The Relationship Between School Quality and the Probability of Passing Standards-Based High-Stakes Performance Assessments

CSE Technical Report 644

Pete Goldschmidt
Jose-Felipe Martinez-Fernandez
CRESST, UCLA

December, 2004

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Project 3.1 Methodologies for Assessing
Pete Goldschmidt, Project Director, CRESST/UCLA

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences, U.S. Department of Education.

THE RELATIONSHIP BETWEEN SCHOOL QUALITY AND THE PROBABILITY OF PASSING STANDARDS-BASED HIGH-STAKES PERFORMANCE ASSESSMENTS

Pete Goldschmidt and Jose-Felipe Martinez-Fernandez

CRESST/University of California, Los Angeles

Abstract

We examine whether school quality affects passing the California High School Exit Exam (CAHSEE), which is a standards-based high-stakes performance assessment. We use 3-level hierarchical logistic and linear models to examine student probabilities of passing the CAHSEE to take advantage of the availability of student, teacher, and school level data. The indicators of school quality are the Academic Performance Index (API) and magnet school status. The results indicate that both indicators of quality improve the probability of passing the CAHSEE, even after accounting for individual student characteristics. Also, the effect of opportunity-to-learn increases with school quality the relationship between school quality and the probability of passing standards-based high-stakes performance assessments.

Introduction

The continuing trend of school accountability at both state and federal levels places increasing emphasis on identifying how well schools are performing. States with existing accountability systems often attempt to tie these into current No Child Left Behind (NCLB) legislation in order to minimize confusion at the district and school level. NCLB legislation has brought to the forefront that all students need to be proficient and that schools need to facilitate student proficiency or be subject to, among other things, parents being allowed to move their children to schools that can effectively teach students. This naturally raises the question of whether or not information available to parents allows them to make an informed decision when choosing where to place their children. California established, prior to NCLB, an academic performance index (API) that explicitly creates a measure of school quality. This measure is used to produce school rankings where schools with higher API scores are theoretically better than those with lower scores. Another independent indicator commonly associated with school quality, is whether or not a school is a magnet school.[1] The assumption is that attending a higher quality school necessarily leads to better individual performance, but is this, in fact, the case? Simply examining this question using the standard assessments given to students confounds student performance with school performance if the same tests results are used to make inferences about both school quality and true student achievement, independently.

Using a separate assessment that is not included in the API school quality index, and is based on different standards, can shed some light on the role of school quality and how school quality might impact student achievement. Given that accountability no longer applies solely to schools, but to students as well, students have a vested interested in attending schools that will facilitate attainment of higher levels of achievement. For the particular case of this paper we are interested in the effect of school quality on the probability of passing the California High School Exit Exam (CAHSEE).

Under the over-arching question of what the relationship between school quality and the probability of passing standards-based high-stakes performance assessments is, we examine the following four issues: a) How much of the variation in CAHSEE scores is actually attributable to schools? Only that portion of the variation in CAHSEE scores that varies among schools can actually be accounted for by differences in school quality. b) For the average student, does school quality affect the probability of passing the high school exit exam? c) If the mechanism through which schools facilitate learning is through opportunity to learn, are there differences in how opportunities to learn translate into student performance between high and low quality schools? d) Are all students equally impacted by school quality?

Background

The API and the CAHSEE are two central components in the State's recent accountability-based efforts towards improvement in school effectiveness and student achievement.[2] The Academic Performance Index measures "the academic performance and growth of schools…A school's score on the 2001 API is an indicator of the school's

[1] Schools are not designated magnet schools based on API scores but usually exist due to the efforts of groups of teachers and parents. There is a general perception that magnet schools are higher quality than common schools as evidenced by the fact that parents usually need to undertake an extensive process in order to apply for their child to be admitted into a magnet school. There is a subset of magnet schools where student performance is a condition of acceptance, but for these the magnet's quality designation would not be wholly independent of the API.

[2] Both are either primary (assessments used to generate API scores are also used in NCLB) or secondary annual measurable objectives (AMOs) included in Adequate Yearly Progress (AYP) calculations.

performance level," (California Department of Education, 2003a). The API is also used to set goals for improvement that schools should meet every year as part of the state's accountability system. The calculations for the API include scores on the Stanford Achievement Test Version 9 (SAT-9) and the California Standards Tests. Although the API is intended to monitor school performance and present easily interpretable information to parents and education policy-makers, the use of unadjusted mean test scores[3] presents misleading results (Aitkin & Longford, 1986; Goldstein & Meyers, 1996). Also, ranking schools on the basis of means artificially attributes all of the variation in achievement to differences among schools (Raudenbush & Bryk, 2002), over-emphasizes the importance of student background characteristics because these characteristics are aggregated (Hanushek, 1986), and in general inadequately describes which differences in the index are actually statistically significant (Goldstein, 1997). Further, measures of school quality may not adequately capture other dimensions of school quality, especially processes occurring within schools (Baker, Goldschmidt, Martinez-Fernandez, & Swigert, 2002). Despite these limitations, this is the current measure of school quality to which parents have ready access.

