

California State University, San Bernardino

**CSUSB ScholarWorks**

---

Theses Digitization Project

John M. Pfau Library

---

2004

## Effects of environmental factors present during the administration of the California High School Exit Exam on students' outcome scores

Kelly Lynn Coumbe

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/etd-project>



Part of the [Education Commons](#)

---

### Recommended Citation

Coumbe, Kelly Lynn, "Effects of environmental factors present during the administration of the California High School Exit Exam on students' outcome scores" (2004). *Theses Digitization Project*. 2597.  
<https://scholarworks.lib.csusb.edu/etd-project/2597>

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact [scholarworks@csusb.edu](mailto:scholarworks@csusb.edu).

EFFECTS OF ENVIRONMENTAL FACTORS PRESENT DURING THE  
ADMINISTRATION OF THE CALIFORNIA HIGH SCHOOL EXIT EXAM ON  
STUDENT'S OUTCOME SCORES

---

A Thesis  
Presented to the  
Faculty of California State University,  
San Bernardino

---

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in  
Psychology:  
Industrial Organizational

---

by  
Kelly Lynn Coumbe  
December 2004

EFFECTS OF ENVIRONMENTAL FACTORS PRESENT DURING THE  
ADMINISTRATION OF THE CALIFORNIA HIGH SCHOOL EXIT EXAM ON  
STUDENT'S OUTCOME SCORES

---


A Thesis  
Presented to the  
Faculty of California State University,  
San Bernardino


---

by  
Kelly Lynn Coumbe  
December 2004

Approved by:

  
Janelle Gilbert, Chair, Psychology

  
Kenneth Shultz

  
Sharon Ward

11/29/04  
Date

## ABSTRACT

Previous research involving environmental factors and test administration has largely been based on case studies. This study looked at the environmental factors present during testing for the spring 2004 administration of the California High School Exit Exam (CAHSEE) in an attempt to quantify some of the factors that were previously only qualitatively reported. Short phone interviews, consisting of 20 questions, were conducted and environmental factors were examined for 90 schools across California. Five factors were examined for their ability to predict passing percentages of students on the CAHSEE at the school level. The results indicated that socioeconomic status was the only significant predictor. There was a lack of statistical evidence concluding that the environmental factors included in this study predicted passing percentages. These findings were not consistent with previous case studies specific to the CAHSEE and further research is suggested including the inclusion of a social desirability scale.

## TABLE OF CONTENTS

ABSTRACT . . . . .	iii
CHAPTER ONE: LITERATURE REVIEW	
Introduction . . . . .	1
Standardized Testing . . . . .	5
Purpose of the Study . . . . .	6
CHAPTER TWO: TESTING ENVIRONMENT	
Distractions . . . . .	9
Size and Type of Testing Room . . . . .	11
Consistency of Administration . . . . .	12
The Number of Tests Given per Year . . . . .	14
Preparation Courses . . . . .	17
Combination of Environmental Factors . . . . .	19
Socioeconomic Status . . . . .	20
CHAPTER THREE: METHOD	
Participants . . . . .	23
Procedures . . . . .	23
Measures . . . . .	24
Definition of Testing Environment . . . . .	25
California High School Exit Exam . . . . .	26
Standardization . . . . .	27
Distractions . . . . .	28
Number of Students . . . . .	29

Consistency of Administration . . . . .	30
The Number of Tests Given per Year . . . . .	31
Preparation Courses . . . . .	31
Socioeconomic Status . . . . .	32
Outcome Scores . . . . .	32
CHAPTER FOUR: RESULTS	
Description of Data . . . . .	34
Test of Hypotheses . . . . .	37
CHAPTER FIVE: DISCUSSION . . . . .	41
Limitations of Study Design and Procedures . . . . .	43
Distractions . . . . .	44
Number of Students . . . . .	46
Type of Classroom . . . . .	46
Consistency of Administration . . . . .	46
The Number of Tests Given per Year . . . . .	48
Preparation Courses . . . . .	48
Future Research . . . . .	48
Implications . . . . .	52
APPENDIX A: SUMMARY OF HYPOTHESES . . . . .	54
APPENDIX B: SUBJECT MATTER EXPERT RATING FORMS . . . . .	57
APPENDIX C: TABLES . . . . .	67
APPENDIX D: INFORMED CONSENT . . . . .	73

APPENDIX E: DEBRIEFING . . . . .	75
APPENDIX F: INTERVIEW QUESTIONS . . . . .	77
REFERENCES . . . . .	81

# CHAPTER ONE

## LITERATURE REVIEW

### Introduction

According to Linn (2001), Americans have had a "love hate relationship with standardized testing," which dates back over 100 years (p. 29). Tests such as the Stanford Binet, the Army Alpha and Army Beta, and educational tests such as the Iowa Basic and the SAT are some of the more prominent standardized tests. Over the years, there have been positive and negative claims made in regards to standardized testing. These tests have been seen as a window into the true ability of the test taker by some advocates, while critics feel that standardized testing hinders the mental health of the examinees, may promote cheating, and waste valuable learning time as well as being seen as unnecessary (Linn, 2001).

Some of the negative claims may be made due to the skepticism towards the actual properties of the test and the quality and usefulness of the results. Some of the properties used to evaluate tests are reliability and validity. If there were no environmental differences between schools (e.g., disturbances, size and type of



classroom), a test given to fifth graders at school A in May should be highly correlated, or produce similar results, as a test given to fifth graders at school B in May of the same year. This test should also be highly correlated with an administration given to another group of fifth graders at school A the following October. If this is the case, the test is said to be reliable. If a test is reliable, you can be confident that the test will produce similar or comparable results between schools and within schools.

It is also important to show consistency within a test. A good internal consistency coefficient suggests that each item correlates with the overall test score. An alpha reliability coefficient of .9 is often reported for standardized tests in education (Miller, 2001). This is sometimes mistakenly interpreted to mean that it is a good test. However, reliability does not provide a full assessment of the quality of a test. Even when the psychometric properties of the test seem to paint a picture of a sound test, there may be other factors that are important in interpreting the results.

Along with the importance of demonstrating evidence of reliability, a test also needs to demonstrate evidence of

validity. Tests are not valid themselves, but demonstrate evidence of validity depending on their use. In order for the testing results to provide meaningful information and for the proposed inferences to be made from those results, the tests need to demonstrate evidence of validity for the population they were intended to measure.

In general, validity defines the meaning of the outcomes or scores of a test. AERA et al. (1999) suggests that a measure should demonstrate content validity, criterion-related validity, construct validity and/or internal consistency. Content validity refers to the ability of the measure to cover a domain of interest. In this study, it would be the adequacy of the California High School Exit Exam (CAHSEE) and its ability to cover the intended material students needed to understand before graduation, such as English and math competency. Construct validity evidence comes from relationships between a test and other measures of the same behavior or attitudes, as well as the relationship between a test and other, non-related tests. Construct validity evidence theoretically determines if a test is able to predict what it was intended to predict. Convergent validity, would predict that similar tests would produce similar results as

the CAHSEE, and divergent validity would suggest that dissimilar tests would not be related. Criterion-related validity would suggest that the results of the test accurately predict the students' level of knowledge. These relationships are what define the meaning of the test (Feshbach, Weiner & Bohart, 1996).

Unfortunately, estimates of reliability and validity are subject to systematic error variance and therefore do not provide a full assessment of the quality of test results. Tests may appear reliable and valid, when in fact there are factors causing systematic differences in the outcome scores. In assessing the quality and usefulness of test results we must consider confounds that may influence test results. One situation in which systematic differences may cause changes in outcome scores could be described as differences in the level of standardization for tests in practice. For instance, most exams include a test manual for the administrator to follow so that each student who takes the test, no matter which school they attend, will receive the same instructions. In practice, the procedures may not be followed exactly the same at all school sites. Some administrators may follow the instruction manual exactly, where others may tailor the

instructions to their needs or the needs of their classroom. While this may seem necessary to them, they may actually be systematically inflating or deflating their students test scores. If environmental factors present during testing are truly causing systematic differences in outcome scores, these scores would be confounded and the psychometric properties such as validity and reliability could be compromised.

