table 3.2 shows comparisons of average 10th grade mathematics and reading scores on the 2000 STAR. This information was not available when the sample was selected in March 2000, but it now provides the most up-to-date basis for comparison on key measures of student achievement. The first row of each table shows the averages for all schools, with each school weighted by the number of 10th grade students so that the averages

are average scores for all 10th grade students in the state. The remaining rows provide means for the target and participating districts, with the participating districts weighted by the number of participants. For mathematics, the statewide average was 698, the average for districts in the target sample was 700 and the average for districts participating in the item review workshops was 699. For the ELA panels, the corresponding numbers were 691 for the statewide average, 693 for the target sample, and 690 for the workshops participants. Table 3.2 also shows standard deviations that indicate the degree of variation in average scores across districts. The standard deviations were quite a bit smaller for the participating districts (7 to 10 compared to 19 for the state as a whole). This means that we had fewer very high scoring and fewer very low scoring districts in comparison to the state as a whole.

Table 3.2
Comparison of Participating Districts to Statewide STAR 2000 Results

Population/Sample	Mathematics		Reading	
	Average	Standard Deviation	Average	Standard Deviation
All Districts[1]	698	19.1	691	19.2
Target Sample[1]	700	12.9	693	14.1
ELA Workshop Participants[2]	697	8.1	691	9.9
Math Workshop Participants[2]	699	7.1	694	9.3

[1] Average scores for each district were weighted by the number of 10th grade students in the district to get overall averages.
[2] Average scores for each district were weighted by the number of workshop participants from the district to get overall averages.

For completeness, Table 3.3 shows comparisons of district averages for the 1999 STAR scores that were used in selecting the target districts. These results were very similar to the comparisons based on the 2000 STAR scores. Statewide means were 697 and 690 for mathematics and reading respectively while the corresponding means for the participating districts were 698 and 690. The differences in standard deviations between the participating districts and the state as a whole were quite a bit smaller than was the case for the STAR 2000 results.

Table 3.3
Comparison of Participating Districts to Statewide STAR 1999 Results

Population/Sample	Mathematics		Reading	
	Average	Standard Deviation	Average	Standard Deviation
All Districts[1]	697	12.5	690	13.2
Target Sample[1]	698	11.0	692	12.6
ELA Workshop Participants[2]	695	8.3	690	10.3
Math Workshop Participants[2]	698	8.2	693	10.3

[1] Average scores for each district were weighted by the number of 10th grade students in the district to get overall averages.
[2] Average scores for each district were weighted by the number of workshop participants from the district to get overall averages.

Table 3.4 presents a final comparison of participating districts to statewide figures. We computed the percentage of students who were Hispanic and the percentage who were

English language learners (ELLs) from the 1999 STAR data. For the state as a whole, the percentages were 39 percent and 16 percent respectively. For the target sample of districts, the corresponding figures were slightly higher (43 percent and 18 percent). For the districts participating in our workshops, the figures were lower. The percentage of Hispanic students in districts participating in the mathematics panels was considerably lower than the statewide figure (29 percent compared to 39 percent).

Table 3.4
Comparison of 1999 Demographics for Participating Districts to Statewide Figures.

Population/Sample	% Hispanic	% English Language Learners (ELL)
All Districts[1]	39	16
Target Sample[1]	43	18
ELA Workshop Participants[2]	33	14
Math Workshop Participants[2]	29	13

[1] Average scores for each district were weighted by the number of 10th grade students in the district to get overall averages.
[2] Average scores for each district were weighted by the number of workshop participants from the district to get overall averages.

In summary, the districts participating in our item review workshops had average mathematics and reading achievement scores that matched statewide averages closely. There was some evidence that very high or low scoring districts and districts with higher percentages of Hispanic students were less likely to have been included in our panels.

Content Standards Not Covered in the Current Curriculum

In our initial report, we counted the number of field test questions for which our workshop participants indicated more than 25% of their students had not received instruction needed to answer the question correctly. Our summary count indicated that more than 25% of the students would not have received sufficient instruction for 50% of the mathematics questions and 90% of the English-Language Arts questions. Tables 3.3 and 3.4 in our June 30 report also showed the number and percentage of questions with low curriculum-alignment ratings for each major content category. For ELA, just over 80 percent of the language convention questions and over 90 percent of the questions in all other content categories had low curriculum alignment ratings. For mathematics, the percentages of questions with low curriculum-alignment ratings ranged from 20 percent for number sense to 80 percent for Algebra 1. We did not provide corresponding statistics for the individual content standards within each of the general categories. Members of the HSEE Panel and others have suggested that more detailed information would be useful to them.

In Tables 3.5a, 3.5b and 3.6, we provide complete results for each of the English language arts (reading and writing) and mathematics content standards. Note that several of the standards in the two tables do not have data. We report data only for those standards for which more than one question was included on the field test and rated in the workshops. In addition, no field test performance data is presented for the writing applications standards. These standards are all measured with essay questions scored on a 4-point scale. No decision

has yet been made as to what score a student has to receive on this scale to have "passed" the standard. Thus, we could not compute a percent passing statistic for the questions used to assess the writing applications standards.

We have highlighted the standards with the lowest passing rates and lowest curriculum-alignment ratings by printing both the standard and the corresponding statistics in **boldface**. Specifically, we highlighted standards where both of the following were true:

1. The average percent passing across all questions developed to assess the standard was less than 40% for mathematics or less than 55% for ELA.
2. The average 10^{th} Grade curriculum-alignment (CA) rating for these questions was less than 2.0. Category 2 of the curriculum-alignment ratings was "50–75% of 10th graders are provided with instruction that would allow them to pass the question." If the average rating was less than 2.0, then the panelists were saying that fewer than 50 percent of their students had been provided adequate opportunity to learn the material covered by the question.

For ELA, the standards for which students are least well prepared involve higher order analysis skills. For example, the ELA standard with the very lowest curriculum-alignment rating was:

> 3.12 (Literary Criticism). Analyze the way in which a work of literature is related to the themes and issues of its historical period. (Assessed with essay questions so no passing rates are available; CA Rating=1.3)

The current proposal is to use essay questions in the assessment of this standard. It would be helpful for the HSEE panel to provide illustrations or examples of questions that might be used to assess this standard and a discussion of the guidelines for scoring responses to these questions. Such information will be critical in determining minimum passing scores as well as useful to districts in increasing their coverage of this standard in the curriculum.

Some other examples of ELA standards that appear particularly problematic are:

> 2.3 (Comprehension and analysis of grade-level-appropriate text). Generate relevant questions about readings on issues that can be researched. (% Pass=49, CA Rating=1.6)
>
> 3.1 (Literary Response and Analysis). Articulate the relationship between the expressed purposes and the characteristics of different forms of dramatic literature (e.g., comedy, tragedy, drama, dramatic monologue). (% Pass=53, CA Rating=1.8)
>
> 1.5 (Writing Strategies). Synthesize information from multiple sources and identify complexities and discrepancies in the information and the different perspectives found in each medium (e.g., almanacs, microfiche, news sources, in-depth studies, speeches, journals, technical documents). (% Pass=48, CA Rating=1.7)

The current proposal is to use multiple-choice questions to assess these standards. Unfortunately, because of test security concerns, we cannot provide examples of questions designed to assess these standards. The Panel, CDE, and AIR should give a high priority to providing detailed examples of how the skills identified in these standards might be demonstrated. CDE and the development contractor are working on an Educators Guide that could meet this need.

Table 3.5a
Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard—Reading

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
1.0 Word Analysis, Fluency, and Systematic Vocabulary Development		
Vocabulary and Concept Development		
1.1 Identify and use the literal and figurative meanings of words and understand word derivations.	61%	2.11
1.2 Distinguish between the denotative and connotative meanings of words and interpret the connotative power of words.	68%	2.10
2.0 Reading Comprehension (Focus on Informational Materials)		
Structural Features of Informational Materials		
8.2.1 Compare and contrast the features and elements of consumer materials to gain meaning from documents (e.g., warranties, contracts, product information, instructional manuals). [NOTE: This is a grade eight standard.]	61%	2.06
2.1 Analyze the structure and format of functional workplace documents, including the graphics and headers, and explain how authors use the features to achieve their purposes.	78%	2.35
2.2 Prepare a bibliography of reference materials for a report using a variety of consumer, work place, and public documents.	59%	1.67
Comprehension and Analysis of Grade-Level-Appropriate Text		
2.3 Generate relevant questions about readings on issues that can be researched.	**49%**	**1.61**
2.4 Synthesize the content from several sources or works by a single author dealing with a single issue; paraphrase the ideas and connect them to other sources and related topics to demonstrate comprehension.	61%	1.98
2.5 Extend ideas presented in primary or secondary sources through original analysis, evaluation, and elaboration.	63%	1.98

(table continues)

Table 3.5a

Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard—Reading (Continued)

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
Expository Critique		
2.7 Critique the logic of functional documents by examining the sequence of information and procedures in anticipation of possible reader misunderstandings.	64%	1.67
2.8 Evaluate the credibility of an author's argument or defense of a claim by critiquing the relationship between generalizations and evidence, the comprehensiveness of evidence, and the way in which the author's intent affects the structure and tone of the text (e.g., in professional journals, editorials, political speeches, primary source material).	**52%**	**1.82**
3.0 Literary Response and Analysis:		
Structural Features of Literature		
3.1 Articulate the relationship between the expressed purposes and the characteristics of different forms of dramatic literature (e.g., comedy, tragedy, drama, dramatic monologue).	**50%**	**1.97**
Narrative Analysis of Grade-Level-Appropriate Text		
3.3 Analyze interactions between main and subordinate characters in a literary text (e.g., internal and external conflicts, motivations, relationships, influences) and explain the way those interactions affect the plot.	64%	2.16
3.4 Determine characters' traits by what the characters say about themselves in narration, dialogue, dramatic monologue, and soliloquy.	63%	2.10
3.5 Compare works that express a universal theme and provide evidence to support the ideas expressed in each work.	68%	1.93
3.6 Analyze and trace an author's development of time and sequence, including the use of complex literary devices (e.g., foreshadowing, flashbacks).		
3.7 Recognize and understand the significance of various literary devices, including figurative language, imagery, allegory, and symbolism, and explain their appeal.	**53%**	**1.77**
3.8 Interpret and evaluate the impact of ambiguities, subtleties, contradictions, ironies, and incongruities in a text.	**54%**	**1.67**
3.9 Explain how voice, persona, and the choice of a narrator affect characterization and the tone, plot, and credibility of a text.	59%	1.96
3.10 Identify and describe the function of dialogue, scene designs, soliloquies, asides, and character foils in dramatic literature.	59%	1.88

(table continues)

Table 3.5a

Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard—Reading (Continued)

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
Literary Criticism		
8.3.7 Analyze a work of literature, showing how it reflects the heritage, traditions, attitudes, and beliefs of its author. (Biographical approach) [NOTE: This is a grade eight standard.]		
3.11 Evaluate the aesthetic qualities of style, including the impact of diction and figurative language on tone, mood, and theme, using the terminology of literary criticism. (Aesthetic approach)		
3.12 Analyze the way in which a work of literature is related to the themes and issues of its historical period. (Historical approach)		**1.29**

Table 3.5b

Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard—Writing

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
1.0 Writing Strategies (Grades 9-10):		
Organization and Focus		
1.1 Establish a controlling impression or coherent thesis that conveys a clear and distinctive perspective on the subject and maintain a consistent tone and focus throughout the piece of writing.	**51%**	**1.94**
1.2 Use precise language, action verbs, sensory details, appropriate modifiers, and the active rather than the passive voice.	**50%**	**1.82**
Research and Technology		
1.3 Use clear research questions and suitable research methods (e.g., library, electronic media, personal interview) to elicit and present evidence from primary and secondary sources.	60%	2.20
1.4 Develop the main ideas within the body of the composition through supporting evidence (e.g., scenarios, commonly held beliefs, hypotheses, definitions).	52%	2.17
1.5 Synthesize information from multiple sources and identify complexities and discrepancies in the information and the different perspectives found in each medium (e.g., almanacs, microfiche, news sources, in-depth field studies, speeches, journals, technical documents).	**48%**	**1.69**
1.6 Integrate quotations and citations into a written text while maintaining the flow of ideas.	57%	1.79

(table continues)

Table 3.5b

Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard—Writing (Continued)

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
Evaluation and Revision		
1.9 Revise writing to improve the logic and coherence of the organization and controlling perspective, the precision of word choice, and the tone by taking into consideration the audience, purpose, and formality of the context.	60%	1.90
2.0 Writing Applications (Genres and Their Characteristics)		
2.1 Write biographical or autobiographical narratives or short stories: a. Relate a sequence of events and communicate the significance of the e vents to the audience. b. Locate scenes and incidents in specific places. c. Describe with concrete sensory details the sights, sounds, and smells of a scene and the specific actions, movements, gestures, and feelings of the characters; use interior monologue to depict the characters' feelings. e. Make effective use of descriptions of appearance, images, shifting perspectives, and sensory details.	Essay	
2.2. Write responses to literature: a. Demonstrate a comprehensive grasp of the significant ideas of literary works. b. Support important ideas and viewpoints through accurate and detailed references to the text or to other works. c. Demonstrate awareness of the author's use of stylistic devices and an appreciation of the effects created. d. Identify and assess the impact of perceived ambiguities, nuances and complexities within the text.	Essay	
2.3 Write expository compositions, including analytical essays and research reports: a. Marshal evidence in support of a thesis and related claims, including information on all relevant perspectives. b. Convey information and ideas from primary and secondary sources accurately and coherently. c. Make distinctions between the relative value and significance of specific data, facts, and ideas. e. Anticipate and address readers' potential misunderstandings, biases, and expectations. f. Use technical terms and notations accurately.	Essay	2.27
2.4 Write persuasive compositions: **a. Structure ideas and arguments in a sustained and logical fashion.** **b. Use specific rhetorical devices to support assertions (e.g., appeal to logic through reasoning; appeal to emotion or ethical belief; relate a personal anecdote, case study, or analogy).** **c. Clarify and defend positions with precise and relevant evidence, including facts, expert opinions, quotations, and expressions of commonly accepted beliefs and logical reasoning.** **d. Address readers' concerns, counterclaims, biases, and expectations.**	**Essay**	**1.85**

(table continues)

Table 3.5b
Field Test Passing Rates and Curriculum-Alignment (CA) Ratings for each English/Language Arts Content Standard – Writing (Continued)

STRANDS/STANDARDS [BOLDED STANDARDS INDICATE PROBLEM AREAS]	PERCENT PASSING	AVERAGE CA RATING
2.5 Write business letters: **a. Provide clear and purposeful information and address the intended audience appropriately.** **b. Use appropriate vocabulary, tone, and style to take into account the nature of the relationship with, and the knowledge and interests of, the recipients.** **c. Highlight central ideas or images.** **d. Follow a conventional style with page formats, fonts, and spacing that contribute to the document's readability and impact.**	**Essay**	**1.80**
1.0 Written and Oral English Language Conventions (Grades 9 & 10):		
Grammar and Mechanics of Writing		
1.1 Identify and correctly use clauses (e.g., main and subordinate), phrases (e.g., gerund, infinitive, and participial), and mechanics of punctuation (e.g., semicolons, colons, ellipses, hyphens).	59%	2.35
1.2 Understand sentence construction (e.g., parallel structure, subordination, proper placement of modifiers) and proper English usage (e.g., consistency of verb tenses).	**49%**	**1.83**
1.3 Demonstrate an understanding of proper English usage and control of grammar, paragraph and sentence structure, diction, and syntax.	**53%**	**1.76**
Manuscript Form		
1.4 Produce Legible work that shows accurate spelling and correct use of the conventions of punctuation and capitalization.		
1.5 Reflect appropriate manuscript requirements, including title page presentation, pagination, spacing and margins, and integration of source and support material (e.g., in-text citation, use of direct quotations, paraphrasing) with appropriate citations.	50%	2.06