The CAHSEE is intended to "significantly improve pupil achievement in public high schools and to ensure that pupils who graduate from public high schools can demonstrate grade level competency in reading, writing, and mathematics" (California Department of Education, 2003b). The test consists of extensive multiple choice and open-ended sections that are aligned to state performance standards. It was initially offered in 2001 and is planned to become a mandatory requirement for graduation in the 2005-2006 school year. Two key elements of the 2001 CAHSEE for the purpose of our study are that the results of this test were not yet included in the API calculations for that year—as mentioned before, after 2002 the CAHSEE is part of the API. Additionally, the CAHSEE matches standards thought important for high school graduates to meet, rather than the grade matched skills assessed by the SAT-9.

Hence, SAT-9 results summarize the current status of performance and how well students' performance compares to what students should have had an opportunity to learn up to 10th grade.[4] The CAHSEE, taken before the final year of high school, assesses how students are making progress towards passing fixed standards-based criteria. In general, SAT-9 results, summarized by the API[5], measure how well students know (or how well schools get students to know) what they ought to know in 10th grade, and this

---

[3] The API is a weighted mean—weighted by student subgroups and content areas.
[4] 9th grade for 20% of our sample.
[5] Five for each high school, of course, the API summarizes school performance based on 9th, 10th, and 11th grade student SAT-9 assessment results.

should in theory bear some relationship to how well students are progressing towards a final absolute criterion—the CAHSEE. The issue of interest in this study is whether school quality impacts the chances of a student passing the CAHSEE.

Given that students can only be held accountable for what they have had, an opportunity to learn (OTL) (Burstein & Winters, 1994), schools should be held accountable for both providing that opportunity and for ensuring that those opportunities produce the expected educational benefits. The effect of OTL is a more concrete mechanism through which we can examine the effects of school quality, beyond simply comparing mean performance differences between high and low quality schools.

Data and Methods

The data are from a large, racially integrated, urban school district in California and consist of approximately 45,000 mainly 10th grade students attending more than 200 schools (a small number of 9th graders were included in the 2001 sample). Descriptive statistics for the sample are presented in Table 1. The outcome measure of interest is the CAHSEE that, in the 2001-2002 school year, was not included in the California accountability system API computations. We have additional student test scores that include reading and mathematics SAT-9 scores. However, as noted, SAT-9 scores were two of the primary factors used in calculating schools' API scores, and for this reason we exclude these from the analysis. Additionally, student demographic information is available that allows us to adjust for enrollment factors typically considered outside the direct control of the school (e.g., gender, ethnicity, language proficiency[6], socioeconomic status, gifted and special education status). The days attended variable is used as a rough measure of opportunity to learn and represents the number of days attended for the fall semester.

The dataset also allows for the inclusion of variables that describe the educational context, both at the classroom and school levels of aggregation. The classroom dataset includes information on teacher qualifications and experience, as well as class size, mean achievement levels, and other aggregates of student enrollment. Another important aspect of school quality included is whether or not the student is attending a classroom that is a magnet school classroom. At the school level we are mostly interested in the API index and its relationship to the probability of passing the

[6] While it appears that this sample does not adequately represent English language learner (ELL) students, it is important to note that by 9th or 10th grade, the majority of ELL students will have been redesignated.

CAHSEE. Additional information is also available about the schools, including school size, aggregates of the socioeconomic characteristics of the students' families, and whether or not the school hosts a magnet school.

As we noted above, another indicator of perceived school quality is whether or not a school is a magnet school. Many magnet schools in this district are part of a host school. In fact, 42% of schools host a magnet school on their campus. Still, this only represents 10% of the students. Given that magnet schools are actually schools within schools, we code students as attending either the magnet school or the host non-magnet at the classroom level. This coupled with the fact that magnet schools do not receive separate API scores from their host schools, allows us to examine the effects of two different quality measures separately. This also allows us to examine whether there are any externalities accruing to non-magnet school students attending the host school.

Table 1

Sample Descriptive Statistics

	Statistics
Students (n=28,801)	
% Pass CAHSEE Language Arts	53%
Mean Score CAHSEE Language Arts	354.16
% Pass CAHSEE Mathematics	31%
Mean Score CAHSEE Mathematics	337.75
10th Grade	80%
9th Grade	20%
Female	51%
ELL	18%
RFEP (Redesignated fully English proficient)	43%
Special Education	3%
Low SES (Free Meal)	61%
Magnet	10%
Days Attended	77
Ethnicity	
American Indian	1%
Asian	5%
African American	11%
Hispanic	67%
White	13%
Other	3%
Classroom / Teacher (n=811)	
Class size	36.02

Table 1 (*continued*)

	Statistics
Magnet Classroom (School)	8%
Teaching Experience	
Less than 3	69%
3 to 10 Years	16%
More than 10	15%
Teacher Education	
5 or Less Years	75%
6 to 10	17%
11 or More	8%
School (n=43)	
Size (10th Grade, Incl. Magnet)	717.7
API	557.58
Average Parental Education	2.45
Hosts a Magnet School	42%