### Standardized Testing

Standardized testing will not be eliminated in the near future. Some of the purposes of these standardized tests today include grade-to-grade promotion, high school exit standards and college entrance exams. These tests are intended to hold students accountable for their learning (Linn, 2001).

In 2002, President Bush signed into law the No Child Left Behind Act, which was established to increase the accountability of students, teachers and the schools. The teachers have standards that are taught and the annual assessments are based on these standards so that the students are being tested consistently across the country. Annual tests or tests that individual teachers use to

assess their teaching are used because research has also shown that past performance is the best predictor of future performance (Gatewood & Feild, 2001). One of the more recent standardized tests being used is the California High School Exit Exam (CAHSEE), which is still in the development stage.

### Purpose of the Study

The purpose of this study will be to examine the effects of several environmental factors on the outcome scores associated with use of the California High School Exit Exam. If the environmental factors systematically affect the outcomes, the psychometric properties of the test would be called into question.

It has been found that differences in testing environments have influenced test scores even when the test characteristics are equal (Ehrenfeld, 2001; Haines, Stansfeld, Head, & Job, 2002; Kim, Baydar, & Greek, 2003). These findings may indicate that there are factors beyond the students' knowledge that are being measured by the standardized test. The size and type of the classroom, number of students in the classroom during the test, and the number of distractions during the test administration

are some of the main differences in testing environments. When the students are able to take tests in a familiar environment with familiar test administrators, it has been reported that they often feel more comfortable (Taylor & Walton, 1998). If the size of the classroom is considerably larger than a typical classroom, students may feel discomfort with the additional number of students present and their outcome scores may be affected.

Additional differences that should be considered include differences in the amount of testing a student receives during the school year, practice sessions that are available before test administration, and disturbances during test administration. These variables may not be standardized between schools and administrations, because they are outside the scope of the administrator's handbook. Administration may also become less standardized when administrators are in charge of a large testing room or large numbers of students, and they are not able to follow the guidelines exactly as stated. Educational Testing Services (ETS) and other large testing agencies can standardize tests to an extent, but there is a large burden placed on the school and the test administrator to provide a consistent environment for testing. Therefore, below I

will discuss the specific environmental factors that were measured in this study and the subsequent effects the factors may have on the outcome scores associated with the use of the California High School Exit Exam.

## CHAPTER TWO

### TESTING ENVIRONMENT

#### Distractions

Kim et al. (2003) found that there were differences in responses from children depending on the noise level while testing. The number of people present in the room, distractions during testing and the light and temperature in the testing environment were among the factors that were also found to create systematic differences in students' standardized test scores. "A quiet, well-lit and comfortable room [was] expected to improve the assessment outcomes" (p. 570) and was supported by their findings. Their findings also supported that students' test results may be a result of not fully understanding how to take a particular test. These factors were all shown to have an impact on performance (Kim et al., 2003).

Cherry and Kruger (2001) found that distractions of any kind decreased performance on standardized tests for all children. Cherry and Kruger (2001) used semantic, linguistic non-semantic and nonlinguistic distracters and found that the semantic distracters, specifically speech forward distracters, had the greatest effect on outcomes.



Wolach and Pratt (2000) found that distracters are processed differently than the memorized items for a task, such as answers to test questions. They included distracters during and after the memorization of items. These distracters, when present, effect the memorization and scanning ability in short term memory. These distracters during test taking may cause a decrease in students' ability to recall previously memorized material. If the distracters are not equal across testing situations in type and frequency, the environments are not standardized. Thus, the factors of the testing environment may have an affect on students' outcome scores in addition to their actual knowledge.

In practice, there may be talking amongst administrators and students, noises from walking, construction, pencil tapping, bells or hallway noise and other variables that would cause the students to be distracted. These distractions may lower their concentration and possibly their ability to answer the questions correctly. Students taking exams in an environment with a high degree of distractions should have lower outcome scores. Conversely, outcome scores will be higher for schools with fewer distractions present during

testing administrations. Thus, the following hypothesis is proposed:

Hypothesis 1: The number of distractions present during testing will be negatively correlated with average passing percentages.

#### Size and Type of Testing Room

The size of the testing classroom may be a factor in addition to the students' knowledge of the test material that affects his or her outcome score. When testing classrooms are similar in size to their typical learning environment, students may feel more comfortable with the number of students in the room. As the room gets larger, not only is the number of students per testing room going to increase, but the comfort level of the student may also decrease (Taylor & Walton, 1998). This may be due to the increased pressure felt on succeeding with increased numbers of students testing at the same time.

These feelings of discomfort may not be removed with the presence of an administrator and they may cause students to score lower on the test. If students are distracted by their discomfort, their scores may be affected. The size of the learning environment will not be

assessed in this study, but the size of the testing room, in itself, will be a factor. While the possibility of cheating may increase in a large setting, more students are likely to be affected by their discomfort with the environment. Due to the fact that most high school students are in a classroom sized learning environment rather than an auditorium or gymnasium setting, outcome scores will be higher for schools that use smaller testing rooms. Thus, the following hypotheses are proposed:

Hypothesis 2: There will be significant mean differences in average passing percentages for schools depending on the type of the testing room (classroom or gymnasium).

Hypothesis 3: The average number of students present in the testing room will be negatively correlated with average passing percentages.

#### Consistency of Administration

The consistency of test administration is also important to increase the standardization of testing procedures. The importance of administration consistency is usually detailed in a handbook or a list of administrator guidelines provided by the test developer.

Test administrators are usually responsible for the instruction, setting up the classroom and providing directions before, during and after the test is administered.

There may be a delay in handing out tests, which could make the administrator rush through the directions, causing students to feel anxious. Even when students are tested in their own classroom, there may be an administrator other than the teacher in charge of testing. It is essential that students get the same treatment, regardless of the administrator. They need to be read the same directions, given the same answers to questions and treated the same way during the test.

Another important aspect of administrators is their knowledge of the testing process and understanding of the procedures. The Standards for Educational and Psychological Testing (AERA et al., 1999) state that administrators need to be properly instructed in the appropriate testing procedures and they need to understand the importance of following the guidelines exactly. They are given materials ahead of time to guide them through the process and it is important that these administrators are clear with regard to what they are expected to do before,

during and after the test. It is crucial that they understand the consequences if they deviate from the standardized test administration process. This would include how they read the instructions, how they announce time limits, how they handle students who finish early and what they do with the tests when they are collected. Thus, the following hypothesis is proposed:

Hypothesis 4: There will be a positive correlation between average passing percentages and degree of consistency with administration guidelines, as outlined by ETS.

#### The Number of Tests Given per Year

Another environmental factor that may influence students results on standardized tests are the number of standardized tests students take per year. The number of tests is not equal across all students or schools, but Haney, Madaus, and Lyons (1993) estimated that the average U.S. public school student is administered between three and eight standardized tests each year.

Crooks (1988) explained the positive effects of annual or semi-annual testing on students. In general, there is an increase in a student's ability to test when he/she is

able to practice what he/she is going to be tested on in format and or content. Students who are familiar with standardized testing in general will have an advantage to those who are not given the opportunity to practice being tested.

There may also be a benefit to taking the same exact test numerous times because students become familiar with the content and the format. Dhar and Marr (1959) found that when groups had prior experience with test material, they showed greater improvement than did the other groups of students due to their practice. Bird et al. (2003) also found that practice effects were present when students were given the same version of a sub-test. Snedden (1931) tested children with the Dearborn Group Intelligence Test and the National Intelligence Test with an interval of one week between the two administrations. He found that the practice effect was fairly large for the group who was given one test one week and a similar test the following week. When children were given an entirely different test, practice effects were still present, but to a smaller extent.