** Curriculum Alignment (CA) rating scale of how many students had the opportunity to learn this material in local district curriculum:

1. <50% 2. 50% – 74% 3. 75% – 94% 4. >95%

Information on the mathematics standards is shown in Table 3.6. It is not surprising that the Algebra 1 standards were most problematic. Currently, students are not required to take algebra to graduate in most districts. Within the Algebra 1 strand, it was particularly difficult to develop questions for some specific standards. Students responded at about chance level (that is, did not do better than random guessing) to all of the questions developed for the following Algebra 1 standards:

> 17.0 Students determine the domain of independent variables and the range of dependent variables defined by a graph, a set of ordered pairs, or a symbolic expression [an equation]. (% Pass=22, CA Rating=1.5)
>
> 23.0 Students apply quadratic equations to physical problems, such as the motion of an object under the force of gravity. (% Pass=33, CA Rating=1.7)
>
> 24.3 Students use counter examples to show that an assertion is false and recognize that a single counter example is sufficient to refute an assertion. (% Pass=26, CA Rating=1.6)

It was also difficult to write questions that many students could answer correctly for some mathematics reasoning standards. An example of a mathematical reasoning standard that had both low passing rates and low curriculum-alignment (CA) ratings was:

> 2.3 Estimate unknown quantities graphically and solve for them by using logical reasoning and arithmetic and algebraic techniques. (% Pass=23, CA Rating=2.1)

After reviewing questions for the fall field test at the July 2000 HSEE Panel meeting, one of the Panel members stated that she wished there was time to go back and clarify the content standards based on what she had learned from reviewing questions written to these standards. Information from field-test results and the curriculum-alignment ratings presented above might also be useful in developing explanatory material for specific standards. Such material is needed to help teachers align their instruction to these standards and to help students and parents understand more clearly the standards they are being asked to meet. Also, item writers can use this information to create questions that are clearly aligned to the content standards.

Table 3.6

Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
Statistics, Data Analysis, and Probability		
(Grade 6)		
1.0 Students compute and analyze statistical measurements for data sets:		
1.1 Compute the range, mean, median, and mode of data sets.	49%	3.17
2.0 Students use data samples of a population and describe the characteristics and limitations of the samples:		
2.5 Identify claims based on statistical data and, in simple cases, evaluate the validity of the claims.	54%	2.82
3.0 Students determine theoretical and experimental probabilities and use these to make predictions about events:		
3.1 Represent all possible outcomes for compound events in an organized way (e.g., tables, grids, tree diagrams) and express the theoretical probability of each outcome.	36%	2.24
3.3 Represent probabilities as ratios, proportions, decimals between 0 and 1, and percentages between 0 and 100, and verify that the probabilities computed are reasonable; know that if P is the probability of an event, 1-P is the probability of an event not occurring.	54%	2.56
3.5 Understand the difference between independent and dependent events.	**44%**	**1.93**
(Grade 7)		
1.0 Students collect, organize, and represent data sets that have one or more variables and identify relationships among variables within a data set by hand and through the use of an electronic spreadsheet software program:		
1.1 Know various forms of display for data sets, including a stem-and-leaf plot or box-and-whisker plot; use the forms to display a single set of data or to compare two sets of data.	56%	2.63
1.2 Represent two numerical variables on a scatter plot and informally describe how the data points are distributed and any apparent relationship that exists between the two variables (e.g., between time spent on homework and grade level).	57%	2.71
1.3 Understand the meaning of, and be able to compute the minimum, the lower quartile, the median, the upper quartile, and the maximum of a data set.	40%	2.17

(table continues)

Table 3.6
Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
Number Sense (Grade 7)		
1.0 Students know the properties of, and compute with, rational numbers expressed in a variety of forms:		
1.1 Read, write, and compare rational numbers in scientific notation (positive and negative powers of 10) with approximate numbers using scientific notation.	54%	2.76
1.2 Add, subtract, multiply, and divide rational numbers (integers, fractions, and terminating decimals) and take positive rational numbers to whole-number powers	60%	2.81
1.3 Convert fractions to decimals and percents and use these representations in estimations, computations, and applications.	49%	2.83
1.6 Calculate the percentage of increases and decreases of a quantity.	39%	2.40
1.7 Solve problems that involve discounts, markups, commissions, and profit and compute simple and compound interest.	38%	2.54
2.0 Students use exponents, powers, and roots and use exponents in working with fractions:		
2.1 Understand negative whole-number exponents. Multiply and divide expressions involving exponents with a common base.	29%	2.36
2.2 Add and subtract fractions by using factoring to find common denominators.	48%	2.38
2.3 Multiply, divide, and simplify rational numbers by using exponent rules.	59%	2.38
2.4 Use the inverse relationship between raising to a power and extracting the root of a perfect square integer; for an integer that is not square, determine without a calculator the two integers between which its square root lies and explain why.	47%	2.33
2.5 Understand the meaning of the absolute of a number; interpret the absolute value as the distance of the number from zero on a number line; and determine the absolute value of real numbers.	57%	2.33

(table continues)

Table 3.6

Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
Algebra and Functions (Grade 7)		
1.0 Students express quantitative relationships by using algebraic terminology, expressions, equations, inequalities, and graphs:		
1.1 Use variables and appropriate operations to write an expression, an equation, an inequality, or a system of equations or inequalities that represents a verbal description (e.g., three less than a number, half as large as area A).	51%	2.39
1.2 Use the correct order of operations to evaluate [simplify] algebraic expressions such as 3 $(2x+5)^2$.	65%	2.74
1.5 Represent quantitative relationships graphically and interpret the meaning of a specific part of a graph in the situation represented by the graph.	60%	2.72
2.0 Students interpret and evaluate expressions involving integer powers and simple roots:		
2.1 Interpret positive whole-number powers as repeated multiplication and negative whole-number powers as repeated division or multiplication by the multiplicative inverse. Simplify and evaluate expressions that include exponents.	51%	2.33
2.2 Multiply and divide monomials; extend the process of taking powers and extracting roots to monomials when the latter results in a monomial with an integer exponent.	32%	2.33
3.0 Students graph and interpret linear and some nonlinear functions:		
3.1 Graph functions of the form Y=nx2 and y=nx3 and use in solving problems.	**33%**	**1.93**
3.3 Graph linear functions, noting that the vertical change (change in-y value) per unit of horizontal change (change in x -value) is always the same and know that the ratio ("rise over run") is called the slope of a graph.	47%	2.41
3.4 Plot the values of quantities whose ratios are always the same (e.g., cost to the number of an item, feet to inches, circumference to diameter of a circle). Fit a line to the plot and understand that the slope of a line equals the [ratio of the] quantities.	**48%**	**1.90**

(table continues)

Table 3.6
Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
4.0 Students solve simple linear equations and inequalities over the rational numbers:		
4.1 Solve two-step linear equations and inequalities in one variable over the rational numbers, interpret the solution or solutions in the context from which they arose, and verify the reasonableness of the results.	62%	2.59
4.2 Solve multi-step problems involving rate, average speed, distance, and time or a direct variation.	43%	2.05
Measurement and Geometry (Grade 7)		
1.0 Students choose appropriate units of measure and use ratios to convert within and between measurement systems to solve problems:		
1.1 Compare weights, capacities, geometric measures, times, and temperatures within and between measurement systems (e.g., miles per hour and feet per second, cubic inches to cubic centimeters).	44%	2.47
1.2 Construct and read drawings and models made to scale.	43%	2.38
1.3 Use measures expressed as rates (e.g., speed, density) and measures expressed as products (e.g., person-days) to solve problems; check the units of the solutions; and use dimensional analysis to check the reasonableness of the answer.	67%	2.80
2.0 Students compute the perimeter, area, and volume of common geometric objects and use the results to find measures of less common objects. They know how perimeter, area and volume are affected by changes of scale:		
2.1 Use formulas routinely for finding the perimeter and area of basic two-dimensional figures and the surface area and volume of basic three-dimensional figures, including rectangles, parallelograms, trapezoids, squares, triangles, circles, prisms, and cylinders.	42%	2.69
2.2 Estimate and compute the [surface] area of more complex or irregular two-and three-dimensional figures by breaking the figures down into more basic geometric objects.	46%	2.36

(table continues)

Table 3.6

Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
2.3 Compute the length of the perimeter, the surface area of the faces, and the volume of a three-dimensional object built from rectangular solids. Understand that when the lengths of all dimensions are multiplied by a scale factor, the surface area is multiplied by the square of the scale factor and the volume is multiplied by the cube of the scale factor.	45%	2.11
2.4 Relate the changes in measurement with a change of scale to the units used (e.g., square inches, cubic feet) and to conversions between units (1 square foot = 144 square inches or [1 ft^2] = {144 in^2}, 1 cubic inch is approximately 16.38 cubic centimeters or [1 in^3] = [16.38 cm^3].	44%	2.26
3.0 Students know the Pythagorean theorem and deepen their understanding of plane and solid geometric shapes by constructing figures that meet given conditions and by identifying attributes of figures:		
3.2 Understand and use coordinate graphs to plot simple figures, determine lengths and areas related to them, and determine their images under translations and reflections.	**42%**	**1.96**
3.3 Know and understand the Pythagorean theorem and its converse and use it to find the length of the missing side of a right triangle and the lengths of other line segments and, in some situations, empirically verify the Pythagorean theorem by direct measurement.	42%	2.18
3.4 Demonstrate an understanding of conditions that indicate two geometrical figures are congruent and what congruence means about relationships between the sides and angles of the two figures.	52%	1.94
Mathematical Reasoning (Grade 7)		
1.0 Students make decisions about how to approach problems:		
1.1 Analyze problems by identifying relationships, distinguishing relevant from irrelevant information, identifying missing information, sequencing and prioritizing information, and observing patterns.	54%	2.55
1.2 Formulate and justify mathematical conjectures based on a general description of the mathematical question or problem posed.	42%	2.36

(table continues)

Table 3.6
Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
2.0 Students use strategies, skills, and concepts in finding solutions:		
2.1 Use estimation to verify the reasonableness of calculated results.	52%	2.93
2.3 Estimate unknown quantities graphically and solve for them by using logical reasoning and arithmetic and algebraic techniques.	23%	2.1
2.4 Make and test conjectures by using both inductive and deductive reasoning.	55%	2.31
3.0 Students determine a solution is complete and move beyond a particular problems by generalizing to other situations:		
3.1 Evaluate the reasonableness of the solution in the context of the original	47%	2.47
3.3 Develop generalizations of the results obtained and the strategies used and apply them to new problem situations.	57%	2.23
Algebra 1		
1.0 Students identify and use the arithmetic properties of subsets of integers and rational, irrational, and real numbers, including closure properties for the four basic arithmetic operations where applicable.		
2.0 Students understand and use such operations as taking the opposite, finding the reciprocal, and taking a root, and raising to a fractional power. They understand and use the rules of exponents.	43%	2.38
3.0 Students solve equations and inequalities involving absolute values.	35%	2.19
4.0 Students simplify expressions before solving linear equations and inequalities in one variable, such as $3(2x-5) + 4(x-2) = 12$.	40%	2.45
5.0 Students solve multi-step problems, including word problems, involving linear equations and linear inequalities in one variable and provide justification for each step.	43%	2.23
6.0 Students graph a linear equation and compute the x-and y- intercepts (e.g., graph $2x + 6y = 4$). They are also able to sketch the region defined by linear inequality (e.g., they sketch the region defined by $2x + 6y < 4$).	40%	2.31

(table continues)

Table 3.6

Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
7.0 Students verify that a point lies on a line, given an equation of the line. Students are able to derive linear equations by using the point-slope formula.	39%	2.24
8.0 Students understand the concepts of parallel lines and perpendicular lines and how those slopes are related. Students are able to find the equation of a line perpendicular to a given line that passes through a given point.	44%	2.24
9.0 Students solve a system of two linear equations in two variables algebraically and are able to interpret the answer graphically. Students are able to solve a system of two linear inequalities in two variables and to sketch the solution sets.	**44%**	**1.86**
10.0 Students add, subtract, multiply, and divide monomials and polynomials. Students solve multi-step problems, including word problems, by using these techniques.	**31%**	**1.55**
15.0 Students apply algebraic techniques to solve rate problems, work problems, and percent mixture problems.	**29%**	**1.76**
16.0 Students understand the concepts of a relation and a function, determine whether a given relation defines a function, and give pertinent information about given relations and functions.	32%	2.14
17.0 Students determine the domain of independent variables and the range of dependent variables defined by a graph, a set of ordered pairs, or a symbolic expression [an equation].	**22%**	**1.50**
18.0 Students determine whether a relation defined by a graph, a set of ordered pairs, or a symbolic expression [an equation] is a function and justify the conclusion.	**33%**	**1.94**
21.0 Students graph quadratic functions and know that their roots are the x-intercepts.	**27%**	**1.53**
23.0 Students apply quadratic equations to physical problems, such as the motion of an object under the force of gravity.	**33%**	**1.65**
24.0 Students use and know simple aspects of a logical argument:		
24.2 Students identify the hypothesis and conclusion in logical deduction.		
24.3 Students use counter examples to show that an assertion is false and recognize that a single counter example is sufficient to refute an assertion.	**26%**	**1.60**

(table continues)

Table 3.6

Opportunity to Learn Ratings and Field Test Performance for Mathematics Strands (Continued)

STRANDS/STANDARDS	PERCENT PASSING	AVERAGE CA RATING
25.0 Students use properties of the number system to judge the validity of results, to justify each step of a procedure, and to prove or disprove statements:		
25.1 Students use properties of numbers to construct simple, valid arguments (direct and indirect) for, or formulate counterexamples to, claimed assertions.		
25.2 Students judge the validity of an argument according to whether the properties of the real number system and the order of operations have been applied correctly at each step.		
25.3 Given a specific algebraic statement involving linear, quadratic, or absolute value expressions or equations or inequalities, students determine whether the statement is true sometimes, always, or never.		

** Curriculum Alignment (CA) rating scale of how many students had the opportunity to learn this material in local district curriculum:

1. <50% 2. 50% – 74% 3. 75% – 94% 4. >95%

CHAPTER 4: PRINCIPAL AND TEACHER SURVEYS

Introduction

The Year 1 Evaluation Report (Wise, et al., 2000) contained a chapter describing preliminary analyses of data (Chapter 5) from our Teacher and Principal Surveys. Due to time constraints, it was necessary to analyze a subset of survey responses at that time; analyses were restricted to surveys that had been returned by June 19, 2000. This supplemental report includes all surveys received (i.e., an additional 8 principal surveys and 45 teacher surveys). This chapter has been written as a replacement for Chapter 5 in our initial report, rather than a supplement. In order to facilitate comparison, the original chapter structure has been kept intact. Findings are little changed by inclusion of the additional surveys, but the descriptions here provide a more complete representation of the opinions of principals and teachers currently working in California high schools. We have included an additional analysis of the representativeness of the schools responding to each survey.