Methodology

The sample consists of a nested structure including data at the student, class, and school levels. Ignoring the multiple levels in the data can lead to substantive errors in inferences regarding estimated effects of variables at different levels (Burstein, 1980). The outcome of interest is dichotomous (passing the CAHSEE or not) so the method employed is a three-level multilevel logistic regression (Raudenbush & Bryk, 2002). In this model, the probability of the event, E, occurring for a student, conditional on the student's characteristics, is placed on a logarithmic odds ratio scale so that the model is defined as:

$$ln\ [P(\ \mathrm{p}ijk)/(1-\ P(\ \mathrm{p}ijk)] = \mathrm{p}_{0jk} + \mathrm{p}_{1jk}{}^{*}a_{1ijk} + \ldots + \mathrm{p}_{njk}{}^{*}a_{nijk} \qquad (1)$$

when the log odds of passing the CAHSEE for student i, in classroom j, in school k are a function of the school mean probability of passing (p_{0jk}) adjusted for effects ($\mathrm{p}_{1jk}\ldots\mathrm{p}_{njk}$) of student characteristics (a_{1ijk}). This is convenient to work with because, it is simply a linear sum of the explanatory variables. Equation 1 (the Level-1 or student model) is expanded to include classroom (Level-2) and school (Level-3) components:

Level-2 (Classroom) Model

$$p_{0jk} = b_{00k} + b_{01k}*X_{1jk} + \ldots + b_{0mk}*X_{mjk} + r_{0jk} \qquad (2)$$

The classroom's mean probability of passing the CAHSEE (p_{0jk}) after controlling for student level characteristics, is represented as a function of a mean classroom effect for a school, (b_{00k}) plus the effect ($b_{01k}\ldots b_{0mk}$) of independent classroom-level variables (Xs), and a component of random error (r_{0jk}). The random effect (r_{0jk}) captures the variables in mean passing rates among classrooms. Similarly, for the student level effects:

$$p_{n1jk} = b_{n0k} + b_{n1k}*X_{1jk} + \ldots + b_{nmk}*X_{mjk} + r_{njk} \qquad (3)$$

The classroom mean student effect (e.g., the mean classroom effect of gender) (p_{n1jk}) after controlling for student characteristics is represented as a function of a mean effect for a school (b_{n0k}) plus the effect ($b_{n1k}\ldots b_{nmk}$) of classroom variables (Xs) and a random error term (r_{njk}). The random effects (r_{njk}) capture the variation in student effects among classrooms. For example, it may be that girls have a different probability of passing the CAHSEE and that this difference varies significantly from classroom to classroom.

Level-3 (School) Model

$$b_{00k} = g_{000} + g_{001}*W_{1k} + \ldots + g_{00p}*W_{nk} + u_{0pk} \qquad (4)$$

where the mean CAHSEE passing rate for a school, adjusted for student and classroom characteristics, is a function of a grand mean passing rate for all schools (g_{000}) the effect ($g_{001}\ldots.g_{00p}$) of school specific characteristics (the W's), and a random error component (u_{0pk}). Unlike ordinary least squares regression coefficients (or single level logistic models), in multilevel models the parameters are subscripted by j and k, indicating that each classroom and school can have a different intercept and slope(s). The Level-1 and Level-2 coefficients can be specified as being either fixed, non-randomly varying, or randomly varying (Raudenbush & Bryk, 2002). A model with several student-level explanatory variables can have any combination of the three specifications.

The pass/no pass decision is based on an underlying score on the CAHSEE. We take advantage of having these scores available to counter the modeling limitations imposed by the logistic specification. In three-level data structures such as the one employed in this study, the number of additional parameters estimated grows very quickly as more predictors are added into the model, and this along with the increased computational needs of non-linear models soon overburdens the data. In particular, estimation of the variance components becomes infeasible for most or all of the student and classroom-level parameter slopes. In other words, we are not able to take full advantage of the random effects and cannot model how classroom and school context might affect student correlates of performance. As expected, estimation of the multilevel logistic models presented above became troublesome as more predictors were added. Only equation (1) needs to be altered in order to simplify the logistic model into a continuous model. Hence, for the continuous outcome Y, we use the following student (Level 1) model:

$$Y_{ijk} = p_{0jk} + p_{1jk}{}^{*}a_{1ijk} + \ldots + p_{njk}{}^{*}a_{nijk} + e_{ijk}, \text{ where } e \sim N(0,s^2).$$

Y_{ijk} is the CAHSEE score for student i in classroom j in school k. The ps are modeled as above (Equations 2-4). Using the continuous three level model also allows us to parse the variation into all three levels. The advantage is that we can directly estimate the extent to which CAHSEE scores' variation can be attributed to students, classrooms, and schools.

Results

The complete results for a three-level model tend to be complex, so we present results, building one level at a time. Hence, we begin with a completely random model without any predictors and subsequently examine the effects of student, classroom, and school variables sequentially.