Students who attend schools where they are tested year round using standardized tests will be more likely to have

higher scores on an exit exam, all other things being equal, due to the fact that they are more familiar with the testing process (Rosner, 2001).

Increasing the number of standardized tests alone may not increase a student's score on all exams. The key to testing and success is a balance of testing to familiarize the students and allow needed classroom learning. If students are taken out of their classes over and over again throughout the school year for testing, yet continue to fail the tests, they are missing much needed classroom time directed at learning (Ehrenfeld, 2001).

However, if students are not familiar with standardized tests, they are not as likely to do well. Taylor and Walton (1998) state the importance of students' familiarity with standardized tests by stating that there may be many opportunities to create a positive testing environment, but nothing substitutes a students' knowledge and an understanding of how to take tests. This is important because researchers have shown that familiarity of standardized tests contribute greatly to the overall standardization of the testing environment. Thus, the following hypothesis is proposed:

Hypothesis 5: There will be a positive correlation between average passing percentages and number of testing sessions per year.

### Preparation Courses

Taylor and Walton (1998) explain the importance of students being familiar with test taking not only by taking the same test, but possibly conducting workshops to discuss their attitudes to test taking, and information about the test. This could include understanding multiple choice tests and how to read and think about them. It could also include things like using an optical mark sheet and using a process of elimination strategy when looking at response options.

Rosner (2001) also found that scores on tests such as the SAT and LSAT can be significantly increased with preparation courses. Many schools offer practice sessions for specific tests as well as general tips for testing. This addition of preparatory courses to students' knowledge base may increase their performance on specific tests. Some schools report that there are preparation courses throughout the year, others have after school programs or blocks of time during the regular school day that are



specifically for teaching standardized testing and exit exam material. There is, however, an inconsistency across schools in regards to the content among the preparatory classes offered (Wise, Harris, Brown, Becker, Sun & Coumbe, 2003). Thus, the following hypothesis is proposed:

Hypothesis 6: There will be a positive correlation between average passing percentages and number of opportunities to participate in test preparation courses prior to test administration.

Whether or not the schools are testing too much, or too infrequently, is a problem that remains to be solved. The fact that there are differences in the amount of testing going on in the schools is cause for concern. When the testing process is not standardized, it is unclear if the differences in outcomes are based on students' knowledge or their familiarity with standardized tests in general. It would seem that students would be able to be given the same test in the same environment, regardless of the school they attend. Also, administrators should be careful in delivering the information and the test to keep procedures standardized. In reality, there are other factors that make all of these things impossible or near impossible to accomplish, even with very careful

administration. This study will look at five of those factors.

#### Combination of Environmental Factors

Each of these predictors is proposed to add unique variance to the prediction of outcome scores. All five factors: distractions, size and type of testing room, consistency of administration, the number of testing sessions, and preparatory class offerings, will be looked at in a combined manner to assess their effects on the outcome scores in regards to the California High School Exit Exam. These factors will be coded as to reflect the negative affects towards outcome scores. Thus, the following hypothesis is proposed:

Hypothesis 7: The linear combination of distractions, size and type of testing room, consistency of administration, the number of testing sessions, and preparatory class offerings, will predict average passing percentages.

Most of the literature on the effects of environmental factors has been based on case studies (Cherry and Kruger, 2001; Crooks, 1988; Ehrenfeld, 2001; Kellaghan, Madaus, & Airasian, 1982; Linn, 2001; Miller, 2001; Taylor & Walton,

1998). Thus, there is a complete lack of empirical research relating environmental factors to differences in outcomes. In the present study, I hope to support some of the previous case studies empirically and capitalize on the knowledge given by the administrators from the 2003 administration. In addition, I hope to find empirical data to support the claims made by the administrators. This test is a high-stakes test and it needs to be as standardized as possible before its inaugural implementation in 2005.

#### Socioeconomic Status

There may be an additional variable, socioeconomic status, which needs to be considered in terms of the factors that influence testing outcomes. Socioeconomic status may be an underlying reason that there are differences in the environment in which students are tested. Often high stakes tests, such as exit exams, are seen as measuring socioeconomic advantages rather than the intended material students were responsible for knowing before they leave high school (Miller, 2001).

Socioeconomic status of the school may affect the equipment and resources available, such as numbers of test

administrators and adequate testing rooms. Socioeconomic status could play a large role in the ability of students to take advantage of preparatory courses. For example, they may attend a school where these courses are not offered at all. Knowing what we know about the importance of practice effects on testing outcomes, this could be a huge hurdle for some students.

Jencks and Phillips (1998) found that there is a racial difference in test scores based on a compilation of numerous individuals and tests. They discuss socioeconomic status and how those who are raised in black neighborhoods actually score lower, but the difference is significantly smaller when black children are raised in non-black dominated environments. They found differences in test scores and they suggested that one reason was genetics, but there was no empirical support for this assumption. These differences could be confounded by environmental factors. It is likely that environmental factors would account for these differences regardless of race. This would also be in support of the fact that socioeconomic status may play a factor in students' scores on standardized tests. While Socioeconomic status may covary with the environment, it

won't account for all of the differences. Thus, the following hypothesis is proposed:

Hypothesis 8: Test environment factors (distractions, size of testing room, consistency of administration, number of tests per year and preparatory classes) will be related to average passing percentages after accounting for variance due to average socioeconomic status.

See APPENIX A for a summary of hypotheses.

## CHAPTER THREE

### METHOD

#### Participants

California Public High School principals were contacted to participate in this study. They were assured that any information given during this study would only be reported at the organizational/school level and there would be no identifying information reported at the individual school level. The goal of 97 schools was set, based on Cohen's power primer for six predictors (Cohen, 1992). Due to the response rate of 32%, a final sample size of 90 schools was attained and used for analyses.

#### Procedures

To reach the goal of 97 schools, a random sample of 300 California Public High Schools was initially contacted and asked to participate in this study. The 300 schools were chosen from the entire list of high schools using a random number table. The final response rate was 32%.

Phone calls were made to the principals of all 300 California Public High Schools selected, describing this study and explaining the information that was to be

collected and the timeline for that collection. This call also addressed the importance of the findings in regards to their students. This phone call informed the principals of the informed consent and the 10-minute phone interview to follow. Up to four attempts were made to contact each school site, at which point a failed attempt was concluded.

During the interview, the same wording and questions were administered in the same order to all principals. At the end of the call an open-ended question was asked in regards to any differences they felt their school experienced during the CAHSEE administration.

### Measures

A structured phone interview was administered to all CAHSEE contacts. The structured interview was designed based on evidence that it showed greater validity for comparing results than a non-structured interview (Gatewood & Feild, 2001). Research on structured interviews was examined in designing this interview (Gatewood & Feild, 2001; Miller, 2003). This phone interview took no more than 10 minutes and included 20 questions developed to assess the environmental factors present during the spring 2004 administration of the CAHSEE (See Appendix A).

### Definition of Testing Environment

Testing environment for this study was defined using five factors: 1) the number of distractions, 2) the size and the type of the testing room, 3) the consistency of administration, 4) the number of tests given per year, and 5) the offering of preparation courses. The procedures used when students finish the test, which may have caused differences in the amount of distractions during testing, was also be included. The precautions taken by the school to minimize disturbances such as diverting traffic in the halls and silencing bells during testing may have contributed to the adequacy of the testing environment.

The number of students present during testing defined the size of the classroom. The type of room included classroom and gymnasium. The consistency of administration looked at the administrator's ability to follow the guidelines and collected information as to whether or not they were followed. Information on the number of tests each school delivered to the students in a given year and the availability of preparation courses was included in the evaluation of testing environment.