Educational reform efforts such as California's high school exit examination will exert an impact beyond just the receipt of a standards-based diploma. By providing feedback about student performance, the reform will serve as a catalyst for change throughout districts and schools. 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 questions 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 (Wise, et al., 2000, Chapter 2) were critical in formulating the 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 the Board has not yet made final decisions on the nature and content of the 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, we used an understanding of the process of "early and late planners and implementers" to develop survey questions.

Survey Development

The following are the main questions addressed in this baseline data collection:

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, including those with special needs and English language learners, for the first administration of the HSEE
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 Wise, et al. (2000). The resulting sample for the principal and teacher 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 one to six high schools we had selected to represent their districts. Based on this information, principal and teacher survey packets were shipped in early May 2000 to 84 schools to the attention of the principal or POC. The packets 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 mathematics courses, and two 9th or 10th 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 each of the survey instruments is included in Appendix A.

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 encourage completion of their evaluation materials.

Findings

Surveys were completed by 42 high school principals and 141 teachers, representing 49 schools. Results are reported in the following areas:

- Representativeness of the Survey Respondents
- Background
- Knowledge
- Preparation Thus Far
- Future Plans
- Expectations
- Other

Representativeness of the Survey Respondents

As described in our original report (Wise, et al., 2000, pages 4–5), a representative sample of 24 districts was selected for intensive study over the course of the HSEE evaluation. Replacements were identified for each district (except for Los Angeles, which is irreplaceable) in case the targeted district could not participate. One to six high schools were selected from each original and replacement district, depending on district size, to create a representative sample of 84 schools. Where possible, replacements were identified for each selected school. In small districts with only one or two high schools, all schools were in the original sample. Sampling ratios were established so that each school would represent approximately the same number of 10th grade students. In this way simple averages across the schools in the sample would provide estimates for all 10th grade students in the state.

The Spring 2000 principal and teacher surveys were distributed to the 84 targeted schools. Three districts, including 8 of the targeted schools, declined to participate, but it was too late to contact the replacement districts for the Spring 2000 surveys. Principal surveys were returned from 42 schools, half of the original sample or 55% of the sample excluding the districts that declined to participate. A few of the schools that did not respond

declined to participate in the evaluation study and will be replaced in subsequent surveys. The remainder of the sample was simply unable to complete the surveys due to heavy staff demands at the end of the school year. One or more teacher surveys were received from 49 schools, including most of the schools participating in the principal survey and also additional schools that did not return principal surveys. In most cases, responses were received from two mathematics teachers and two language arts teachers.

We made several comparisons to determine how well the responding schools represented the original target sample and the state as a whole. Table 4.1 shows a comparison of the distribution of the responding schools on the key stratification variables used in selecting the sample. For the principal survey, slightly fewer schools with a high percentage of English language learners (ELL), high STAR 1999 mathematics schools, and schools from large districts responded in comparison to the target sample. For the teacher survey, fewer high ELL schools, but more high STAR 1999 math schools and more small district schools responded. For both surveys, the responding schools did include both high and low ELL schools, high and low STAR 1999 mathematics schools, and schools from large, medium, and small districts in proportions that matched the target sample reasonably well.

STAR 1999 data were used in the original selection of districts and schools. Recently, school means for the STAR 2000 examination have become available. Table 4.2 shows a comparison of the target and responding schools to statewide averages for the STAR 2000 10^{th} grade mathematics and reading scores. The average scores match to within one or two points. In addition, the standard deviations are quite similar. This indicates that the distribution of schools with average scores at specific levels above or below the overall average also matches. For the teacher respondents, the school averages are slightly more variable, evidence that a few more schools were significantly below and above the overall average in comparison to statewide distributions. This result may be related to the slight overrepresentation of schools from small districts where school averages would be expected to be more variable since they are based on fewer students.

Table 4.1
Comparison of Responding Schools to the Target Sample

School Statistics	Target Schools	Responding Schools	
		Principals	Teachers
Percent High % English Language Learners	57	52	48
Percent High Average STAR 1999 Math Score	43	40	52
Percent in Large Districts	52	48	48
Percent in Medium Districts	29	33	26
Percent in Small Districts	19	19	26
Number of Schools	84	42	49
Number of Survey Respondents		42	141

Table 4.2

Comparison of Survey Respondents to Statewide Averages: STAR 2000 Mean Scores and Standard Deviations

Population/Sample	Mathematics Average	SD	Reading Average	SD
Statewide[1]	698	16.0	691	16.7
Target School Sample[1]	700	16.8	693	17.0
Principal Respondents[2]	696	17.4	690	17.7
Teacher Respondents [2]	697	18.1	691	18.2

[1] School averages were weighted by the number of 10th grade students to estimate averages for all students in the state.

[2] School averages were weighted by the number of survey respondents (principals or teachers).

Tables 4.3 and 4.4 show comparisons of 10th grade averages and demographics from the STAR 1999 data. These data also show close correspondence between responding schools and statewide averages. In summary, the comparison data indicate that the schools from which survey responses were received are reasonably representative of the state as a whole. Based on sample size, the sampling error in estimates of statewide percentages is less than 8 points for the principal survey and less than 7 points for the teacher survey[4].

Table 4.3

Comparison of Survey Respondents to Statewide Averages: STAR 1999 Mean Scores and Standard Deviations

Population/Sample	Mathematics Average	SD	Reading Average	SD
Statewide[1]	697	16.3	690	16.6
Target School Sample[1]	698	15.2	692	16.1
Principal Respondents[2]	694	15.5	689	16.6
Teacher Respondents [2]	695	16.3	690	16.9

[1] School averages were weighted by the number of 10th grade students to estimate averages for all students in the state.

[2] School averages were weighted by the number of survey respondents (principals or teachers).

Table 4.4

Comparison of Respondents to Statewide Averages: Key 1999 10th Grade Demographics

Population/Sample	% Hispanic	% English Language Learners
Statewide[1]	39	16
Target School Sample[1]	43	18
Principal Respondents[2]	38	16
Teacher Respondents [2]	38	14

[1] School averages were weighted by the number of 10th grade students to estimate averages for all students in the state.

[2] School averages were weighted by the number of survey respondents (principals or teachers).

4 The sampling error for a proportion is given by the square root of p*(1-p)/N, where n is the sample size. The maximum sampling error occurs when p = .5 (50%) and is one half the square root of N.

Background

Principals were asked to provide demographic information on themselves. Over half of the respondents (57%) were male, 69% were White, 21% Hispanic, 5% African-American, and 5% declined to specify; 94% reported education beyond a bachelor's degree (7% some graduate school, 79% master's degrees, 10% doctoral degrees) and 5% 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 (21%). The respondents reported 1–31 years of experience as a principal (mean = 12.95, SD = 7.70) and 3–33 years teaching experience (mean = 13.46, SD = 8.19). They had 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 (59%) of the respondent teachers were female; 84% were White; 6% were Hispanic; 5% were Asian/Pacific Islander; 1% were Black; and 3% were other or declined to specify; 9% reported having only a bachelor's degree; most respondents reported education beyond a bachelor's degree (40% some graduate school, 44% master's degrees, 4% doctoral degrees); 4% 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." Seventy-seven percent indicated that their college training was in their primary subject area.

Principals were asked to provide background information on their schools. The current number of teachers on staff ranged from 3 to 200, with a mean of 75 (SD = 52). Principals reported that the percentage of teachers with advanced degrees ranged from 18% to 80%. Counselor-student ratios ranged from 1:1 to 1:1000, with a median of 400:1. Forty-eight percent of the responding schools currently have a testing coordinator; an additional 5% reported plans to have one by September 2000. Most schools (79%) operate on a semester basis; 12% configure their school year in quarters and 5% operate year-round schools. The majority of principals (67%) reported that their schools hold 6–7 academic periods per day. They reported, on average, a graduation rate of 80%, with rates varying by racial/ethnic group. Post-graduation attendance in 2-year colleges averaged 29% and 4-year colleges, 28%.

Principals were asked to indicate whether their schools offered various specialty education programs. Sixty-two percent offer remedial courses; 26%, magnet programs; 74%, special education; 52%, English-language learners; 19%, multicultural/diversity-based; 48%, Advanced Placement; 2%, International Baccalaureate; 43%, school/community/ business partnerships; 31%, targeted tutoring; and 12%, other.

Teachers were asked to provide some information about their own classes. Asked to provide average enrollment per class period, they reported 1–40 students, with a mean of 26 (SD =6.5). Seventy-nine percent report that they create groups within classes for instruction. Of these, 53% assign students to these groups randomly; 8% use ability grouping; 6% allow students to choose their groups; and 14% indicated that they assign students to groups on some other basis. Twenty-four percent of teachers reported that 100% of their students were fluent English speakers; 45% indicated that 90–99% were fluent in English; 21% reported

75–89%; 7% reported 50–74%; and 1% indicated that less than 50% of their students were fluent in English.

Teachers were asked about various instructional practices. Forty percent of teachers require students to maintain a portfolio; an additional 11% indicated that they require another product in lieu of the portfolio. Three-quarters of teachers (78%) estimated that students spend ½ hour or more of class time each week working with a partner or in a small group.

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 (51%) estimated ½ to 3 hours; 19% estimated more than 3 hours; 20%, less than ½ hour; and 8%, none.

Teachers were asked to indicate the importance of specific instructional techniques. Techniques frequently endorsed as "very important" were: using questioning techniques to promote interaction and discussion (79%), developing students' abilities to make connections among content topics (76%), using problem-solving as a means and a goal (76%), 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 (87%), do work from supplemental materials (77%), apply subject area knowledge to real-world situations (72%), write a few sentences (70%), and work in pairs or small groups (72%).

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. Twenty-two percent reported that they were very familiar with the exam, while 2% expressed no familiarity. Teachers reported less familiarity with the exam than the principals: 11% claimed to be very familiar, 66% generally familiar, and 22% 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: 4% very familiar, 62% generally familiar, and 31% 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. Sixty-seven percent of principals said they were very familiar with the state content standards and 31% reported general familiarity. Teachers reported more familiarity with state content standards than did principals: 65% very familiar, 29% 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 that of other teachers whom they rated: 36% very familiar, 48% generally familiar, and 4% not at all familiar.

One possible source of information on the HSEE and state content standards for teachers was the Item Rating Workshops conducted as part of our evaluation. We asked teachers whether they had participated in either of the May 2000 workshops; only 8% 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: their district (93%), the state (76%), newspaper (60%), professional associations (52%), education organizations (33%), computer-based sources (29%), and other (5%). Two percent of principals indicated that they had no sources of information on the HSEE. Teachers reported that their information came from: school-provided information (57%), district-provided information (40%), newspaper (33%), state-provided information (21%), education organizations (15%), professional associations (13%), computer-based sources (9%), and other (11%). The other sources of information included the workshops in May and conversations with other staff. Nine 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 were with the exit exam. Responses indicated a belief that the exit exam was virtually unknown outside the educational community. Two percent of principals responded that students/parents were very familiar or familiar with HSEE. Twelve percent of principals estimated that students/parents were somewhat familiar; 48% not very familiar; and 38% replied that students/parents were not at all familiar.

Preparation Thus Far

Although the HSEE will not be administered operationally until March 2001, 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 students are being taught what will be tested. Thus respondents were queried about alignment with state content standards. In short, most principals indicated that they are already moving in the direction of alignment, but still have a way to go. All principals (100%) reported that their districts/schools encourage use of the content standards to organize instruction, and 81% 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-six percent stated that their textbooks do not align well with the content standards; 38% 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 (69%) responded that their districts have adopted the state standards, and another 19% reported that their district standards include more than the state content standards. Thus, a total of 88% indicated that their district standards encompass all state standards. However, 7% reported that the state standards include more than the district standards, and 5% indicated that they could not judge. No respondents indicated that the two sets of standards were different or 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 (50%) or 16–35 hours (24%). Nineteen percent reported fewer than 6 hours; 7%, more than 35 hours. Most teachers reported fewer hours than principals: 23% none, 57% fewer than 6 hours, 11% 6–15 hours, 3% 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: 25% none, 39% fewer than 6 hours, 18% 6–15 hours, 6% 16–35 hours, and 9% more than 35 hours.

Respondents were asked to identify the specific activities they have undertaken to prepare students for the first administration of the HSEE. Although the students who will participate in the HSEE had not yet entered the ninth grade, most principals reported initiating some activities; only 17% indicated that they have implemented none. Figure 4.1a indicates the percentage of principals who reported implementing each activity, in descending order of endorsement; Figure 4.1b indicates teachers' responses, in the same order as Figure 4.1a to facilitate comparison. In general, fewer activities were reported by teachers; 36% indicated that none had taken place. This may mean principals were aware of some individual teachers implementing activities even though implementation was not school-wide.

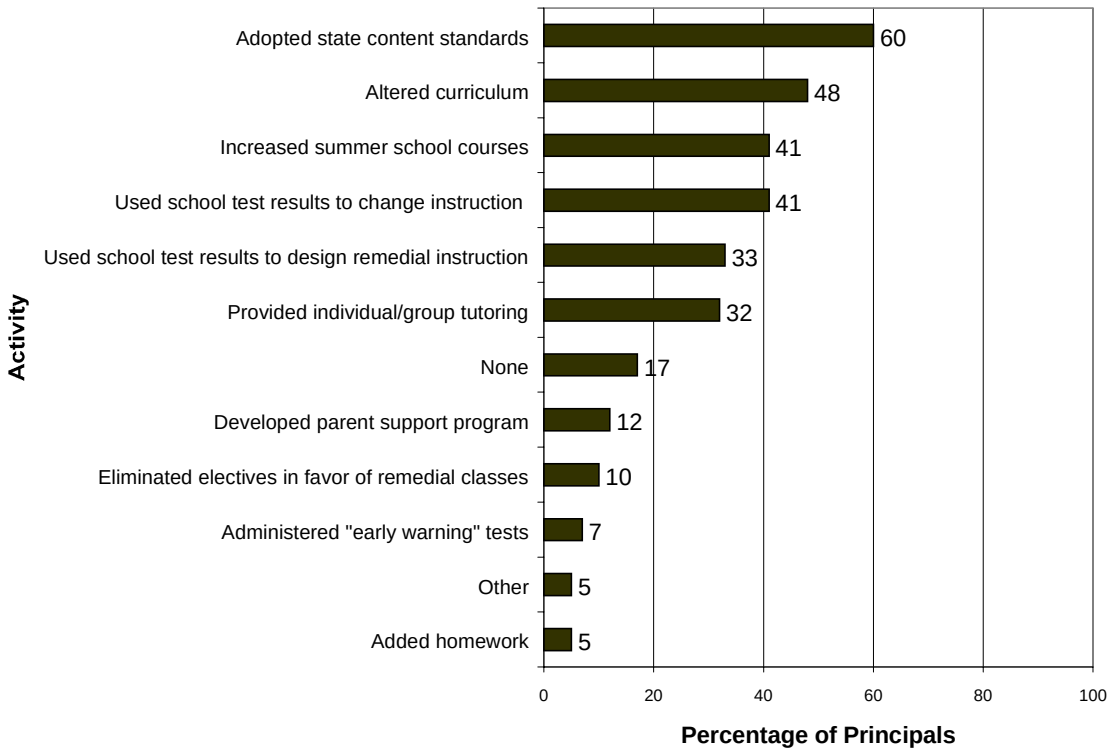

Figure 4.1a. Percentage of principals reporting activities already underway to prepare students for the HSEE.