The variance components in Table 2 reveal significant and considerable variation in the probabilities of passing the language and mathematics exams arising from classroom differences within schools. The results, in fact, suggest that between-classroom variability is at least as, or more important, than that observed between schools (Level-3). This provides support for one of the basic tenets of this study, namely that classroom attendance inside a school is a key determinant of performance on the CAHSEE, or student performance in general.

Table 2

Three-Level Multilevel Logistic Model—Unconditional Model Results

		CAHSEE Language				
Fixed Effect		Estimate	Standard Error	T-ratio	d.f.	*approx. p*
Baseline	G000	-0.31	0.15	-2.05	42	0.05
Random Effect		Standard Deviation	Variance	c2		
Classroom	R0	1.39	1.93	1815.93	775	0.00
School	U00	0.88	0.78	241.68	42	0.00
		CAHSEE Mathematics				
Fixed Effect		Estimate	Standard Error	T-ratio	d.f.	*approx. p*
Baseline	G000	-1.35	0.14	-9.88	42	0.00
Random Effect		Standard Deviation	Variance	c2		
Classroom	R0	1.16	1.35	1492.31	775	0.00
School	U00	0.81	0.65	239.35	42	0.00

The intercept presented in Table 2 can be interpreted as the unconditional, or baseline, log-odds, of passing the CAHSEE. The results indicate that the average expected log-odds of passing the CAHSEE for the students in the sample are -0.307 for Language, and –1.34 for Mathematics. Although it is mathematically convenient for modeling binary data, the log-odd scale is more intuitively interpretable if transformed to the familiar 0 to 1 scale for the probability of an event. The average probability of passing is given by the expression: exp(G000)/[1+ (G000)], which for Language Arts is {exp(-0.307)/[1+ exp(-0.307)]} = 0.42 indicating that less than half of the students in the sample pass the Language Arts exit exam. In Mathematics, the average probability of passing is considerably lower at {exp(-1.346)/[1+ exp(-1.346)]} = 0.21.

Table 3 presents the estimated parameters for the effects of student background characteristics on the probability of passing the CAHSEEs—parameters excluded from the table were not significant in the model. By centering the student predictors around their grand means, the intercepts in these models represent the log-odds of passing for the average student in the sample. The probability that an average student attending a typical school passes the CAHSEE in Language Arts can be obtained as {exp(-0.047)/[1+ exp(-0.047)]} = 0.48. Similarly, the probability of passing the Math CAHSEE for the average student is {exp(-1.35)/[1+ exp(-1.35)]} = 0.21. Note that these probabilities are adjusted for student characteristics and school environment, and therefore can be slightly different from the raw descriptive statistics presented in Tables 1 and 2 above.

Table 3
Three-Level Multilevel Logistic Model—Effect of Student Characteristics in Probability of Passing

	Effects of Covariates [1]	
	CAHSEE Language Arts	CAHSEE Mathematics
Intercept (γ_{000})	-0.047	-1.35.
Days Attended (OTL) (γ_{100})	0.013	0.02
ELL (γ_{200})	-2.40	-1.60
RFEP (γ_{300})	-0.32	-0.26
Ethnicity		
American Indian (γ_{400})	-	-
Asian (γ_{500})	-	0.61
African American (γ_{600})	-1.20	-1.39
Hispanic (γ_{700})	-0.50	-0.86
Other (γ_{800})	-0.42	-0.47
Special Education (γ_{900})	-3.03	-3.15
Low SES (γ_{1000})	-0.21	-0.16
Gender (γ_{1100})	0.36	-0.61
Grade (γ_{1200})	1.00	1.13

Note. (1) all covariables in this table with values are significant at $p < .05$.

The parameter estimates for the covariates represent the difference, in terms of log odds of passing, from the average student as represented by the intercept. Thus for example, the log odds of passing for Hispanic students are, on average, 0.50 lower in

Language Arts, and 0.86 lower in Math. Other things being equal, Hispanic students have on average a {exp(-0.047-0.50)/[1+ exp(-0.047-0.50)]} = 0.36 probability of passing the Language Arts CAHSEE. In other words, a Hispanic student who has the same background characteristics as an average student has an excepted probability of passing the CAHSEE that is 12 percentage points lower than the average student. It is important to note that this accounts for the potential effects of being an ELL student. Similarly, Hispanic students have only a {exp(-1.35-0.86)/[1+ exp(-1.35-0.86)]} = 0.10 probability of passing the Math CAHSEE on their first try, 11 percentage points less than the average. The results for the student covariates are consistent with expectations.