## California High School Exit Exam

The California High School Exit Exam (CAHSEE) is an exam that is still in the development phase, and is scheduled to be implemented in 2005. This exam is intended to measure students' knowledge of educational material necessary for graduation. Upon passing the exam, it suggests the student has the necessary knowledge base to succeed in college or the workforce.

The test has two parts. It includes English-language arts and Mathematics. The English-language arts part of the exam includes multiple-choice questions, a writing component, and a reading and decoding section that covers vocabulary. Students complete the multiple-choice questions and an essay on one specific topic from text provided. The second part of the test is Mathematics. This section is multiple-choice only. This section includes fractions, working with decimals, basic arithmetic, percentages and mathematical reasoning. The CAHSEE is not a timed test. Students are allowed to take the time necessary to finish the exam.

To pass the CAHSEE, a score of 350 or higher is required on each section. Students need to pass both parts of the exam prior to graduation. They begin this testing

in 10<sup>th</sup> grade and continue until 12<sup>th</sup> grade. They do not need to pass both parts of the test at the same time. Only the sections that have not been passed previously by students will be offered during the following administration for retesting.

### Standardization

Research has been done by Human Resources Research Organization (Wise et al., 2003; Wise, Harris, Koger, Bacci, Ford, Brown, Becker, Sun, Koger, Deatz & Coumbe, 2004) to assess the perceptions of test administrators. The administrators were asked questions relating to some environmental factors during testing and other general perceptions. These studies have been conducted for three years in hopes of helping the CAHSEE get to a point where testing is standardized at all schools. There were, however, only small amounts of empirical data to analyze and the comments made by administrators to open ended questions were only content analyzed. These comments led to the belief that many testing administrations were not standardized. There appeared to be differences in testing room, testing size, levels of disturbances and other environment factor differences in administration. There

was no empirical follow up to see how these factors influenced the outcomes of the students.

### Distractions

Distractions were measured by six questions constructed for this study. The factors were rated as having differing impacts on outcome scores by four Subject Matter Experts. Professionals familiar with the CAHSEE and test administration procedures in general, were asked to rate which of the factors they thought would have more impact. The SME's consisted of two professionals with a PhD in educational assessment, a Human Resources Coordinator who deals with testing, and an administrator at the district office. They were asked to rate each of the variables based on the questions asked during the interview with regards to their effects on the outcome scores. The relevancy and clarity of the items was also assessed by the raters in addition to the degree each factor may influence the outcomes of the test.

The ratings were used to weight each of the questions. Question 1 was weighted .2, Question 2 was weighted .22, Question 3 was weighted .15, Question 4 was weighted .08, Question 5 was weighted .15 and Question 6 was weighted .2 (See APPENDIX C). Weighted scores were summed to create a

composite number reflecting the disturbances during testing administration. A higher number related to an environment with increased disturbances during administration. This technique was based on Kim et al's (2003) procedure.

#### Number of Students

The number of students present during testing defined the size of the classroom. The number was at the school level. Only one school offered more than one administration and was excluded from the final analysis, because averaging their results would give no further information into the testing environment.

The type of classroom was measured with one multiple response option question constructed for this study. The same four Subject Matter Experts rated the following questions. They were asked to rate each of the types of testing rooms based on the questions asked during the interview with regards to their effects on the outcome scores. The relevancy and clarity of the items was also assessed by the raters in addition to the degree each factor may influence the outcomes of the test.

The ratings were used to weight each of the questions. Typical classrooms were weighted .15, auditoriums were weighted .25, cafeterias were weighted .25, and gymnasiums

were weighted .35. No schools offered administration in testing rooms other than classrooms and gymnasiums, so weights were not used in the final analyses.

#### Consistency of Administration

Consistency of Administration was assessed using six questions constructed for this study. These questions assessed the extent to which an administrator followed the guidelines given to them by ETS. The same four Subject Matter Experts rated these questions. They were asked to rate each of the questions based on the consistency of administration with regard to their effects on the outcome scores. The relevancy and clarity of the items were also assessed by the raters in addition to the degree each factor may influence the outcomes of the test.

These responses were summed to create a composite score for each school reflecting their consistency of administration. A higher score corresponded to more consistency in administration (following ETS guidelines). There were not significant differences between SME ratings of the four questions, so all were weighted equally in creating the composite score.

### The Number of Tests Given per Year

The number of standardized tests given per year was recorded as the exact numbers of standardized tests given each year at each school contacted.

### Preparation Courses

Test preparation courses were assessed using four questions constructed for this study. The same four Subject Matter Experts rated these questions. They were asked to rate each of the questions dealing with preparation courses with regard to their effects on the outcome scores. The relevancy and clarity of the items was also assessed by the raters in addition to the degree each factor may influence the outcomes of the test. The ratings were used to weight each of the questions. These responses were added to create a composite score for each school reflecting their ability to help the students prepare for the test. Question 15 was weighted .30, Question 16 was weighted .25, Question 17 was weighted .25 and Question 18 was weighted .20 (See APPENDIX C). There were two follow up questions to the items in this section, which were not weighted. They were used to give a better idea of the schools ability to prepare the students for the exam.

Based on the feedback from the Subject Matter Experts (SME's), all proposed questions were included in the structured interview. There were no additional questions suggested by the SME's. The California High School Exit Exam is not a timed test, so questions regarding timing and announcing time periods throughout the test were not added to the interview questions.

#### Socioeconomic Status

Socioeconomic status (SES) was assessed based on information regarding the average SES for each school. This information was gathered using the DataQuest Web site, an online compilation of school data for the state of California. DataQuest reports the percentage of students taking advantage of the free and reduced price lunch program, which is an already accepted form of measuring average school SES (California Department of Education: DataQuest, 2004). Those schools with higher percentages reflected more students being offered a free or reduced priced lunch, which translated to lower SES scores.

#### Outcome Scores

Outcome scores were gathered from the DataQuest Web site using passing percentages at the school level. The percentage of students who passed the CAHSEE during the

spring 2004 administration was recorded as the passing percentage outcome score. This gave a solid percentage of students who passed and failed the spring 2004 administration of the CAHSEE.

An additional variable, Academic Performance Index (API) scores for each school, was gathered at the school level from the DataQuest Web site. The API is an index that is reported for each school in the state of California. It is based on the schools overall scores on five standardized tests (California Science Standards Test - Science CST, California Standards Tests - CST's, California Alternate Performance Assessment - CAPA, California Achievement Test, Sixth Edition Survey - CAT/6 Survey, and the California High School Exit Exam - CAHSEE) as well as their growth performance from the previous year. These API scores were reported in whole number increments that related to the schools overall performance for 2003. The test process relating to the CAHSEE also may have been related to the API. There may have been an effect on the outcome scores of the API due to the testing environment similar to the affects the testing environment has on the outcome scores for the CAHSEE. This was beyond the scope of the study, but was collected for further understanding.



## CHAPTER FOUR

### RESULTS

#### Description of Data

Before running any analyses, the data were examined for outliers, out of range values and missing data. The sample included 90 cases with an average number of students per classroom of 28.56 (as reported to the California Department of Education), and an average API score of 662 (See APPENDIX C for mean scores and standard deviations). Ninety percent of the schools reported having testing session with 35 or fewer students present during the 2004 administration of the CAHSEE. Ten percent of the schools reported 80-250 students present at any given testing session. Eighty one of 90 schools reported holding testing sessions in a typical classroom while the other 9 schools reported holding testing sessions in the gymnasium. Eighty six of 90 schools sampled reported administering four standardized tests per year.

The percentage of students taking advantage of free or reduced price lunch at any given school ranged from 1.1% to 89.9% with an average of 35.25%. English passing percentages for the 2004 CAHSEE ranged from 46%-98%, Math

passing percentages for the 2004 CAHSEE ranged from 50%-99%. API scores ranged from 497 to 841 for the sample of schools in this study. An independent samples t-test was run on all 300 schools to see if there were differences between the schools that responded and those that did not respond. No significant mean differences in API scores were found for respondents ( $\mu = 662.22$ ) and non-respondents ( $\mu = 680.40$ ,  $t = -1.45$ ,  $p > .05$ ).