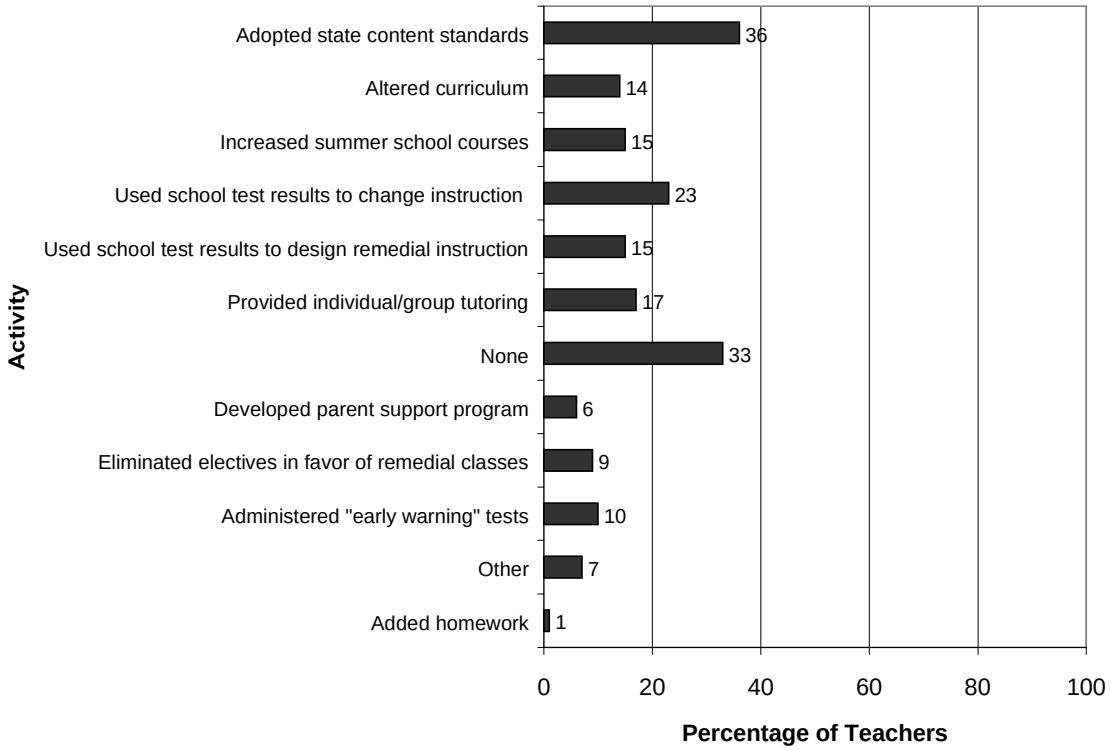

Figure 4.1b. Percentage of teachers reporting activities already underway to prepare students for the HSEE.

Teachers were asked to provide, in their own words, a list of "any specific changes 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." Of the 69 open-ended responses, nearly 30% indicated that they had made no changes to accommodate the HSEE. Another 19% reported that they already teach to the State Content Standards; 19% specified that they were focusing on higher-level subject content and 14% are focusing on test-taking techniques in the hope that these techniques will apply to the HSEE. Fewer than 10% reported that they were focusing more on basic math skills, participating in the task force to modify curriculum, or other responses.

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 4.2a lists the percentage of principals who endorsed each activity (in descending order of endorsement) and Figure 4.2b reflects teacher responses to the same question (in the same order as the principal graph). Similar to the pattern of preparatory steps, more principals reported activities than did teachers. For example, only 14% of

principals indicated that no plans had been made for remediation, compared to 38% of teachers.

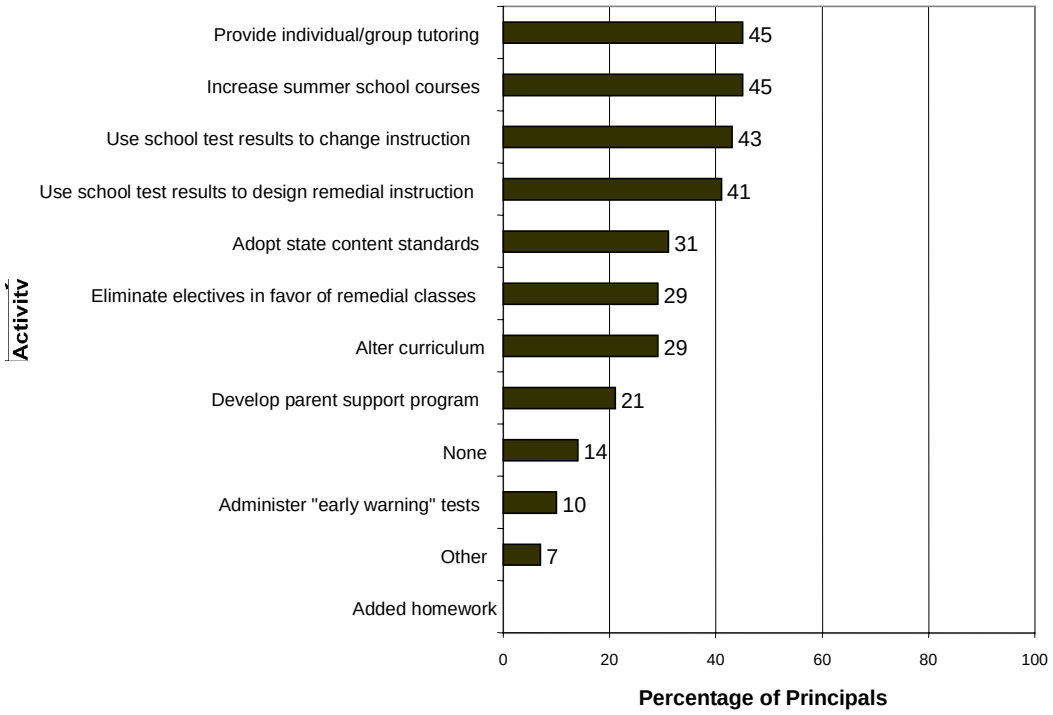

Figure 4.2a. Percentage of principals reporting plans for remediation of students who do not pass the HSEE.

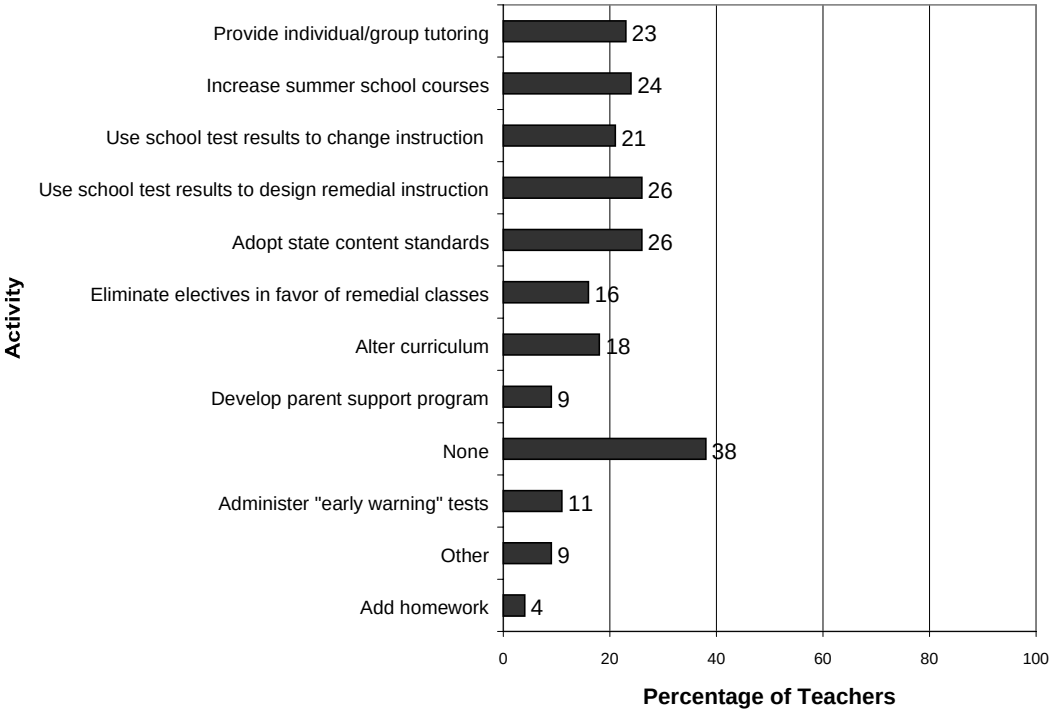

Figure 4.2b. Percentage of teachers reporting plans for remediation of students who do not pass the HSEE.

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, in-service 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 4-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.

Along similar lines, teachers were asked to indicate, in their own words, responsibilities they believed they were likely to be assigned to get students through successful completion of the exam. Eighty-one teachers responded. The most common response, by far, was that they expected to be called upon to provide tutoring, remediation, added instruction, or teach summer school (58%). Fewer respondents indicated that they would be asked to revise curriculum (15%) or identify students' strengths and weaknesses or place students in appropriate classes (11%). Fewer than 10% reported that they would be assigned to work on test-talking strategies or give sample tests, work with parents, or have general higher responsibilities

Over half of the principals' 21 open-ended responses 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 to the law. One-fourth of the principals 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.

Teachers were asked to specify any specific curricular or instructional changes they planned to make in the future. Seventy-six teachers provided responses to this open-ended question. Responses varied widely: 24% plan to focus on higher-level subject content; 22% reported that they don't know or are not familiar with the test content yet; 14% plan to modify their course content according to what is and is not tested; 11% plan to implement to State Content Standards. Fewer than 10% indicated that they plan to focus on more basic math skills, practice more test-taking techniques, select new textbooks, or depend on district changes or mandates. Another 11% provided other responses that could not be readily categorized.

Expectations

Several survey questions 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 10th grade students (Class of 2002) who would earn a passing grade on the upcoming exam.[5] As Figure 4.3 indicates, responses were generally guarded. Half (50%) of principals predicted that fewer than 50% of students would pass the exam; 29% predicted 50–74% of students would pass; 14% predicted 75–95%; and 5% of principals predicted that more than 95% of students would pass.

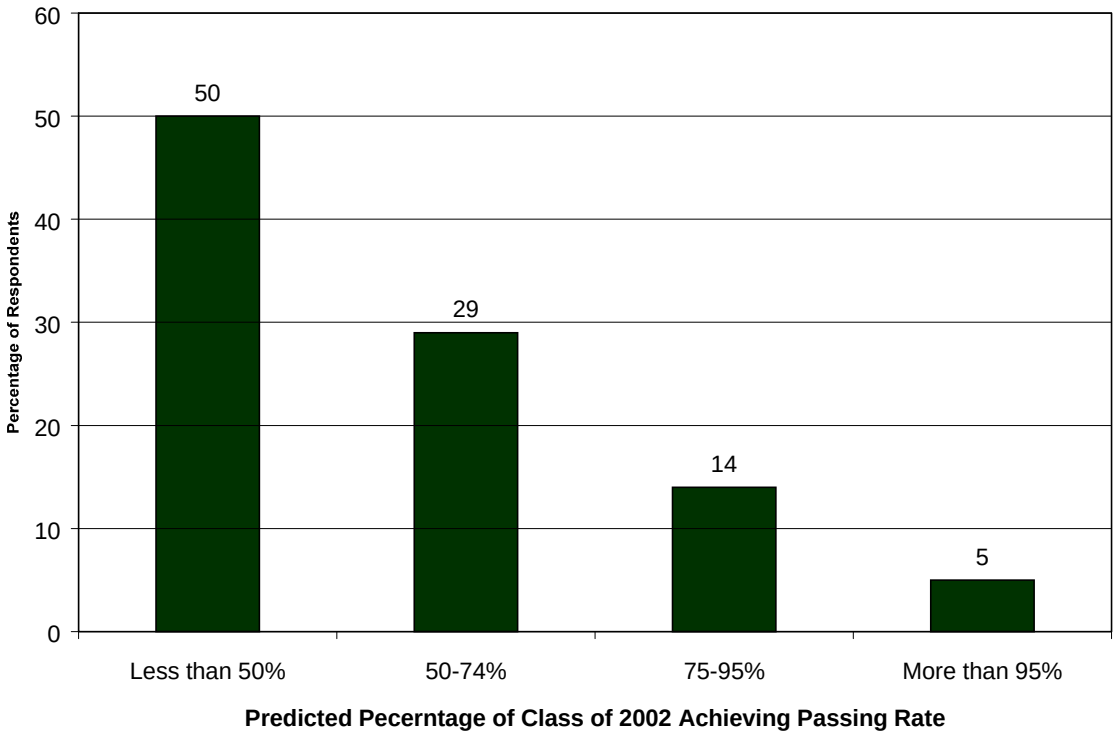

Figure 4.3. Principals' predictions of pass rates if the Class of 2002 were to take the exam.

Teachers were asked two variants of the same question. They were asked to estimate the preparedness of students to pass the HSEE in the 9th grade and in the 10th grade, based upon the teacher's knowledge of the feeder schools. As Figure 4.4 indicates, nineteen percent of teachers responded that students were prepared (or better) in the 9th grade; 40% indicated that students were prepared or better in the 10th grade. Although the structure of the questions asked of principals and teachers differed, the responses were similar.

[5] 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 9th grade. Because the first participating group is not yet in high school, principals were asked to assess current 10th graders (Class of 2002) as a proxy for the Class of 2004.

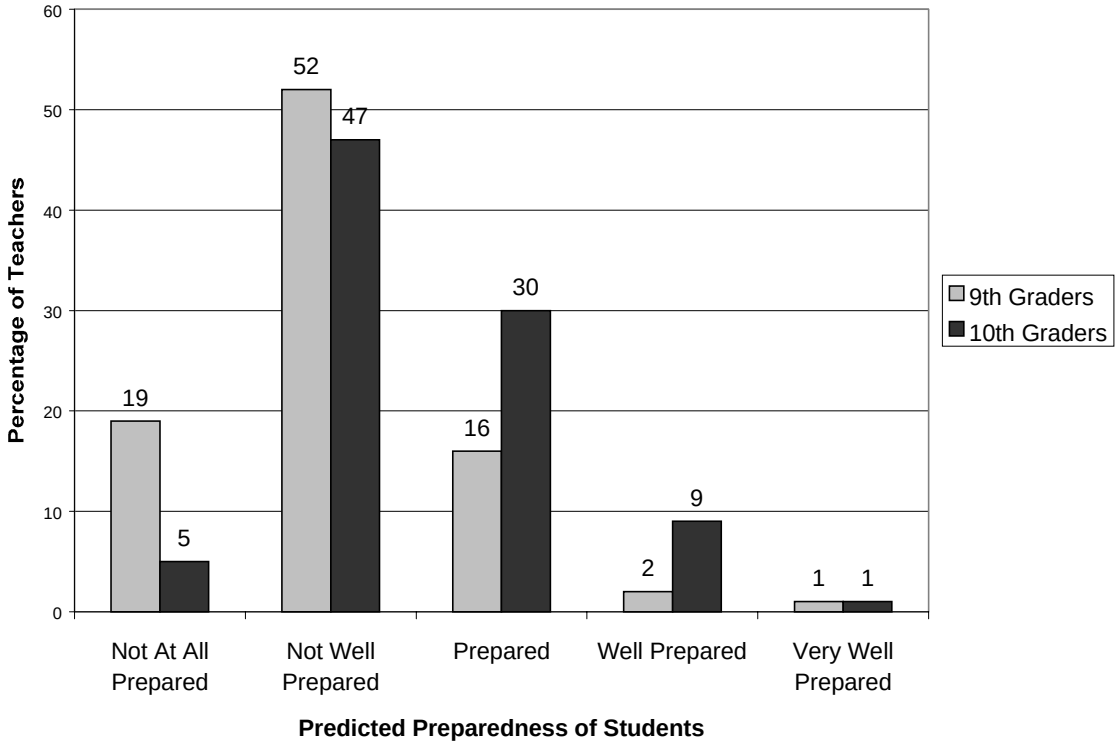

Figure 4.4. Teacher's estimates of preparedness of students to pass the HSEE in the 9th and 10th grades.