A rough first cut indication for OTL is grade, as students in higher grades have more opportunities to learn the content required for graduation. We would expect 10th grade students to have a greater probability of passing than 9th grade students. In the model, grade level was recoded as a binary variable in such a way that Grade 9 was assigned a 0, and Grade 10, a 1. The grade parameter in the model therefore represents the expected advantage in the log-odds of passing for students in the 10th grade. The results show 10th graders have, on average, a {exp(-0.047+1.00)/[1+ exp(-0.047+1.00)]}=0.72 probability of passing the Language Arts CAHSEE, and a {exp(-1.35+1.13)/[1+ exp(-1.35+1.13)]} = 0.44 probability of passing the Math CAHSEE. A second, more refined indicator of OTL is the number of days attended. The results displayed in Table 3 indicate that there is a statistically significant effect for attendance. In fact, there is a 10 percentage point difference in the chances of passing the CAHSEE (results are approximately the same for language arts and mathematics) between high attending and low attending students[7]. That is, among students in the same grade, students attending class more regularly have greater chances of passing the CAHSEE.

We next turn to our first indicator of school quality, magnet school status. As noted, we code magnet attendance at the classroom level because magnet schools are generally located within a host school and do not receive a separate API. The results in Table 4 present the unadjusted effects associated with attendance in a magnet classroom. The parameter estimates for magnets represent an advantage in terms of probability of passing for students attending magnet schools, other things being equal. On average, the log odds of passing the exam for students attending magnet schools are 1.29 higher in Language Arts, and 1.61 higher in Math. In terms of probabilities, these effects translate to a {exp(-0.30+1.21)/[1+ exp(-0.30+1.21)]} =0.71 probability of passing

[7] We consider students to be high attendees if they attend school one or more standard deviations above average. We consider students low attendees if they attend one or more standard deviations below average.

the Language Arts CAHSEE, (a 0.29 advantage), and {exp(-1.31+1.61)/[1+ exp(-1.31+1.61)]} = 0.57 probability of passing the Math CAHSEE (a 0.36 advantage).

Table 4
Three-Level Multilevel Logistic Model—Magnet Schools Predictors

CAHSEE Language

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	-0.31	0.15	-2.03	42	0.05
MAGNET	G010	1.29	0.24	5.30	816	0.00

CAHSEE Mathematics

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	-1.31	0.14	-9.43	42	0.00
MAGNET	G010	1.61	0.36	4.51	42	0.00

By way of comparison, we also briefly examine the unadjusted effect of school quality, as measured by the API. These results are displayed in Table 5. The API parameter estimates represent the increase in probability of passing, per 100 point increase on the 800-point API scale, holding all other factors constant. As an example, we take a hypothetical school with an API score 100 points higher than another school. The log odds of passing the exam for students attending the school with the higher API are 0.94 higher in Language Arts, and 0.89 higher in Math. This translates to a {exp(-0.26+0.94)/[1+ exp(-0.26+0.94)]} =0.66 probability of passing the Language Arts CAHSEE, a 23 percentage point advantage between the two. For Mathematics, the probability is {exp(-1.28+0.89)/[1+ exp(-1.28+0.89)]} = 0.40, a 19 percentage point advantage.

We next examine the combined effects of both of the school quality indicators, magnet status and API. These results are displayed in Table 6. The results indicate that students who attend schools with higher API's have a higher probability of passing the Language Arts and Mathematics Exit Exams, even after taking into account whether they attend magnet classrooms. The log odds of passing for a student in a school with an API 100 points higher than another are on average 0.93 higher in Language Arts, and 0.86 higher in Math, even after taking Magnet schools into account.

The results indicate that the effect of magnet schools remains considerable. On average, students who attend magnet classrooms have an advantage in their probability of passing the CAHSEEs, even if they attend host schools with similar APIs. Interestingly, the API of the host school does not have an effect on the advantage of magnet classrooms operating within it, as indicated by a statistically insignificant γ_{011} parameter. In other words, students attending a school that does not host a magnet school benefit from school quality as measured by the API; students who attend a school hosting magnet school also benefit from school quality as measured by the API, but to a lesser extent (the parameter estimates in Table 4 for the API effect on the mean probabilities are larger than the parameter estimates for API effect on the mean probabilities in Table 5). While the marginal effect of each API point is lower for these students, a host school is likely to have slightly higher API due to hosting a magnet, than it otherwise would if it did not host a magnet. Students attending magnet classrooms do not derive any additional benefit from the magnet class's host school being of better quality.

Table 5

Three-Level Multilevel Logistic Model—API Schools as Predictors

CAHSEE Language

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	-0.27	0.08	-3.33	41	0.00
API01	G001	0.01	0.00	9.94	41	0.00

CAHSEE Mathematics

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	-1.28	0.05	-24.94	41	0.00
API01	G001	0.01	0.00	14.41	41	0.00

Finally we combine all of the student, class, and school characteristics into a single three-level logistic model. Due to limitations noted above, all variables displayed in

Table 1 are not included in this analysis, but will be used subsequently in the continuous outcome models.