Using the value of  $\pm 3.3$ , there were nine cases in the data set that were identified as univariate outliers. Only one school reported significant outside noises during testing,  $z = 9.38$ . Only three cases reported having bells or phones ring during testing,  $z = 5.36$ . One school reported 250 students present during testing,  $z = 5.51$ . A second school reported 190 students present during testing,  $z = 3.94$ . Only two schools reported that they needed to paraphrase the instructions given to them by the California Department of Education,  $z = 6.60$ . Only three of the 90 schools did not offer preparation courses of any kind prior to the 2004 administration of the CAHSEE,  $u = 5.36$ . Using  $\pm 3.3$  as criteria for continuous variables, there were no outliers found. There was nothing unusual about the survey and all participants were from the population of interest.

The outliers were due to the lack of variance on items; therefore, no outliers were removed.

There were no out of range values found in any case. While analyzing missing values, only two variables were missing data. 1.1% of the schools did not answer the question about online preparation offered. 4.4% of the schools did not have a documented API score for 2003. All other variables were missing less than 1% of the data.

Variables were combined to create a composite score for the following variables: Distractions, Administration, Preparation Classes and the average passing percentage of the CAHSEE 04. The following variables: significant disturbances, construction noise, outside noise, bells and phones, leaving after the second test, and talking during testing, were combined to make the variable distraction to signify the overall level of distractions. A higher number was related to more distractions present during the CAHSEE administration.

The following variables: inventory, video, instructions, extra testing time and moving locations, were combined to make the variable administration to reflect the overall consistency of administration for each school. Inventory, video, and extra testing time were reverse

scored. A higher number related to less consistency of administration.

The following variables: preparatory courses, teacher preparation and online preparation were combined to make the variable preparation to reflect the overall amount of preparation classes offered by the school prior to administration. All three variables were reverse coded. A higher number related to fewer options for preparation course offered by the school.

The following variables: passing percentage for English and passing percentage for math, were combined to create the average passing percentage variable for the CAHSEE 04 to reflect the overall average passing percentage for each school. This score was computed by adding the two passing percentages for math and English and dividing by two.

### Test of Hypotheses

Hypotheses 1 and 3 through 6 were analyzed using a Pearson Correlation. The variables number of students in room, number of tests per year, and the composite factors distractions, administration and preparation, were looked at to see if they were significantly correlated with the

average passing percentage on the CAHSEE 04. There were no significant correlations. Analyses were run on the weighted composites and no significant correlations were found. For exploratory purposes, the relationship between the environmental factors (weighted and unweighted) and API scores were examined. There were no significant results (See APPENDIX C for a table of correlation coefficients).

Hypothesis 2 was analyzed using a t-test. There were no significant mean differences in passing percentages based on the type of room a school used for testing ( $t = 1.16, p > .05$ ). Students testing in a classroom ( $N = 81$ ) had an average passing percentage of 78.45 and students taking the CAHSEE in a gymnasium ( $N = 9$ ) had an average passing percentage of 73.67.

There were also no significant mean differences in API scores based on the type of room a school used for testing ( $t = .513, p > .05$ ). Schools who reported testing in a classroom had an average API score of 663.84. Schools reporting testing in a gymnasium had an average API score of 648.33.

Hypothesis 7 was analyzed using multiple linear regression. The linear combination of predictors (number of tests per year, students in testing room, preparation,

administration, distractions) did not significantly predict passing percentages ( $R^2 = .045$ ,  $p > .05$ ). A multiple linear regression was also used to analyze the weighted linear combination of predictors based on the SME ratings. The linear combination of predictors (number of test per year, students in testing room, preparation (w), administration and distractions(w)) did not significantly predict passing percentages ( $R^2 = .049$ ,  $p > .05$ ).

Similar to hypothesis 7, API scores were examined. The linear combination of predictors (number of tests per year, students in testing room, preparation, administration, distractions) did not significantly predict the schools API score ( $R^2 = .069$ ,  $p > .05$ ). The weighted linear combination was also unable to significantly predict school API scores ( $R^2 = .075$ ,  $p > .05$ ).

Hypothesis 8 was analyzed using multiple linear regression. When controlling for socioeconomic status, the following predictors: number of test per year, students in testing room, preparation, administration, and distractions, did not significantly add to the prediction of passing percentages ( $R^2$  change =  $.018$ ,  $p > .05$ ). The model including the weighted predictors did not significantly add to the prediction of passing percentages

( $R^2$  change = .019,  $p > .05$ ). Socioeconomic status significantly predicted passing percentages ( $R^2 = .540$ ,  $p < .001$ ).

Similar to hypothesis 8, API scores were examined. The model including the following predictors: number of test per year, students in testing room, preparation, administration, and distraction, did not significantly add to the prediction of API scores ( $R^2$  change = .021,  $p > .05$ ). The model including the weighted predictors did not significantly add to the prediction of API scores ( $R^2$  change = .012,  $p > .05$ ). Socioeconomic status significantly predicted passing percentages ( $R^2 = .464$ ,  $p < .001$ ).

## CHAPTER FIVE

### DISCUSSION

The findings of this study lead us to believe that there are no differences in environmental factors present during testing. This is not consistent with previous case studies in which test administrators, teachers and principals were asked to discuss the environmental factors present during previous administrations of the CAHSEE. They reported differences in administration, type of testing room, noises present and the preparedness of the students prior to taking the test (Wise, et.al, 2003; Wise, et.al, 2004).

The results also showed that environmental factors did not affect the passing percentages of students on the CAHSEE or the API scores of the schools in the study. This is not consistent with previous research that has shown that environmental factors, such as room size and distractions may decrease students' performance on tests (Ehrenfeld, 2001; Haines, Stansfeld et.al, 2002; Kim, et.al, 2003 and Taylor & Walton, 1998). Those individuals interviewed by HumRRO (2003, 2004) also believed that the differences in environmental factors affected the outcome



scores for the students who previously took the CAHSEE. They indicated that there were factors beyond the students' knowledge that they believed were being measured by the standardized test.

In this study, the only significant predictor was socioeconomic status. Socioeconomic status was an objective measure that was not included in the interview. This measure differed between schools and was able to significantly predict outcome scores on the CAHSEE and the API. These findings were consistent with previous findings related to socioeconomic status and the subsequent affects it may have on students standardized test scores (Jencks & Phillips, 1998; Miller, 2001). The socioeconomic status of the school may affect the equipment and resources available, such as numbers of test administrators and adequate testing rooms. The room condition, adequacy of the desks, overall resources and number of test administrators per student was not measured in this study.

The environmental factors present during testing that were measured in this study were the following: distractions, room type, number of students present, number of tests per year, preparation and administration. These factors did not produce enough variance to support

hypotheses leading to the significant prediction of outcome scores. Based on the understanding that environmental factors are a way of measuring socioeconomic status, there is still reason to believe that these factors predict outcome scores.

#### Limitations of Study Design and Procedures

This study was conducted using an interview rather than a survey which may have led to less trust on behalf of the participant that the process was anonymous. The phone interview may have helped make the process less personal than a face to face interview, however; specific schools could still be identified. Participants may have been concerned about negative ramifications for their school and answered based on this concern rather than the specific environmental factors present during testing.

This study was intended to speak specifically with testing coordinators and administrators for the spring 2004 California High School Exit Exam administration. However, the interviews were not conducted until June and the principals were the ones who were referred to when the interview was conducted. Many of the principals were very willing to answer questions and excited for the results.

The lack of variability in the responses may have been due to the principal's inability to recall the details from the testing session, their absence of true knowledge of what went on in the testing session or their unintentional favorability in responses due to social desirability.