Principals and teachers were also asked to predict the impact of the HSEE on student motivation and parental involvement, under various circumstances. Figures 4.5a and 4.5b 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 4.5a and 4.5b indicate that teachers are somewhat more pessimistic than principals about the impact of the HSEE on student motivation and parental involvement.

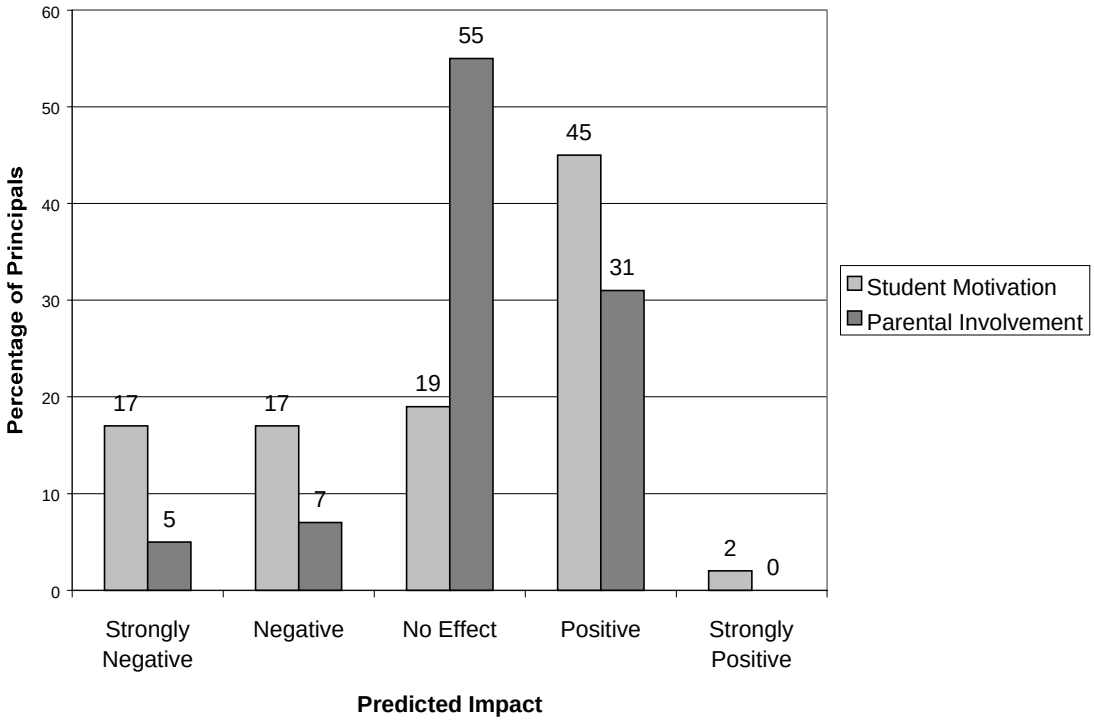

Figure 4.5a. Principals' predicted impact of the HSEE on student motivation and parental involvement of students who pass the exam on the first attempt

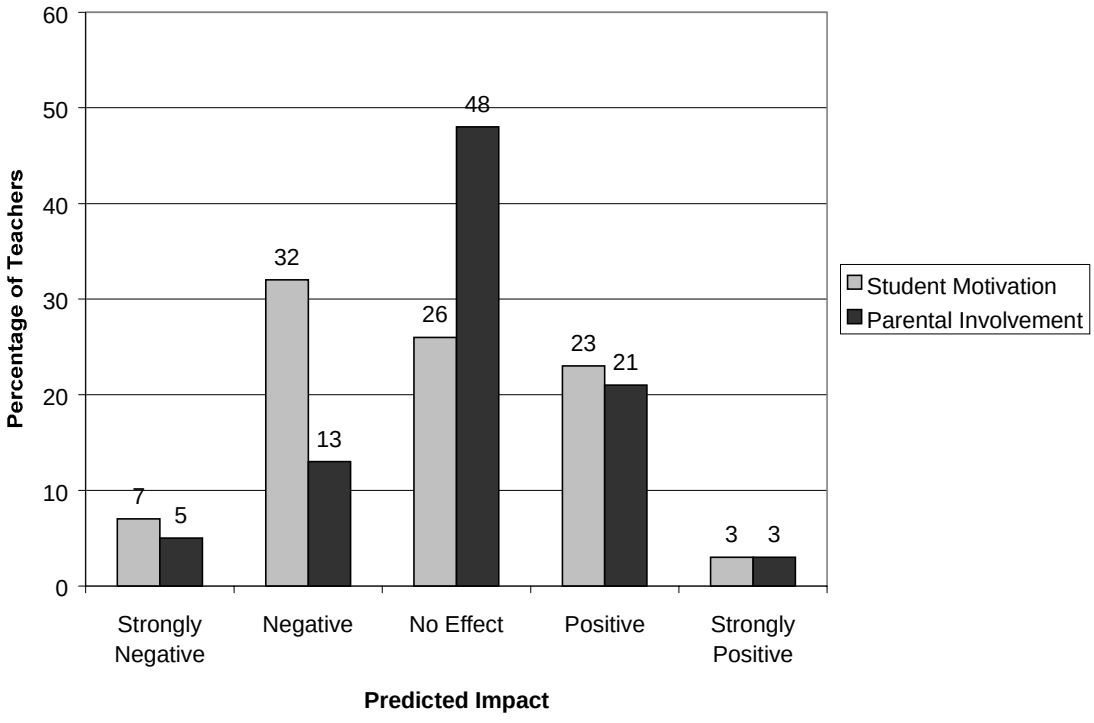

Figure 4.5b. Teachers' predicted impact of the HSEE on student motivation and parental involvement of students who pass the exam on the first attempt

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. As Figure 4.6a depicts, only 5% of principals expected that student motivation would drop after students cleared the hurdle of the HSEE. Thirty-three percent of principals predicted that student motivation would be unaffected by passing the exam; 62% predicted a positive or strongly positive effect. Half of principals expected no impact on parental involvement; 33% predicted a positive effect, 12% a strongly positive impact, and 2% a strongly negative impact on parental involvement for those students who pass the exam early in their high school careers.

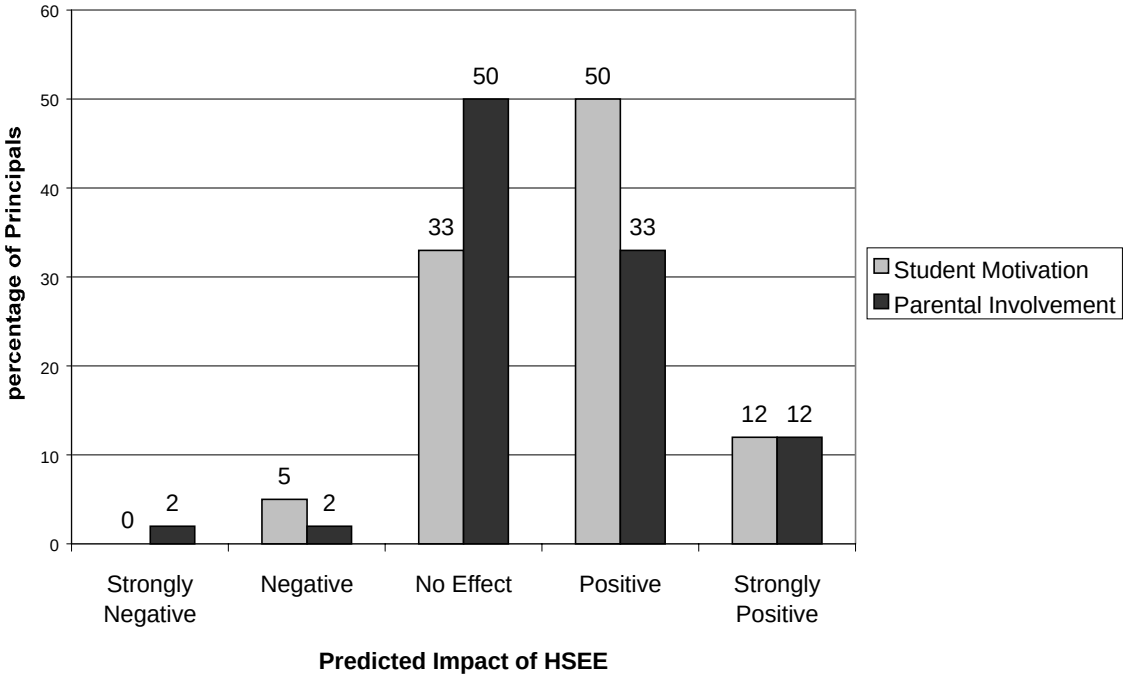

Figure 4.6a. Principals' predicted impact of the HSEE on student motivation and parental involvement of students who pass the exam on the first attempt

Here again, principals were more optimistic than teachers. Figure 4.6b indicates that fourteen percent of teachers expected a negative or strongly negative impact on student motivation after passing the exam on the first attempt. Thirty-eight percent of teachers predicted that student motivation would be unaffected by passing the exam; 39% predicted a positive or strongly positive effect. Half of teachers (49%) expected no impact on parental involvement; 8% expected a negative or strongly negative effect; 29% predicted a positive effect and 6% a strongly positive impact on parental involvement for those students who pass the exam early in their high school careers. Nine percent of teachers declined to estimate the impact of passing the test on student motivation or parental involvement.

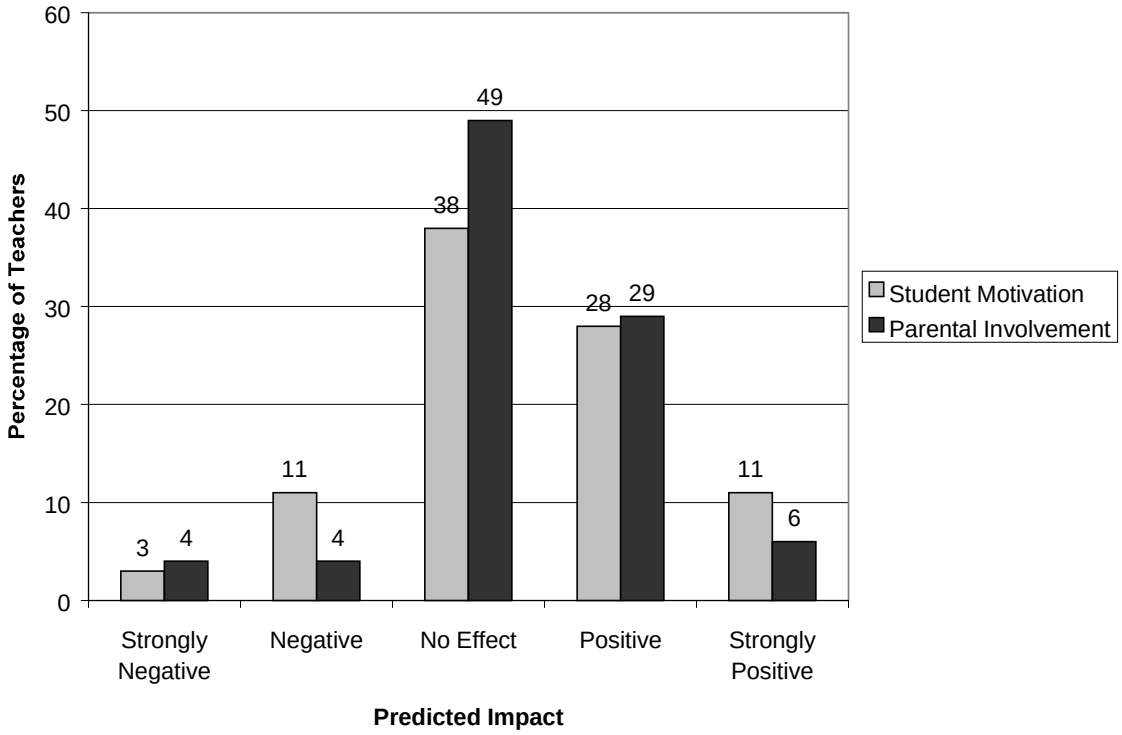

Figure 4.6b. Teachers' predicted impact of the HSEE on student motivation and parental involvement of students who pass the exam on the first attempt

For those students who fail the exam on the first try, the principals' and teachers' predictions were quite different from pre-examination predictions. Figures 4.7a and 4.7b illustrate response patterns for principals and teachers, respectively. Principals were split on whether the impact of failing the exam would have a negative effect on student motivation; 10% predicted a strongly negative effect; 36%, negative; 17%, no effect, 33%, positive, and 2% strongly positive. Predictions for parental involvement were very similar to those of student motivation: 7% predicted a strongly negative effect; 36%, negative; 14%, no effect; 40%, positive; and 2%, strongly positive. There was a similar pattern for teacher responses, albeit slightly more negative overall: regarding student motivation, 7% predicted a strongly negative effect; 30%, negative; 16%, no effect, 33%, positive; and 4%, strongly positive. As for parental involvement, 6% of teachers predicted a strongly negative effect; 21%, negative; 28%, no effect; 32%, positive; and 2%, strongly positive.