Table 6

Three-Level Multilevel Logistic Model—Magnet Schools and API as Predictors

CAHSEE Language

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	-0.30	0.08	-3.66	41	0.00
API01,	G001	0.01	0.00	9.65	41	0.00
Magnet,	G010	1.33	0.39	3.46	41	0.00
API01,	G011	0.00	0.00	0.38	41	0.71

CAHSEE Mathematics

Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Intercept,	G000	-1.30	0.05	-24.61	41	0.00
API01,	G001	0.01	0.00	13.72	41	0.00
Magnet,	G010	0.84	0.18	4.72	816	0.00
API01,	G011	0.00	0.00	1.54	816	0.12

The results in Tables 7 and 8 display the effects of the school quality indicators, controlling for the student covariates. That is, accounting for differences in student enrollment, does school quality have an effect on the probability of passing the CAHSEE? Although differences in student enrollment can account for some of the between-school variation in the probability of passing the CAHSEE, the marginal effects of magnet school status and the API remain statistically significant and positive. The major exception is that once student enrollment and school quality are taken into account, magnet school students no longer have a higher probability of passing the Math CAHSEE.

Another way to understand the effects of the API and magnet school status is presented in Figure 1. Figure 1 presents the probability of passing the Language Arts CAHSEE for different levels of school quality as measured by the API, as well as for magnet and non-magnet school students. Figure 2 presents the results for the Mathematics CAHSEE. There is only one line displayed because magnets and non-magnet students have equal chances of passing the Mathematics CAHSEE, ceteris paribus.

Table 7

Summary of Multilevel Logistic Model Results Language Arts CAHSEE (standard error in parentheses)

	Model 0 (ANOVA)	Model 1 (Student Covariates)	Model 2 (Magnet and API)	Model 3 (Slopes as outcomes. Fixed)
Intercept	-0.307	-0.080	-0.299	-0.300
Fixed effects				
Days Attended (OTL) (γ_{100})		0.02 (0.00)		0.02 (0.00)
ELL (γ_{200})		-2.44 (0.07)		-2.38 (0.07)
RFEP (γ_{300})		-0.30 (0.04)		-0.29 (0.04)
Ethnicity				
American Indian (γ_{400})		-		-
Asian (γ_{500})		-		0.34 (0.11)
African American (γ_{600})		-1.27 (0.08)		-1.05 (0.08)
Hispanic (γ_{700})		-0.60 (0.06)		-0.39 (0.06)
Other (γ_{800})		-0.38 (0.12)		-
Special Education (γ_{900})		-2.99 (0.33)		-2.65 (0.38)
Low SES (γ_{1000})		-0.16 (0.04)		-0.16 (0.04)
Gender (γ_{1100})		0.42 (0.03)		0.43 (0.03)
Magnet			1.33 (0.38)	0.74 (0.17)
API (100 Points)			0.93 (0.00)	0.6 (0.00)

Note. Only Significant parameters ($p < .05$) are presented.

Table 8

Summary of Multilevel Logistic Model Results (Mathematics CAHSEE)

	Model 0 (ANOVA)	Model 1 (Student Covariates)	Model 2 (Magnet and API)	Model 3 (Slopes as outcomes. Fixed)
Intercept	-1.31	-1.34	-1.29	-1.48
Fixed effects				
Days Attended (OTL) (γ_{100})		0.03 (0.00)		0.03 (0.00)
ELL (γ_{200})		-1.37 (0.08)		-1.62 (0.08)
RFEP (γ_{300})		-0.24 (0.04)		-0.24 (0.04)
Ethnicity				
American Indian (γ_{400})		-0.58 (0.29)		-
Asian (γ_{500})		0.60 (0.09)		0.86 (0.10)
African American (γ_{600})		-1.45 (0.07)		-1.30 (0.09)
Hispanic (γ_{700})		-0.94 (0.05)		-0.74 (0.06)
Other (γ_{800})		-0.46 (0.10)		-
Special Education (γ_{900})		-3.12 (0.46)		-
Low SES (γ_{1000})		-0.12 (0.04)		-0.09 (0.04)
Gender (γ_{1100})		-0.53 (0.03)		0.54 (0.03)
Magnet			0.83 (0.17)	-
API (100 Points)			0.86 (0.00)	0.80 (0.00)

Note. Only significant parameters ($p < .05$) are presented.