### Distractions

There is an unlimited amount of possible distractions that can be present in a testing environment. The interview questions captured those distractions that were brought up in previous research (Ehrenfeld, 2001; Kim, et al., 2003) and others that were created for this study to add to previous research. There may also be other distractions that were not captured in this study that affect students' performance.

The distractions included in the interview may not have been noted during the testing sessions and forgotten by the principal before the interview occurred. Due to the dichotomous response option of yes or no, the participants may have felt that the distractions were so minimal that they responded with "no." In this study, there was not a way to determine the level of distractions present based on individual items.

Social desirability may have also played a role in the responses for this factor. Most schools would strive to create a quiet testing room with minimal outside noise and talking. This may have been the case in the schools included in this study, but principals may have answered questions consistent to what they prepared for or desired, rather than the actual environment. If there was no documentation of distractions and the principal had forgotten, they may also have responded to the interview questions in a socially desirable way.

The HumRRO studies (Wise, et al., 2003; Wise, et al., 2004) indicated that school administrators were concerned about the consistency of administration and other environmental factors present during testing. The administrators indicated that there was variance in environmental factors in their open-ended comments. Subject matter experts showed their belief that different environmental factors are present during testing as well as previous administrators suggesting those differences. There may be reason to believe that people weren't reporting the variance in environmental factors that were present during the administration, which may be explained by social desirability.

### Number of Students

The measure of the number of students in the classroom may be unreliable based on the responses given. Some schools kept exact numbers of students present, while others appeared to give an estimate. This may be due to a lack of documentation. There was also a lack of variability in this factor due to the fact that only two types of rooms were reported in this study as being used for the CAHSEE administration.

### Type of Classroom

This study intended on capturing variance in this factor by differentiating between size of classrooms and other types of rooms available at a school site for testing. However, the 90 schools who participated in this study only reported using two of those rooms: typical classrooms and gymnasiums. Eighty one schools used a classroom setting to administer the CAHSEE, while only nine used a gymnasium.

### Consistency of Administration

There was very little instruction given to the administrators prior to administration, which may have caused inconsistency in administration. This was one of the items that came up during the content analysis portion

of the HumRRO studies (Wise, et al., 2004). Due to the lack of guidance given to administrators prior to the test, there are not a lot of items that we could use for this study in regards to consistency of administration.

ETS produced the materials for the administration of the CAHSEE and the guidelines were to be followed exactly. Many schools reported that they followed the guidelines exactly which could have led to the low levels of variability. Social desirability may have also played a role in the responses for this factor. All schools are supposed to follow the guidelines set forth by ETS, so admitting that they were not followed may be undesirable. The principals may have answered in a way consistent with how they feel they should have administered the test rather than how the test was actually administered. If any of the schools questioned their anonymity, they may have been even more likely to reply with favorable responses. Some of the information in the questions regarding the administration may have been forgotten and a desirable response may have been provided.

### The Number of Tests Given per Year

The number of standardized tests given to high school students is governed by the state which may have led to the low variability in this factor.

### Preparation Courses

There was minimal variability in responses to any of the preparation course offering questions. This may have been due to the fact that so many schools offered test preparation as part of the normal curriculum and so many resources are now available on-line. The interview questions may not have tapped into the actual differences in preparation courses and preparation material between schools in California.

Overall, there was a lack of variability in the environmental factors examined in this study. This lack of variability reduced the possibility of finding any effects on outcome scores. Social Desirability and the methodology used for collecting data may have aided in this overall lack of variability.

### Future Research

This study was conducted using a phone interview 3-4 months after the completion of the spring administration of

the California High School Exit Exam. Future studies may benefit from a data collection window closer to the close of the administration dates. Poor memory and lack of documentation may have caused some of the lack of variability in responses that may be overcome by collecting data sooner.

In addition, future studies should be conducted using an on-line or paper survey rather than an interview. This would allow a greater distribution of the measures and may produce greater feelings of anonymity and allow more data to be gathered. They could be distributed the week following the administration and the details from the administration may be easier to recall for the participant. The survey may also allow for participants to look up or verify any information that they may not remember or know off-hand during a phone interview. Surveys may be less intrusive than a phone interview and participants would have a window of time to complete it rather than immediately when the phone rings.

Another option for data collection would include observations. This way the details of the environment would be captured in the moment. This could be done along with the survey to get a more detailed understanding of the



factors that were present during testing that may have affected the student's performance.

The participants in this study were all principals, but they may or may not have actually been present in the testing room. If they were in the room, they may not have stayed the entire time. Follow-up studies could include specific questions about the participant's position in the school and their involvement in the testing process (e.g., did they coordinate the process, did they receive the materials personally, were they responsible for the administration or did they assist with the administration).

The response options for this study were all yes/no. This did not allow for a range in the factors. Further studies may include a 5-point strongly disagree to strongly agree scale, which may increase the variability in responses.

There may have been many other factors that were related to the overall testing environment that were not included in this study, which may have caused the lack of significant results. A future study could be conducted using different testing environment factors or clarification of the factors that were previously included.

The type of classroom the students are in on a normal day versus the testing day may significantly predict outcome scores. The person who administers the test and the actual students in the room may be a factor. If students are tested with their homeroom class in their familiar classroom with a familiar teacher, they may score significantly different than schools that test students in an unfamiliar classroom with less familiar students and teachers.

The type of the classroom could be specified to a greater extent than it was in this study. The room type could allow for a range of size in a classroom setting based on maximum capacity to allow for more variability in responses. Other questions could get at specific details of the room such the use of desks or tables for testing.

Social desirability was a concern in this study. Due to this concern, future research may benefit from the inclusion of a social desirability scale. This would help to identify the extent to which participants answer questions in a socially desirable way versus what environmental factors were actually present during the administration.

Socioeconomic status was the strongest predictor of passing percentages for the spring 2004 CAHSEE administration. Future research may include more than one objective measure of SES in an attempt to better identify the environmental factors that are present during testing that effect outcome scores. The inclusion of measures that would get at the specific resources available at each school and the demographics of the area would aid in the understanding of the effects of environmental factors during testing.

#### Implications

While the findings of this study were not statistically significant, they do have practical implications. For example, this study helped to quantify some of the effects and the degree of influence environmental factors may have on testing environments. Social desirability most likely played a role in this study. The results are not consistent with previous research on the CAHSEE administration. Based on HumRRO's research (Wise et al., 2003, Wise et al., 2004) it is believed that testing administration for the CAHSEE was not as standardized as this study found. This may lead to a

need for further administration manuals, a way to check if guidelines are followed and a more reliable way to measure the environmental factors during testing.

Socioeconomic status was found to be a driving factor in the prediction of students passing percentages on the CAHSEE. The participants did not show variance in the factors as they were described in the interview questions, but there is evidence that they exist. As previously stated, subject matter experts and previous test administrators have suspected and reported greater variance in environmental factors than the variance in factors found in this study.

This may be evidence that additional factors exist above and beyond the results. There may be a need to assess the factors being measured rather than just measuring them. The content and effectiveness of the preparation courses may need to be assessed rather than just making sure it is available. The quality of administration and the detail of documentation may also be important to look into.

APPENDIX A  
SUMMARY OF HYPOTHESES

## Summary of Hypotheses

Hypothesis 1: The number of distractions present during testing will be negatively correlated with average passing percentages.

Hypothesis 2: There will be significant mean differences in average passing percentages for schools depending on the type of the testing room (classroom or gymnasium).

Hypothesis 3: The average number of students present in the testing room will be negatively correlated with average passing percentages.

Hypothesis 4: There will be a positive correlation between average passing percentages and degree of consistency with administration guidelines, as outlined by ETS.

Hypothesis 5: There will be a positive correlation between average passing percentages and number of testing sessions per year.