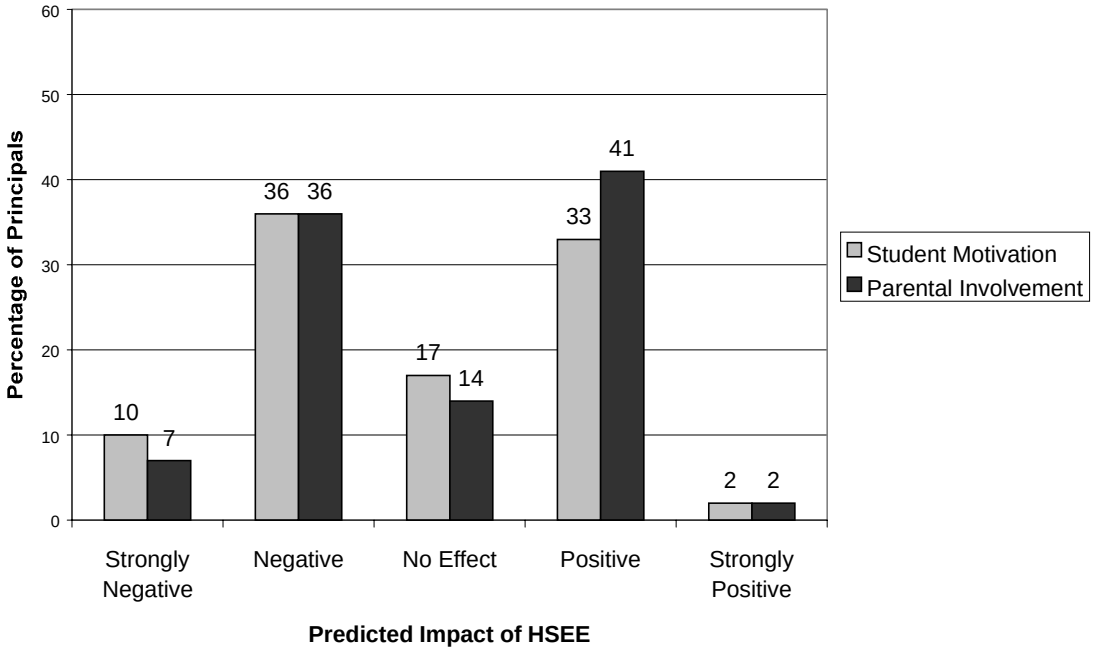

Figure 4.7a. Principals' predicted impact of the HSEE on student motivation and parental involvement of students who fail the exam on the first attempt.

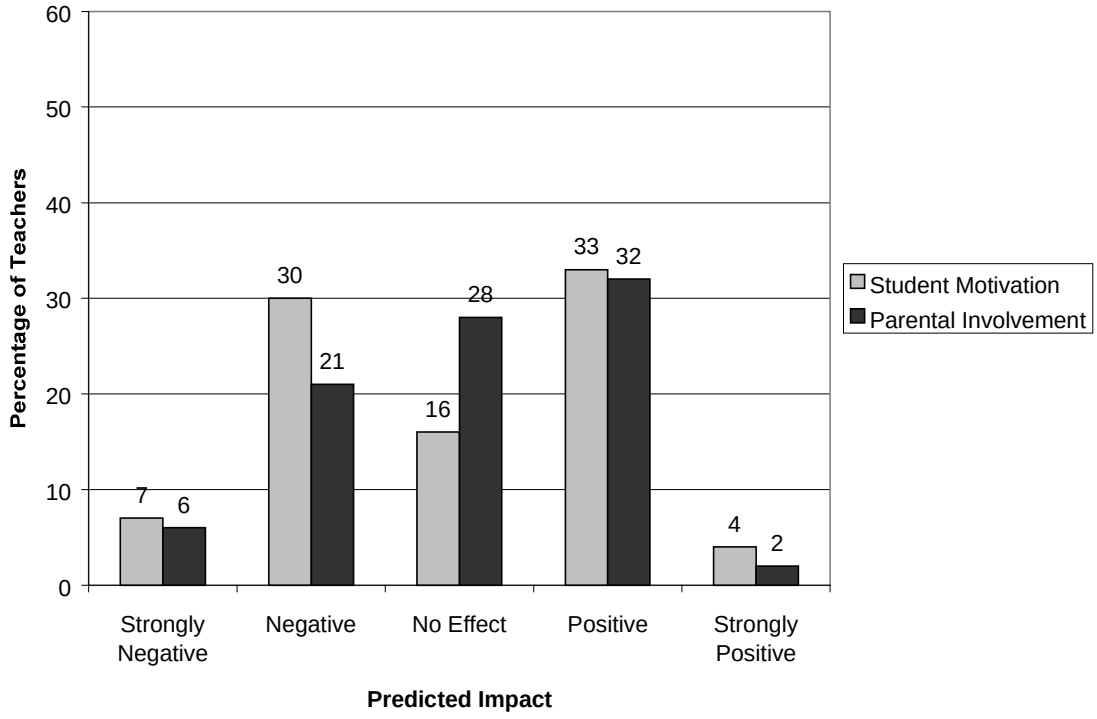

Figure 4.7b. Teachers' predicted impact of the HSEE on student motivation and parental involvement of students who fail the exam on the first attempt.

Principals and teachers were also asked to predict the impact of the HSEE on student retention and dropout rates. Responses were somewhat negative overall. Figures 4.8a and 4.8b reveal that predictions followed a similar pattern on both questions. Fifty-five percent of principals anticipated a strongly negative or negative impact on student retention rates; 64% predicted a strongly negative or negative impact on student dropout rates. Twenty-nine percent predicted no effect on student retention and 21% predicted no effect on student dropouts. Seventeen percent anticipated a positive or strongly positive effect on student retention rate and 14% 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.

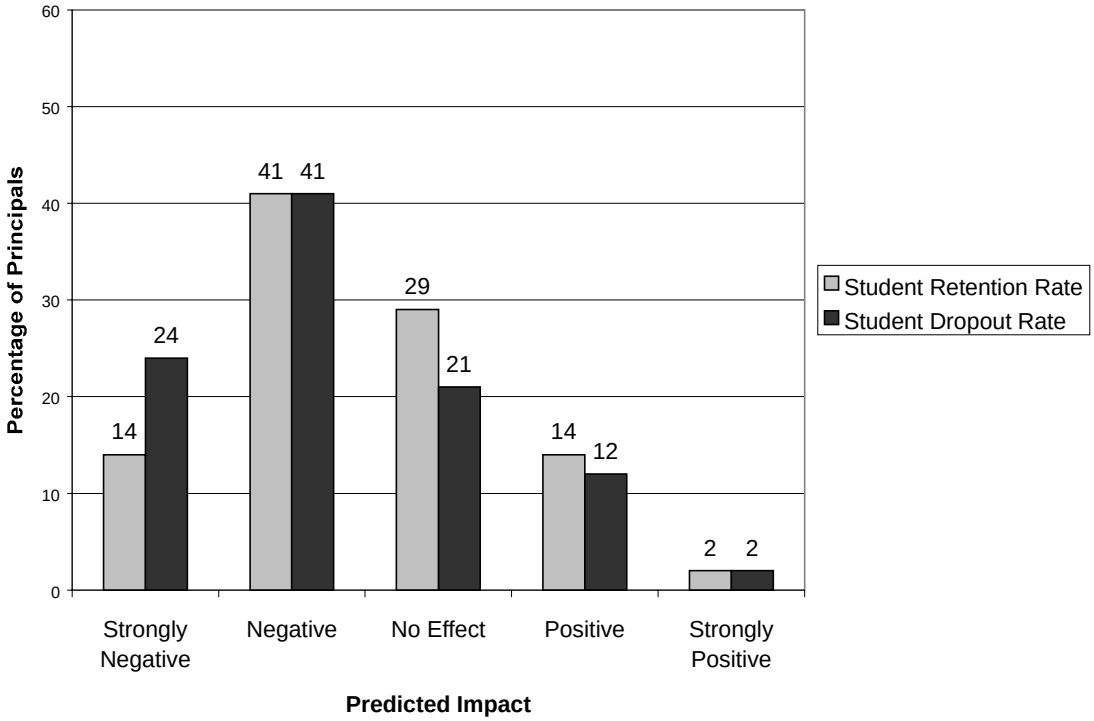

Figure 4.8a. Principals' predicted impact of the HSEE on student retention and dropout rates.

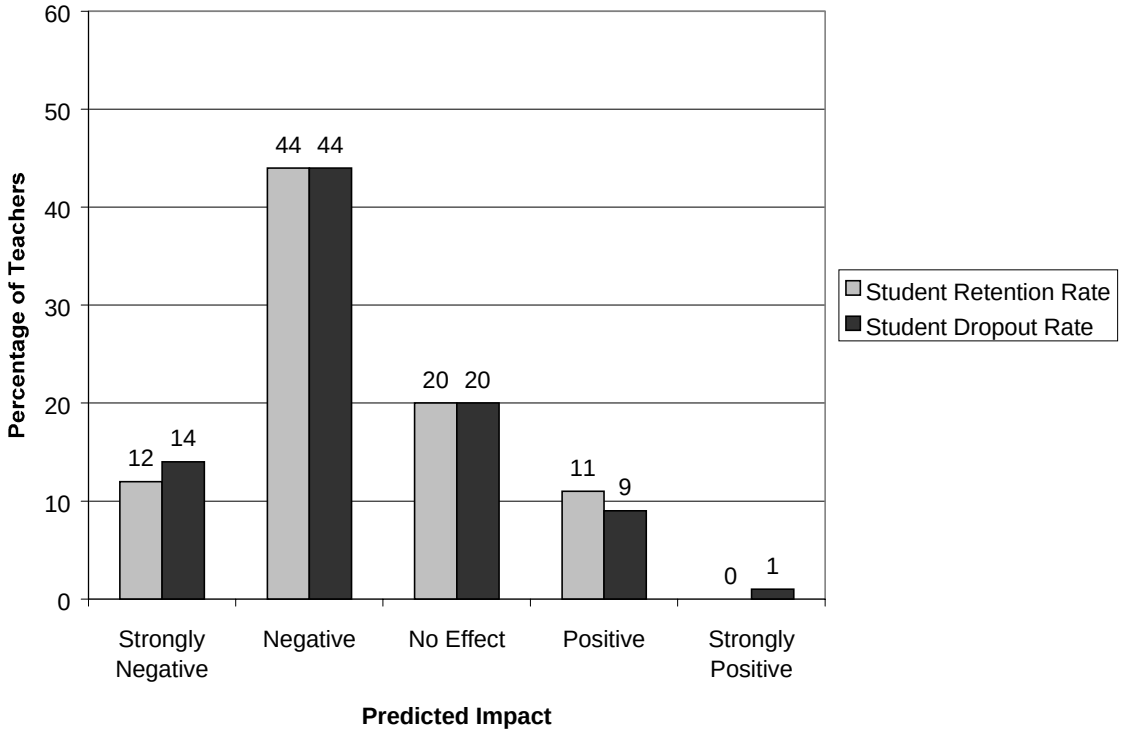

Figure 4.8b. Teachers' predicted impact of the HSEE on student retention and dropout rates.

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: 74% responded that practices would be improved, 10% predicted no effect, and 2% said extremely weakened. No respondents chose the options of strongly improved or weakened and 14% declined to respond.

Teachers were asked the same question about the influence of the HSEE on instructional practices, but they were asked to provide separate estimates for 3 school years. Figure 4.9 provides the responses for all 3 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.

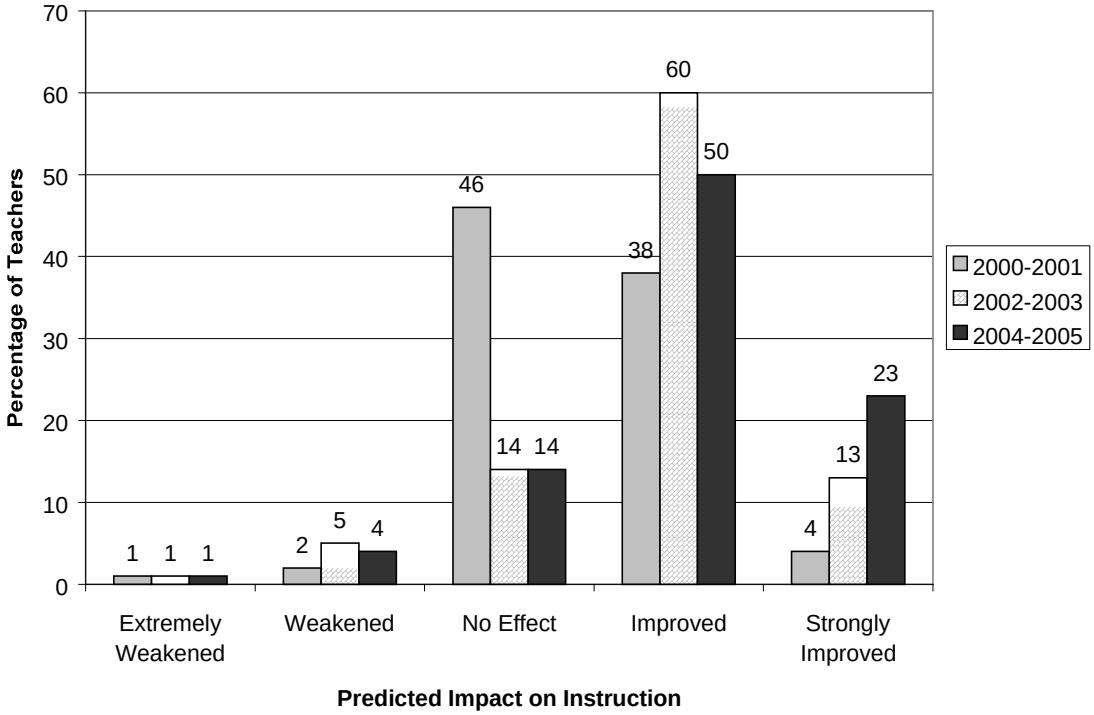

Figure 4.9. Teachers' prediction of influence of the HSEE on instructional practices over time.

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. Five percent of principals indicated that they were "not sure of the effect on the total student population;" 17% reported an excellent opportunity to learn; 26% selected good; 31%, adequate; and 19%, 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 19% for all students, to 31% for students with disabilities, 36% for English-language learners, 29% for English-language learners in targeted subjects areas, and 24% 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. Figure 4.10a depicts principal and teacher responses regarding the full student population: 5% of principals expressed that they were unsure; 12%, excellent; 26%, good; 40%, adequate; and 10% poor. No principals selected a response of "none." Teachers provided similar responses, with the bulk of predictions falling in the "adequate" and "good" categories.

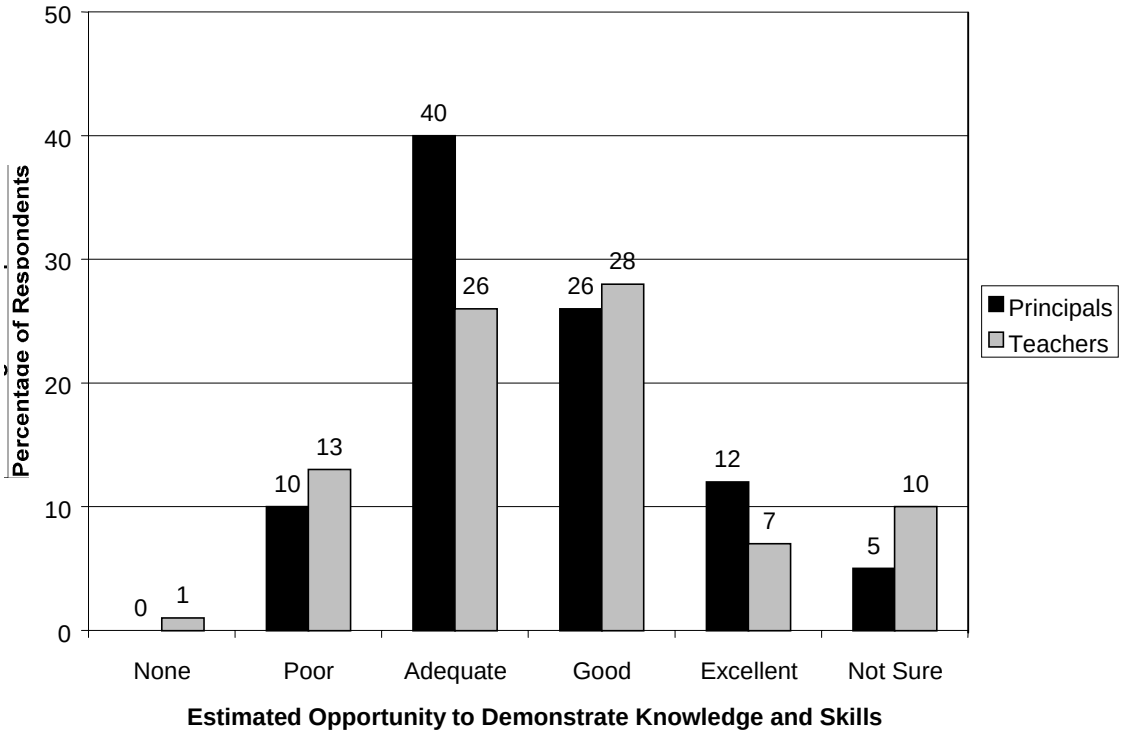

Figure 4.10a. Principal and teacher estimates of the opportunity for all students to demonstrate their knowledge and skills on the exam.