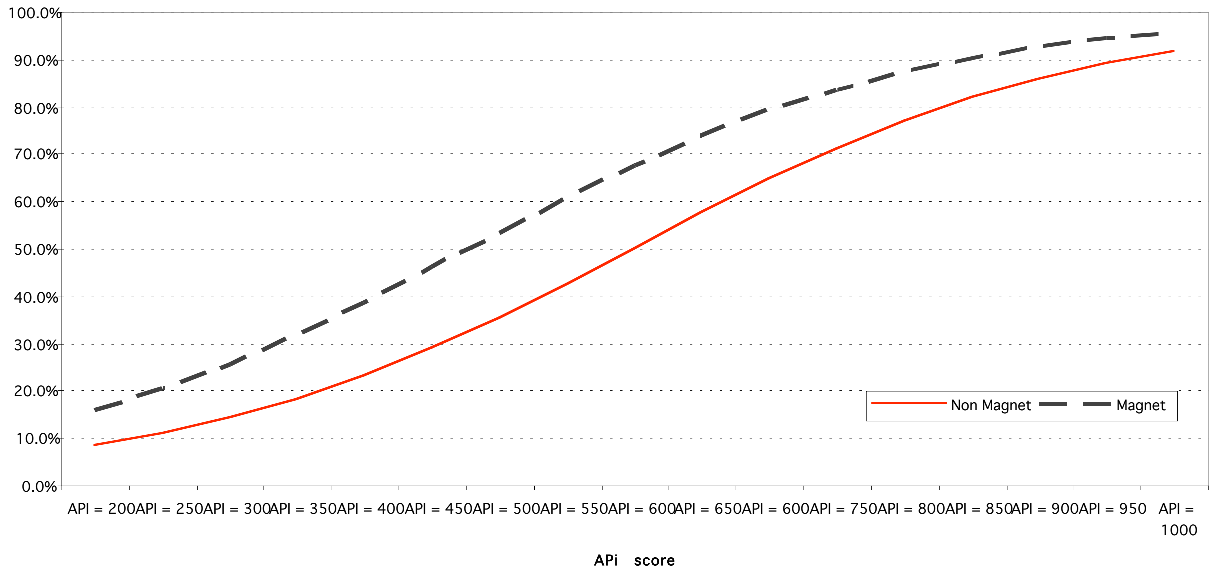

Figure 1. Probability of passing the Language Arts CAHSEE.

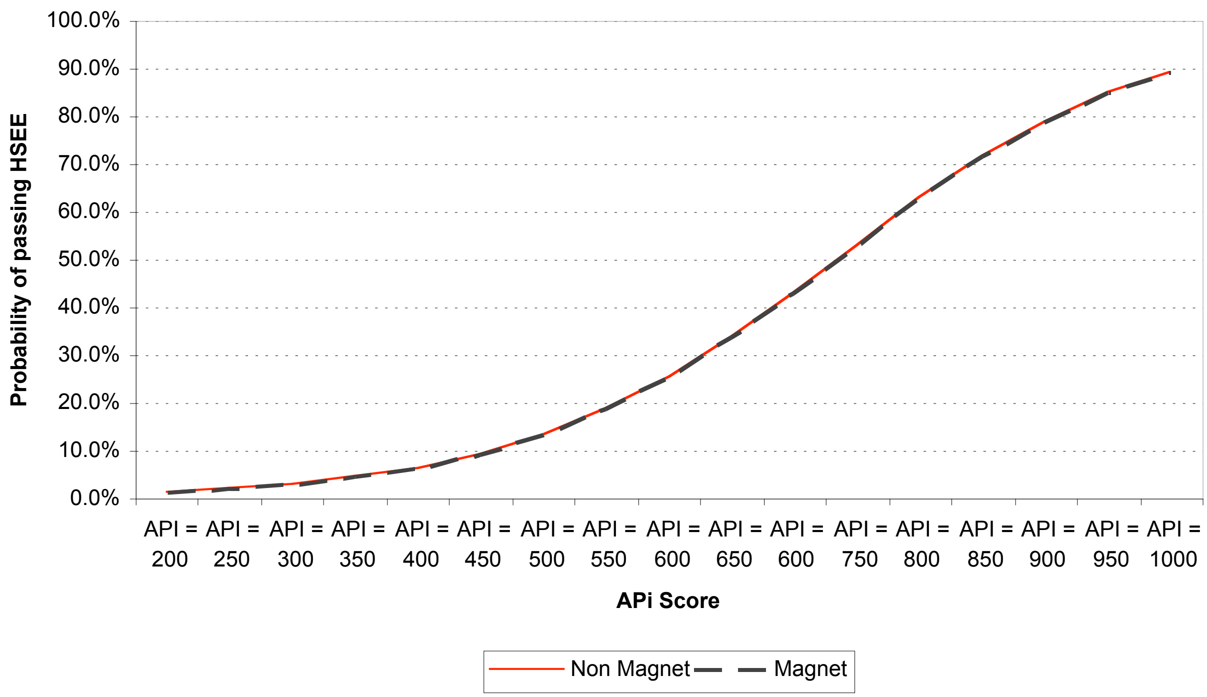

Figure 2. Probability of passing the Mathematics CAHSEE.

Turning to the role of OTL and school quality, we present Figures 3 through 6. Figures 3 and 5 compare the effect of attendance on the probability of passing the CASHEE. For Language Arts, there are increasing probabilities of passing the CAHSEE as attendance increases, and these probabilities increase further with school quality. This is also the case for mathematics, except to a larger extent. The extent to which differences in attendance matter by levels of school quality is displayed in Figures 4 and 6. Generally, for both Language Arts and Mathematics, the difference in the probability of passing between students with high and low attendance rates increases with school quality. That is, OTL has a more beneficial effect on passing the CAHSEE in better quality schools.