Hypothesis 6: There will be a positive correlation between average passing percentages and number of opportunities to participate in test preparation courses prior to test administration.

Hypothesis 7: The linear combination of distractions, size and type of testing room, consistency of administration, the number of testing sessions, and preparatory class offerings, will predict average passing percentages.

Hypothesis 8: Test environment factors (distractions, size of testing room, consistency of administration, number of tests per year and preparatory classes) will be related to average passing percentages after accounting for variance due to average socioeconomic status.

APPENDIX B

SUBJECT MATTER EXPERT RATING FORMS



## Distractions

The following questions will be asked of test coordinators during a phone interview.

1. Were there significant disturbances in the hallways during testing?
2. Was there any construction noise during testing?
3. Were there outside distractions during testing other than construction?
4. Did any classroom bells/phones ring during testing?
5. Were students allowed to leave the testing room after they finished each section? If no (question 5), were students allowed to leave the testing room after the completion of the test?
6. Was there talking or other classroom noise during testing?

Please rate the extent you feel distractions during testing, as described in each question above, would affect outcome scores on the California High School Exit Exam.

1 would not affect outcome scores	2 may affect outcome scores slightly	3 would strongly affect outcome scores
--	---	---

- |    |               |
|----|---------------|
| 1. | 1-----2-----3 |
| 2. | 1-----2-----3 |
| 3. | 1-----2-----3 |
| 4. | 1-----2-----3 |
| 5. | 1-----2-----3 |
| 6. | 1-----2-----3 |

Please make notes in regards to any wording changes, and the clarity and relevance of the items. If there are any other relevant items you feel should be added to the measure of disturbances, please add them below and rate them (1,2,3).

---

---

---

---

---

---

---

## Size and Type of Testing Room

The following questions will be asked of test coordinators during a phone interview.

7. What is the average number of students present in each testing room during the March administration of the CAHSEE?

i. What type of room are the students tested in (answer all that apply)?

1. Typical classroom setting
2. Auditorium
3. Cafeteria
4. Gymnasium
5. Other: \_\_\_\_\_

Please rate the extent you feel size and type of testing room, as described in each question above, would affect outcome scores on the California High School Exit Exam.

1 would not affect outcome scores	2 may affect outcome scores slightly	3 would strongly affect outcome scores
--	---	---

7.        1-----2-----3

i.

1. 1-----2-----3

2. 1-----2-----3

3. 1-----2-----3

4. 1-----2-----3

5. 1-----2-----3

Please make notes in regards to any wording changes, and the clarity and relevance of the items. If there are any other relevant items you feel should be added to the measure of the size and type of the testing room, please add them below and rate them (1,2,3)..

---

---

---

### Consistency of Administration

The following questions will be asked of test coordinators during a phone interview.

8. Did you inventory all test material as it arrived at your school to make sure you had the necessary resources for administration?
9. Did you watch the test administration video?
10. Did you read the instructions exactly as they were written when giving the test?
11. Did you plan and provide space for students who needed extra time to take the test?

Please rate the extent you feel the consistency of administration, as described in each question above, would affect outcome scores on the California High School Exit Exam.

1 would not affect outcome scores	2 may affect outcome scores slightly	3 would strongly affect outcome scores
--	---	---

8. 1-----2-----3

9. 1-----2-----3

10. 1-----2-----3

11. 1-----2-----3

Please make notes in regards to any wording changes, and the clarity and relevance of the items. If there are any other relevant items you feel should be added to the measure of the consistency of administration, please add them below and rate them (1,2,3).

---

---

---

---

### Number of Tests Given Per Year

The following questions will be asked of test coordinators during a phone interview.

12. How many standardized tests are administered at your school each year?

Please rate the extent you feel number of tests given per year, as described in the question above, would affect outcome scores on the California High School Exit Exam.

1	2	3
would not affect outcome scores	may affect outcome scores slightly	would strongly affect outcome scores

12. 1-----2-----3

Please make notes in regards to any wording changes, and the clarity and relevance of the items. If there are any other relevant items you feel should be added to the measure of number of tests given per year, please add them below and rate them (1,2,3).

---

---

---

---

## Preparation Courses

The following questions will be asked of test coordinators during a phone interview.

13. Did your school offer preparation courses for the students prior to test administration?
  - a. What percentage of students participated in these classes?
14. Did the teachers offer preparation during the school day to help aid students in test taking strategies?
15. Were students directed to on-line or print resources to help them pass the CAHSEE?
16. Were there any other options that students could take advantage of to prepare for the CAHSEE that your school offered?
  - a. Please describe these options.

Please rate the extent you feel opportunities for prep courses prior to testing, as described in each question above, would affect outcome scores on the California High School Exit Exam.

1 would not affect outcome scores	2 may affect outcome scores slightly	3 would strongly affect outcome scores
--	---	---

13.      1-----2-----3
- a.    1-----2-----3
14.      1-----2-----3
15.      1-----2-----3
16.      1-----2-----3
- a.    1-----2-----3



Please make notes in regards to any wording changes, and the clarity and relevance of the items. If there are any other relevant items you feel should be added to the measure of preparation courses, please add them below and rate them (1,2,3).

---

---

---

---

## APPENDIX C

### TABLES

Table 1

*Mean Scores and Standard Deviations for Responding Schools*

<hr/>		
Factor		
	M	SD
Number of students per room	39.42	38.20
Room type	1.10	.30
Distractions	1.01	.03
Administration	1.91	.14
Preparation	1.69	.23
Number of tests per year	3.94	.27
Percentage of students taking advantage of free or reduced price lunch	35.25	23.22
<hr/>		
Outcome		
	M	SD
Average percentage of students who passed the CAHSEE 2004	77.99	11.79
Average API Score	662.22	85.54
<hr/>		

Table 2

*Correlation Coefficients between Environmental Factors and Outcome Scores*

Factor	API	CAHSEE
Number of students per room	.01	-.04
Room type	-.06	-.12
Distractions	.13	.13
Administration	.16	.11
Preparation	.17	.11
Number of tests per year	-.10	-.10
Percentage of students taking advantage of free or reduced price lunch	-.69**	-.74**
Weighted Factor	API	CAHSEE
Distractions	.13	.13
Preparation	.17	.11

\*  $p < .05$

\*\*  $p < .01$

Table 3

*Testing Environment Factors with Weights Based on Subject Matter Experts Ratings of Importance*

---

Distractions	
--------------	--

---

	Weight
Disturbances in the hallways	.20
Construction noise	.22
Outside distractions	.15
Classroom bells/phones	.08
Leave after each section	.15
Talking during testing	.20

---

Preparation	
-------------	--

---

	Weight
School offered prep courses	.30
Teacher prep during school	.25
On-line or print resources	.25
Other prep options	.20

---

Table 4

*Regression Coefficients for Weighted and Unweighted  
Environmental Factors and Average Passing Percentages*

Environmental Factor	Average Passing Percentage	
	Weighted	Unweighted
<hr/>		
Distractions		
R <sup>2</sup>	.025	.081
M	1.15	1.01
SD	.02	.03
Preparation		
R <sup>2</sup>	.025	.081
M	1.31	1.69
SD	.20	.23
<hr/>		

\*  $p < .05$

\*\*  $p < .01$

Table 5

*Hierarchical Regression Results for Average Passing  
Percentages on the 2004 CAHSEE*

Predictor Variables	Beta	R	R <sup>2</sup>	R <sup>2</sup> Change
<b>STEP ONE Socioeconomic Status</b>		<b>.74</b>	<b>.54</b>	<b>.54**</b>
Percentage of students Taking advantage of free/ reducedprice lunch	-.74**			
<b>STEP TWO Environmental Factors</b>		<b>.75</b>	<b>.56</b>	<b>.02</b>
Preparation	.04			
Number of students in room	-.13			
Administration	-.02			
Distractions	.08			
Number of test per year	.06			

\*  $p < .05$

\*\*  $p < .01$

APPENDIX D  
INFORMED CONSENT



## Informed Consent

You are invited to participate in a study to investigate the relationship between testing environment and passing percentages of students taking the California High School Exit Exam. This study is being conducted by Kelly Coumbe under the supervision of Janelle Gilbert, professor of Psychology. This study has been reviewed and approved by the Psychology Department Review Board of California State University San Bernardino. It is in fulfillment of part of the requirements for my Masters Thesis in Industrial/Organizational psychology from California State University, San Bernardino. It will be looking at environmental factors present during the administration of the California High School Exam. I will be asking you questions in regards to the administration and the testing environment during the March CAHSEE administration sessions. There will be no record of your personal information or your school information. After completing the phone interview, I will assign you an identification number. This number will be the only way that your data will be identifiable. My advisor and myself will be the only two individuals with access to your identification number, but it will not be attached to you or your school. If you wish to make any changes to your data during the study or withdraw your data, you will need to contact me and provide your identification number. I will have no other way of tracing your information back to you or your school. There are no foreseeable risks to participating in this study and you have the right to quit at any time. Would you like to continue with the study?