For the various student subgroups, responses were less optimistic; a none-to-poor opportunity to demonstrate knowledge and skills was anticipated for students with disabilities by 29% of principals; English-language learners, 29%; English-language learners in targeted subject areas, 24%; and economically disadvantaged students, 19%. Teachers provided similar responses, although teachers estimated the proportion of each group having none/poor opportunity as about 2–5 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 none/poor opportunity. Figure 4-10b compares the principal and teacher responses of "none" or "poor" opportunity for each of these student subgroups.

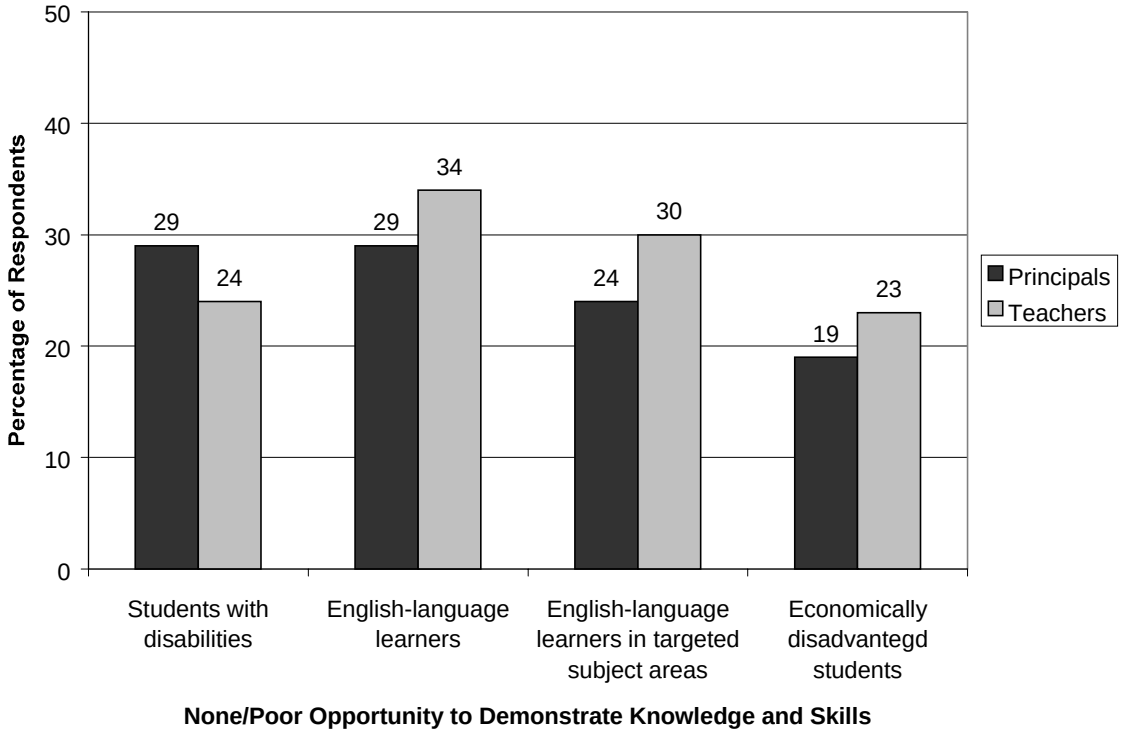

Figure 4.10b. Principal and teacher estimates of none/poor opportunity for various student subgroups to demonstrate their knowledge and skills on the exam.

Challenges

When asked to "describe challenges facing the school and students in successfully meeting the exam requirements," 30% of the 30 open-ended responses from principals and 38% of the 80 open-ended responses from teachers commented on the low levels of student competency and skills of present incoming high schoolers—especially for Continuation and Community Day schools. Also 30% of the principals described alignment issues, and 13% referred to the difficulties of meeting algebra and English/language arts proficiencies—especially for English-language learners. Fourteen percent of teacher comments indicated that low attendance or some aspect of home life (e.g., lack of parental involvement, unstable home lives, transience) as important challenges; 11% of teacher comments reflected lack of student motivation. Of the 20% who cited time requirements and the burden of testing, two comments particularly captured this challenge and underscored the lack of knowledge about the purpose of the test:

> "We test too much behavior Stanford 9, 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 9th graders. We will visit the homes of the incoming 9th 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 principals who commented 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. Teachers, on the other hand, emphasized creating standards, defining expectations, or improving the curriculum (30% of 50 comments), motivating students and improving their performance (26%), accountability (16%), and causing graduating students to be academically prepared (12%).

Other

Principals were asked to add any comments about specific factors at their schools that they felt would influence the exit examination. Of 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 questions. One of these stated that the questions 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 would not be a true reflection of their predominantly Caucasian school.

Similarly, teachers were asked to add any comments about "factors specific to you, your classes, or your school that are influencing the exit examination." Thirty-six percent of the 44 open-ended comments indicated that their students are "at risk" because of home lives, language barriers, low socioeconomic status, or under-achievement at other high schools. Another 27% commented on the test itself, which they haven't yet seen and thus preparedness is affected; some of these respondents questioned the validity of the test for predicting future success or thought that the test would be either difficult or unchallenging for their students. Sixteen percent of teachers felt that the highly-involved community and parents would be helpful. Fewer than 10% specified that examinations were a bad idea or should be taken outside school hours; will raise standards; will change their approach to teaching; will reduce their control over the curriculum; will depend on pre-high school preparedness; the dropout rate would increase; or expressed appreciation for their small class sizes.

Summary

Not surprisingly, 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 were more knowledgeable about the HSEE than the typical teacher, a possibility that 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 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 preparing staff through such efforts as planning curriculum workshops and inservice training. Most principals also reported initiating some type of activity to prepare students for the first administration of the HSEE; efforts included 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. Both principals and teachers expressed some concern 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 student 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 would be positive. Further, teachers were asked to estimate effects next year and in 3 and 5 years; they predicted greater improvement with time.

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

In short, the principals and teacher survey responses 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 5: CONCLUSIONS AND RECOMMENDATIONS

Updated Conclusions

In our June 30 report, we concluded that a great deal of progress had been made in defining and developing the HSEE and that results to date were quite positive. We also pointed out that much needed to be done before the HSEE could be administered operationally and that there was good reason to be concerned that many students are not well prepared to succeed on this examination. Nothing in our supplemental analyses leads us to change these conclusions.

As was the case for the multiple-choice questions, the quality of the essay questions was found to be high. The majority of the essay questions for reading passed the statistical screens, as did all of the writing scores. The essay questions appear to function in the same way for both males and females and for Hispanic and white students. Field test sample sizes were too small, however, to test for equivalent functioning in other groups.

The average scores for the essay questions were consistent across prompts, but generally low. Few students received maximum credit for their responses to these questions. The finding that students did not do well on these questions was no worse than the low levels of performance on the multiple-choice questions. We note, however, that while almost all students attempted to answer all of the multiple choice questions, a significant number of students (at least 10%) did not respond to both essay prompts. Some students may not have been highly motivated to do well on these questions. Scores may be higher when test results count, but it is not possible to estimate how much difference this would make. Even among students who did respond fully, fewer than half received 3 or 4 of the 4 possible points.

In Chapter 3, we presented passing rates and curriculum-alignment information for each of the proposed content standards. In mathematics, both passing rates and curriculum-alignment ratings were low for the Algebra 1 standards. This should not be surprising since students are not now required to take algebra in many districts and algebra is not mandated in state legislation. In English/language arts, passing rates and curriculum-alignment ratings were generally lower for standards involving higher levels of integration and analysis. There were few, if any, content standards for either test on which the majority of students performed well.

The results from the Teacher and Principal Surveys are largely the same as those presented in our June 30 report. There was considerable variation across the 40 to 50 participating schools in the nature and extent of current preparation for the HSEE and in expectations for its impact on the schools and their students.

In the next two sections of this chapter, we offer suggestions and recommendations concerning the large amount of work remaining to be done in developing the HSEE. We also offer suggestions for responding to concerns that students are currently not well prepared to pass the HSEE. These sections are followed by more specific recommendations on technical issues based on our review of the test developer's plans and proposals and then further discussion of the specific recommendations in our June 30 report.

Action Plans for a Tight Timeline

The overarching recommendation in our June 30 report was that the Legislature, Governor, Board, and Department consider delaying full implementation of this new graduation requirement. Since the timeline for implementation has been written into the California Education Code, the primary audience for this recommendation was the Legislature. Superintendent Eastin and the Board have pointed out, quite correctly, that they have no choice but to follow the law in this matter. So far, there is no indication that the Legislature will consider changes prior to initial implementation in 2001.

We continue to be concerned about the timeline for implementing this graduation requirement for two reasons. The first concern is that, on the proposed schedule, the net outcome for students may well be negative. Based on our assessment of current student achievement relative to the standards likely to be adopted, it seems probable that a significant number of students in the first few high school classes subject to this requirement will not be adequately prepared. Without adequate preparation, the number of students who are denied diplomas will increase sharply as will the number of students who lose hope and drop out. The consequences for these students will necessarily be quite negative.

The Board, the HSEE Panel and others expressed concerns that, because of our initial recommendation, schools might relax efforts to implement programs to teach the state content standards. Schools please note: It is not likely that this program will be delayed and you have a lot to do. Please do not interpret either of our reports as a reason to back off from full and rapid implementation of the state content standards.

If the concern is with what schools should do, we suggest it would be more reasonable to begin by establishing a system of consequences for schools and give that system some time to work before enforcing consequences for students. The Academic Performance Index (API) is designed as part of just such a school-consequence system. Schools receive rewards or sanctions based on student performance as measured by the API. To date, however, the API only includes Stanford 9 scores. The API does not yet include measures of student achievement relative to the state content standards. It would be reasonable to move to include standards-based measures in the API as soon as possible and hold schools accountable for student performance on these measures before implementing important student consequences.

Our second concern is that, because timelines are so tight, implementation without a full dress rehearsal may well be flawed and the whole program then rejected before it can achieve its desired effects. If initial passing rates are quite low, there is likely to be a significant political backlash against the program. In particular, the experience in other states suggests a reasonable probability that the fate of the HSEE will be decided by the courts. Chapter 2 of our June 30 report identified several legal concerns, the most critical of which is the need to demonstrate that all students are provided adequate instruction in the material covered by the exam. (Note that “adequate” is likely to mean adequate quality as well as adequate content.) It seems risky, at best, to implement the HSEE requirement without first assembling evidence that schools are providing adequate instruction on the content of the test. Under California law, suits are likely to be directed at schools and districts as well as at the state. Individual

districts need to be given time and assistance to assemble evidence supporting the adequacy of their standards-based instruction.

Based on these concerns, our first recommendation continues to be:

Recommendation 1. The Legislature and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.

We recognize that the primary task for the independent evaluation is to report on consequences of the HSEE, not necessarily to try to predict them in advance. Balancing risks associated with implementing a program too quickly with risks associated with implementing too slowly requires policy decisions. Such decisions are appropriately made by the Legislature, Governor, Board, and Superintendent who are charged with responsibility for setting such policy. The present implementation schedule for HSEE reflects the current judgment of these groups.

An important role for the independent evaluation that is clearly within the scope of our charge is to make recommendations for improving the development and implementation of the HSEE. In the remainder of this section, we assume that the current schedule will be followed and offer suggestions for maximizing the quality and minimizing the risks of the HSEE program under the constraints implied by this schedule. Most of these suggestions would also be appropriate should a delay be approved.

The co-chairs of the HSEE Panel suggested that it would be helpful to follow the discussion of issues experienced by other states (in Chapter 2 of our June 30 report) with a suggested agenda or plan of action for addressing these issues. In discussing our recommendation to postpone implementation, we did list many of the critical steps that need to be completed. We did not, however, suggest specific responsibilities and timelines for these steps and our list was more suggestive than comprehensive. As evaluators, there are limits to the extent to which we believe we should be responsible for creating, as well as evaluating, HSEE development plans. Nonetheless, we are in full agreement that a more detailed and public plan of action for addressing issues in implementing the HSEE is needed. We therefore offer the following recommendation:

Recommendation 2. CDE should develop and seek comment on a more detailed timeline for HSEE implementation activities. This timeline should show responsibility for each required task and responsibility for oversight of the performance of each task. The plan should show key points at which decisions by the Board or others are required along with separate paths for alternative decisions that may be made at each of these points.

Since the June 30 report, CDE has made considerable progress in planning the remaining steps that must be completed prior to implementation of the HSEE. They have received and reviewed AIR's timeline for the fall field test and the development of initial forms, set a date for convening a panel to recommend a minimum passing score, and worked with the HSEE Panel in their development of recommendations for accommodations for special needs students and English language learners. Plans are in place for disseminating information about HSEE to districts and schools immediately following a Board decision on adoption. There is not, however, a public document that lists all of the remaining steps with a timeline

for their completion. The development contractor's proposal covers many, but not all of these steps and there have been numerous updates to the timelines for the activities that are covered. We believe that wide review of plans and timeline for remaining development activities will help to identify potential issues and omissions in these plans while there is still time to address them.

Under current legislation, the Board must adopt an exit exam at its September 2000 meeting. The Board must subsequently review and approve implementing regulations and recommendations on detailed decisions such as how the minimum passing score is established and how test results are reported to students, parents, and schools. A more detailed timeline should inform the Board when required decisions are needed and when information for those decisions will be available. Others, including the HSEE Panel and districts and schools themselves, would similarly benefit from a clearer understanding of timelines and plans for remaining HSEE development and implementation activities.

We believe that a development timeline for HSEE would be most useful if it recognizes roles and tasks for as wide a range of stakeholder groups as possible, not just the test developer. Figure 5.1 lists a number of different groups with important roles in HSEE implementation and indicates some of the tasks each of these groups must complete. Figure 5.1 illustrates the range of activities that would be useful to include in a comprehensive timeline. In most cases, the indicated activities are well underway. This list is intended as an example only, not as a comprehensive list of timeline tasks.

In implementing the HSEE, California is setting ambitious goals for student achievement. These goals must be backed up with significant funding to help students achieve them. As suggested by the above list, the range of activities needed to prepare for HSEE implementation is quite broad. The Legislature has already approved funding for HSEE development and for programs to help districts prepare students for the HSEE, including funding for textbook adoption and teacher preparation. Additional resources will surely be needed for remedial programs and for other state and district efforts to ensure a fair opportunity for all students to pass the exam.

Recommendation 3. CDE and the Board should work with districts to identify resource requirements associated with HSEE implementation. The Legislature must be ready to continue to fund activities to support the preparation of students to meet the ambitious challenges embodied in the HSEE.

Examples of Tasks for Inclusion in the HSEE Implementation Timelines

For Districts and Schools:

1. Implement or improve instruction that covers the state content standards
2. Ensure that teachers are well prepared to deliver this instruction
3. Demonstrate that all students have access to this instruction
4. Provide notice to parents and students about HSEE requirement

For the HSEE Panel:

1. Complete recommendations on accommodations and administration issues
2. Review essay questions with complete scoring guides and benchmark papers
3. Consider recommendations for scoring and reporting
4. Consider recommendations for setting minimum passing scores
5. Recommend materials describing test content and procedures
6. Recommend appropriate district/school preparation activities

For CDE:

1. Work with the other stakeholders to maintain implementation timelines
2. Disseminate information to districts and work with them to prepare for HSEE
3. Monitor test development activities and facilitate communication on issues requiring decisions
4. Obtain legal advice on issues with HSEE implementation

For the State Board:

1. Adopt blueprint specifying test content standards
2. Review and approve implementing regulations including recommendations for test accommodations, scoring, and reporting
3. Approve minimum passing scores

For the Test Development Contractor:

1. Complete development and field testing of HSEE questions
2. Assemble test forms
3. Create sample test form and other test information/preparation materials
4. Develop detailed timelines for pre- and post-administration activities
5. Recommend/implement inclusion and accommodation policies
6. Recommend/implement procedures for recommending minimum scores
7. Recommend/implement procedures for scoring and reporting

For the Independent Evaluators:

1. Review and comment on results from the fall field test
2. Review and comment on plans for testing, setting minimum passing scores, scoring, and score reporting
3. Collect data on preparations for HSEE and on the possible consequences of HSEE implementation

Figure 5.1. Examples of Items for Inclusion in the HSEE Development Timelines.

Setting Achievable Standards

At its September meeting, the Board will be asked to approve the content to be assessed by the HSEE (content standards). At a subsequent meeting, the Board will be asked to approve recommendations on the minimum passing score (performance standard) needed to demonstrate adequate achievement relative to these content standards. One way to lessen risks associated with the current implementation schedule would be to ensure that the content and performance standards are not set so high that it will be difficult for most students to meet them.

There are two possible views of the general aims of the HSEE. One view is that California wants the HSEE to identify a moderately small number of students who are not reaching very minimum levels of achievement in mathematics and English/language arts and to help these students reach at least these minimum levels by the end of their high school years. An alternative view is that California wants the HSEE to motivate all students to achieve a broader range and higher levels of mathematics and language arts skills. These two views have very different implications for setting content and performance standards for the HSEE. This leads us to our next recommendation:

Recommendation 4. The Board should adopt a clear statement of its intentions in setting HSEE content and performance standards. This statement should describe the extent to which these standards are targeted to ensure minimum achievement relative to current levels or to significantly advance overall expectations for student achievement.

The content standards being recommended by the HSEE Panel appear to reflect the second view of HSEE goals. For example, a significant number of Algebra 1 standards are recommended while state law does not currently required algebra for graduation. It would appear that the proposed graduation standards are being used to drive changes toward a new curriculum rather than to identify minimal performers under the current curriculum.

Data reported in Chapter 3 above and in our June 30 report suggest that the average 10th grader from the Class of 2002 was likely to answer only about half of the HSEE questions correctly. While this information is tentative, and passing standards have not yet been established, these results suggest that it is likely that, even if the passing standard is as low as 50 % correct, roughly half of the students who take the test will fail on their first try[6]. There are several ways to reduce the probable failure rate:

- Develop easier questions. This is problematic because the HSEE Panel and others have consistently judged the current questions to be good measures of the standards they were designed to assess. Further, there is not time to develop and field-test new questions prior to the March 2001 operational administration.

- Set low standards for passing the exams. It is difficult to imagine setting a performance standard below the 50-percent-correct level. This would imply that

[6] The estimated passing rate is based on all 10th grade students. In 2001, 9th grade students will take the HSEE on a voluntary basis. Passing rates for this group may be significantly higher or lower depending on who volunteers.

students could pass the exam, even though they had not satisfactorily achieved more than half of the content standards. Note also that a frequent difficulty in setting minimum passing scores is that content experts expect the students to demonstrate at least a minimal level of achievement on every one of the content standards.

- Exclude content standards for which students are not yet well prepared. These standards could be introduced for later classes, after instruction related to them has been introduced or improved.

Unless one of the above approaches is taken, California appears to be trying to legislate the Lake Woebegon effect by requiring all students to be above average. Given better instruction and higher motivation, we would expect 10th graders in the Class of 2004 to perform somewhat better than 10th graders in the Class of 2002 did during the Spring 2000 field test. Requiring all students to score above a level that is likely to be higher than the Class of 2002 average would appear to be an unreasonable expectation relative to typical year-to-year gains in student performance. Yet a very plausible interpretation of current data and recommendations is that this is exactly where we are headed.

In 1980, President Bush and the nation's Governors established a set of National Educational Goals. Goal 5 was "By the year 2000, United States students will be first in the world in mathematics and science achievement." Meeting this goal would have required unprecedented gains in student achievement, and available evidence suggests this did not happen. Would a more plausible goal have been taken more seriously and resulted in greater gains? Unfortunately, there have not been controlled studies that would allow us to answer this question. Policy-makers face a difficult dilemma in attempting to set challenging goals that will maximize improvements in student achievement while avoiding expectations that are too unrealistic to be taken seriously. Parallels for the HSEE standards are obvious.

There should be a clear difference between setting goals for school-accountability where the focus tends to be on average achievement and setting goals for student-accountability that focus on the minimum acceptable level of achievement. To illustrate this difference, consider student grades in an Algebra 1 course. On average, we expect students to exhibit a level of performance that will earn them a C+ or B-, but students can pass the course with a minimum grade of D. Will the HSEE be targeting passing standards at the B- or D level? We offer the following recommendation consistent with the use of HSEE as a measure of minimum levels of student achievement:

Recommendation 5. The Board should exhibit moderation in selecting content standards and setting performance standards for the initial implementation of HSEE. Standards should be subsequently expanded or increased based on evidence of improved instruction.

We recognize that this recommendation may greatly increase the difficulty of meeting the current HSEE implementation schedule. If the Board makes significant changes to the proposed content standards at its September meeting, there are likely to be serious consequences for spring 2001 testing. There would not be time to develop and try out questions for any new content areas that are added. Fortunately, our recommendation for

moderation is aimed at limiting rather than expanding standards for minimum performance. If coverage is eliminated or reduced for some of the currently proposed standards, it would seem desirable to increase coverage of the remaining standards to maintain test score accuracy. Such an increase is likely to put a strain on the supply of available questions, but may still be feasible.

After the Board approves HSEE content standards, the contractor can assemble test forms for the spring 2001 administration that conform to these standards. Currently this is scheduled to happen about December of this year. At that point it will be possible to set performance standards, defined as minimum passing scores for each form. AIR is recommending a standard setting process for identifying a minimum score on a base form. This score level would then be equated to scores at the same level of difficulty on succeeding test forms. The "Bookmark" process being proposed is widely used in state assessment programs for setting achievement levels. We generally endorse the proposed approach, although we would like to see more detail on how and when consequence data (expected passing rates) will be introduced.

We believe that the panels charged with recommending performance standards should have access to the important considerations discussed in the setting of the content standards. For that reason, we offer the following recommendation for the process of setting HSEE performance standards.

Recommendation 6. Members of the HSEE Panel and its Technical Advisory Committee should participate in developing recommendations for minimum performance standards.

The Panel could also play a useful role in developing descriptions of the resulting content and performance standards for distribution to students, parents, and teachers. Many of the recommended standards are stated at a general level. It would be helpful to both students and the teachers charged with preparing the students to have a more extended description of each of these standards with examples and sample questions. Given limited time availability, the Panel's role would be at most advisory, with detailed development work carried out by CDE and the development contractor.

Technical Suggestions for Improving the HSEE

We have reviewed general summaries of AIR's plans for conducting the fall field test, developing operational forms, and for spring 2001 operational administration activities. Selected development issues have been discussed with CDE's Technical Studies Group. We are frankly concerned, however, that neither we, nor the HSEE Technical Committee, nor anyone else beyond the contractor's staff, has had a chance to comment on technical issues in the contractor's development plans. We think it would be highly advisable for these plans to be reviewed before they are implemented. In addition, given the real possibility of legal challenges to the HSEE, efforts to establish the defensibility of technical decisions would seem prudent. We offer a specific recommendation for obtaining additional technical review.

Recommendation 7. CDE should move swiftly to establish an independent Technical Issues Committee (TIC) to recommend approval or changes to the HSEE development

contractor's plans for item screening, form assembly, form equating, and scoring and reporting.

CDE has already initiated efforts to obtain independent experts to review technical issues in implementing the HSEE. This would be a significant improvement over the current situation, and could fully address our concern with the need for independent review of technical issues. There is a need for timely advice and a large committee might make take too long to reach consensus. The group that we envision would differ from current CDE efforts by including one or two additional technical reviewers with complementary expertise.

An example may illustrate one need for a broad consensus on technical issues in HSEE development. Field test plans were developed without a clear review of item analysis and screening procedures. Sample sizes were adequate for the overall screening and calibration of the test questions. However, no effort was made to emphasize minority students in sampling and, as a consequence, sample sizes for detecting differential item functioning for African American students were marginal for multiple choice questions and completely inadequate for the essay questions. The state could be legally vulnerable if scores are used operationally to fail students without an adequate check for differential functioning across key demographic groups. It should be feasible to use data from the operational administration of HSEE forms to check for inappropriate group differences. So long as no problems are found, this might not impact the reporting schedule significantly. An independent review of technical aspects of the field test plans might have improved the design considerably and eliminated the need for important analyses during the initial operational score reporting cycle. More importantly, such review can still improve the process as it continues toward full implementation.

A related concern is that very limited information will be available on possible passing rates as the minimum passing score is being established. Approximate information can be pieced together from the field test results on an question-by-question basis, but there will not be much basis for estimating, in advance, the degree of adverse impact for some minority groups (e.g., African Americans) resulting from alternative passing scores.

One of the requirements for the independent evaluation is to provide recommendations for improving the HSEE. Given our reporting schedule, we cannot provide timely advice on all of the technical issues that must be addressed. We are recommending one or more independent reviewers who can be available to address issues quickly as they arise. In the remainder of this section, we do offer advice on two technical issues of immediate concern. The issues are constructing test forms so that comparable scores can be obtained (test equating) and providing feedback to schools that participate in the field tests.

<u>Equating scores from different test forms</u>. Plans for equating different test forms need to be reviewed and finalized before the initial forms are assembled. The contractor's current plan appears to rely on a "pre-equating" approach, where item statistics from the two field tests provide the only basis for creating comparable (equated) scores. We do not believe that total reliance on this form of equating will be sufficient. There are plenty of examples of item position and context effects that would not be accounted for under this plan. It is also likely that differences in student motivation between the field test and operational administrations will affect different questions differently.

In many programs, a common set of "anchor" questions is included in two or more different forms to provide a basis for estimating differences in performance on these forms. For example, achievement differences between the March and May 2001 examinees could be assessed on the basis of a common set of questions. These estimated differences would then be compared to the differences estimated from all of the test questions, based on the "pre-equated" scores and adjustments introduced if needed.

It may not be feasible to use some of the same scored questions in both the March and May forms because of test security concerns or because of a desire to release all of the test questions from the initial administration. AIR's current plan calls for development of a number of different versions of each operational test form with a different set of about 20 non-operational (unscored) questions included in each version. These questions are being administered as a field test and would not count in determining each student's operational score. We recommend that one or two additional versions be developed with a common set of previously screened (all good) questions in the unscored positions. The same set(s) of questions could also be included as unscored questions in the May 2001 test form, providing a basis for linking (or checking the linkage of) scores across these administrations.

Feedback to schools participating in the Field Test. Given the amount of operational testing that exists in schools today, it is understandably difficult to recruit schools to participate in non-operational testing, such as the HSEE field tests. Currently there are no plans to provide any information to participating schools on the performance of their students in the field test. The field test booklets include questions that will be screened out and they are not operational forms. In addition, passing standards have not yet been set. While it is not possible to provide reliable information on individual students, it may still be possible to provide some information to schools. A large number of questions are administered in the field test, roughly 400 for each subject rather than only 100 in an operational administration. It should be possible to provide some information on relative passing rates for questions assessing each specific content standard, pooling data across both students and the questions that assess these standards. The tables in Chapter 3 of this report show passing rates for the questions under each content standard from the spring 2000 field test. It should be possible to construct similar tables showing passing rates by content standard for a particular school in comparison to passing rates for the state as a whole. We believe that such information would be highly valuable to schools in identifying areas where additional preparation is needed. Offering to provide such information could be a strong incentive for participation in the field test.

Other Recommendations in the June 30 Evaluation Report

Clarifying Relationships Among State High School Testing Programs**.** Now that some scores for the STAR standards-based exam have been reported, it seems more important than ever to find a basis for comparing them to HSEE scores. Schools will want to know how performance for individual students on one exam predicts performance on the other. Indeed, proposals have been discussed for exempting students from the HSEE on the basis of scores on the standards-based STAR assessment or scores from the Golden State exams. We had recommended linking STAR and HSEE scores for specific students, but there are some issues with the feasibility of this approach due to difficulties in identifying particular

students. There are many ways the scores from the two assessments might be linked. One alternative would be to administer some of the standards-based STAR questions to HSEE examinees, perhaps as non-scored questions. The information obtained would help in ensuring consistency in standards and expectations for the two assessments.

Dissemination and Assistance to Local Districts. Our June 30 report contained a recommendation on the importance of dissemination and assistance for local districts. CDE has worked to develop dissemination plans. Full implementation of these plans awaits the Board's approval of the examination in September. We have nothing further to add at this time, except to say that funding issues with respect to dissemination and assistance activities should definitely addressed under Recommendation 3 above.

Appropriate Accommodations. The HSEE Panel has collected a great deal of information on options for accommodating special needs students and English language learners. Specific recommendations were discussed at their July 2000 meeting and will be presented to the Board in September. More work is clearly needed to prepare for the possibility that the state or individual districts may be sued by students who feel that they were inappropriately included or excluded or did not receive a comparable score due to some issue with accommodations.

REFERENCES

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159-176.

Scheffe, Henry (1959). *The analysis of variance*. New York: Wiley & Sons, Inc.

Shavelson, Richard J. & Webb, Noreen M. (1991). *Generalizability theory: a primer*. Newbury Park, CA: SAGE Publications.

Wise, L. L., Harris, C. D., Sipes, D.E., Hoffman, R.G., and Ford, P. F. (2000). *The high school exit examination (HSEE): Year 1 evaluation report*. Alexandria, VA: Human Resources Research Organization.

Zwick, Rebecca, Thayer, Dorothy T., and Mazzeo, John (1997). Descriptive and inferential procedures for assessing DIF in polytomous items. *Applied Measurement in Education, 10*, 321-344.