Yes \_\_\_\_\_

No \_\_\_\_\_

## APPENDIX E

### DEBRIEFING

## Debriefing

Thank you for participating in this study. This study was conducted in hopes of furthering what we know about testing environment factors on the passing percentage of students taking the California High School Exit Exam. This information will be given in summary form to the California Department of Education and any interested schools. I hope to help standardized testing procedures, if necessary, before this test is implemented in 2005. If you have any questions, or would like to withdraw your data, please use ID number \_\_\_\_\_ for further information. Please call Dr. Janelle Gilbert at 909-880-5587. The results of this study will be available in November of 2004.

APPENDIX F  
INTERVIEW QUESTIONS

### Distractions

1. Were there noticeable disturbances in the hallways during testing?
2. Was there any construction noise during testing?
3. Were there outside distractions during testing other than construction?
4. Did any classroom bells/phones ring during testing?
5. Were students allowed to leave the testing room after they finished each section? If no, were students allowed to leave the testing room after they completion of the test but before everyone completed the test?
6. Was there talking or other classroom noise during testing?

### Size of Testing Classroom

7. What is the average number of students present in each testing room during the March administration of the CAHSEE?

#### Type of Testing Classroom

- i. What type of room are the students tested in (check all that apply)?
  1. Typical classroom setting
  2. Auditorium

3. Cafeteria

4. Gymnasium

5. Other: \_\_\_\_\_

Consistency of Administration

9. After completing the inventory of all test materials as they arrived at your school, did you have all the necessary resources for administration?

10. Were you given an administration video to help answer any questions?

11. Were you able to watch the video?

12. Sometimes it is hard to read the instructions exactly as they are printed. Did you find that you used your own words or needed to paraphrase the instructions during administration?

13. Although the CAHSEE is not a timed test, some students need much more time than others. Were you able to provide space for students who needed extra time to take the test?

14. Did you need to move them to a new location?

Number of Tests per Year

15. How many standardized tests are administered at your school each year?

### Preparation Classes

16. Did your school offer preparation courses for the students prior to test administration?
  - a. What percentage of students participated in these classes?
17. Did the teachers offer preparation during the school day to help aid students in test taking strategies?
18. Were students directed to on-line or print resources to help them pass the CAHSEE?
19. Were there any other options that students could take advantage of to prepare for the CAHSEE that your school offered?
  - a. Please describe these options.

### Other Information

20. Is there any other information that you would like to share that helped or hindered the test environment that you were able to provide to students during the spring 2004 administration of the CAHSEE?

## References

- American Educational Research Association/American Psychological Association/National Council on Measurement in Education (1999) Standards for Educational and Psychological Testing. Washington, D.C.: American Educational Research Association.
- Bird, C.M., & Papadopoulo, K., Ricciardelli, P., Rossor, M.N., Cipolotti, L. (2003) Test-retest reliability, practice effects and reliable change indices for the recognition memory test. *British Journal of Clinical Psychology*, 42(4), 407-425.
- California Department of Education: Data Quest. (2004). Retrieved February 15, 2004, from <http://data1.cde.ca.gov/dataquest/>
- Cherry, R.S., & Kruger, B. (2001). Selective Auditory Attention Abilities of Learning Disabled and Normal Achieving Children. *Journal of Learning Disabilities*, 16(4), 202-205.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Crooks, T.J. (1988). The impact of classroom evaluation practices in students. *Review of Educational Research*, 58, 438-481.
- Dhar, C., & Marr, E. (1959). Study of the practice effect of taking Raven's Matrices on performance on the N.I.I.P. Group Test 70/23 and of the N.I.I.P. on the Raven's Matrices. *Journal of Vocational and Educational Guidance - Bombay*, 5(4), 187-189.
- Ehrenfeld, J. (2001). The Trouble With Testing. *Education Week*, 21(8), 43.
- Educational Testing Services (1999): The uses of standardized tests in American education. Princeton: ETS.



- Feshbach, S., Weiner, B., & Bohart, A. (1996). *Personality* (4th ed.). Chapter 7. Lexington, MA: Heath.
- Gatewood, R.D., & Feild, H.S. (2001). *Human Resource Selection*. Fifth edition. Chapter 4, pp. 113-160, Chapter 5, pp. 161-226. Orlando, FL: Harcourt, Inc.
- Haines, M.M., Stansfeld, S.A., Head, J., & Job, R.S.F. (2002) Multilevel modeling of aircraft noise on performance tests in schools around Heathrow Airport London. *Journal of Epidemiology and Community Health*, 56(2), 139-144.
- Haney, W.M., Madaus, G.F., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kluwer.
- Jencks, C., & Phillips, M. (1998). *The Black White Test Score Gap*. Brookings Institute Press: Washington, D.C.
- Kellaghan, T., Madaus, G.F., Airasian, P.W. (1982). *The Effects of Standardized Testing*. Kluwer-Nijhoff Publishing: Boston
- Kim, H. Baydar, N., & Greek, A. (2003). Testing Conditions Influence the Race Gap in Cognition and Achievement Estimated by Household Survey Data. *Journal of Applied Developmental Psychology*, 23(5), 567-582).
- Linn, R.L., (2001). A Century of Standardized Testing: Controversies and Pendulum Swings. *Educational Assessment*, 7(1), 29-38.
- Miller, K. (2003). Conducting Cognitive Interviews to Understand Question-Response Limitations. *American Journal of Health Behavior*, 27(3), S264-S272.
- Miller, D.W. (2001). Scholars Say High-Stakes Test Deserve A Failing Grade. *Chronicle of Higher Education*, 47(25).
- Rosner, J. (2001). Disparate Outcomes by Design: University Admissions Tests. *Berkeley La Raza Law Journal*, 12(2), 377-386.

- Snedden, D. (1931). Practice effect. *Journal of Educational Research*, 24, 376-380.
- Taylor, K., & Walton, S. (1998). *Children at the Center*. New Hampshire: Heinemann
- Wise, L.L., Harris, C.D., Brown, D.G., Becker, D.E., Sun, S., & Coumbe, K. (2003). California High School Exit Examination (CAHSEE): Year 4 evaluation report. (FR-03-64r). Alexandria, VA: Human Resources Research Organization.
- Wise, L.L., Harris, C.D., Koger, L.E., Bacci, E.D., Ford, J.P., Brown, D.G., Becker, D.E., Sun, S., Koger, M.E., Deatz, R.C., & Coumbe, K.L. (2004). Independent evaluation of the California High School Exit Examination (CAHSEE): Second biennial report. Working draft (DTR-04-01). Alexandria, VA: Human Resources Research Organization.
- Wolach, I., & Pratt, H., (2000). The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. *Clinical-Neurophysiology*, 112(1), 186-197.