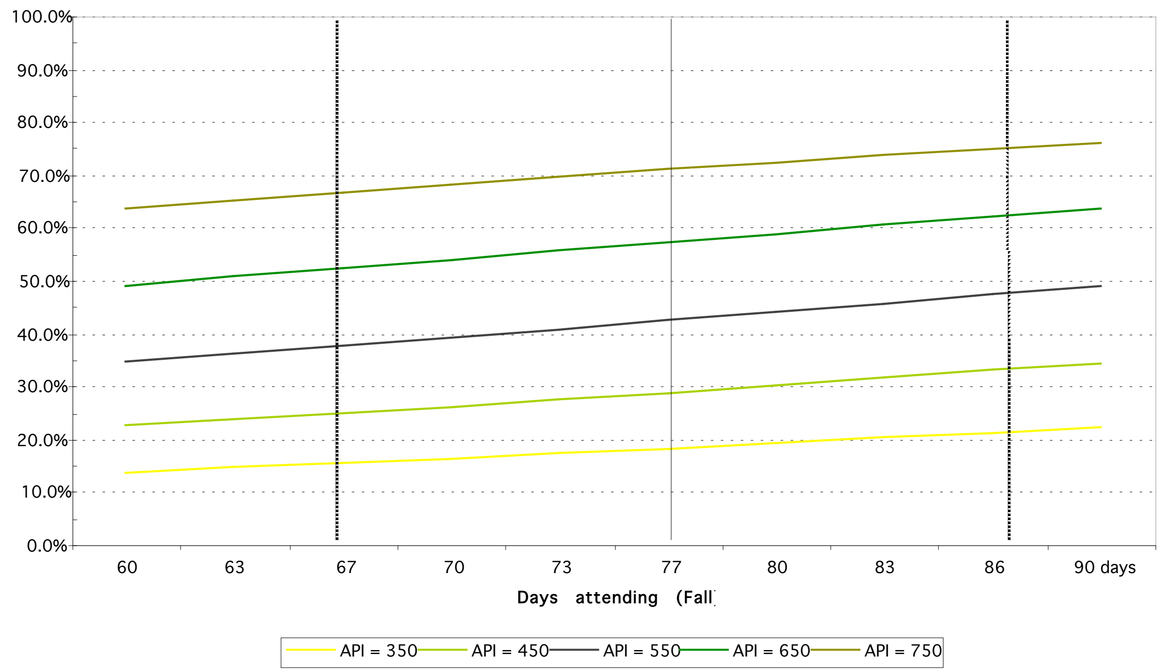

Figure 3. Probability of passing the Language Arts CAHSEE.

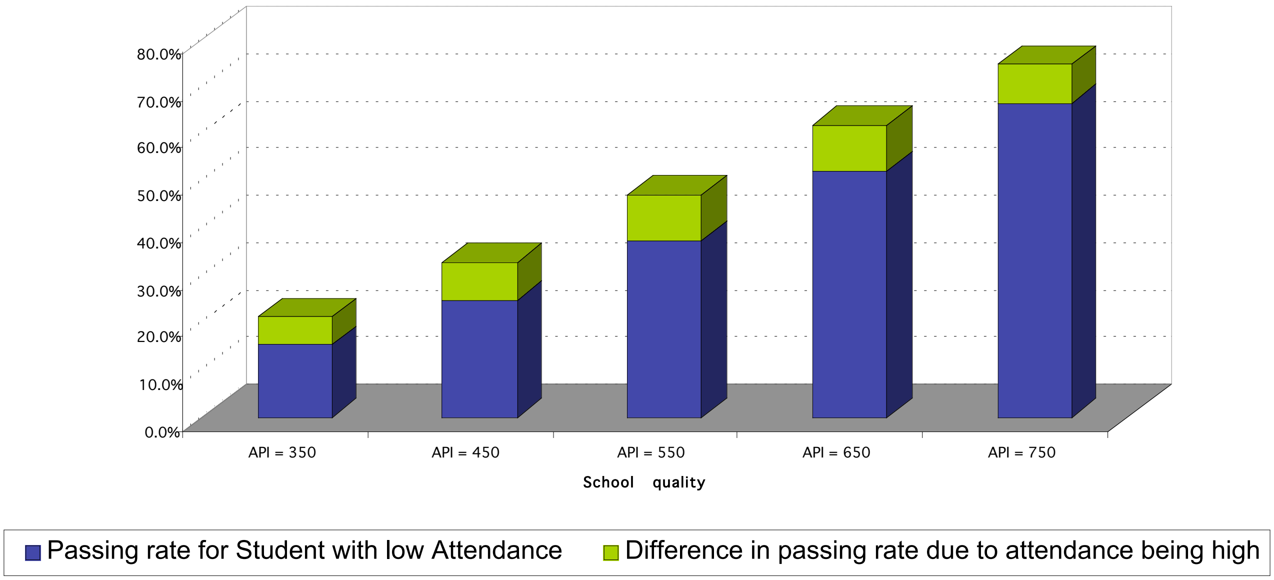

Figure 4. Marginal effect of OTL on passing rates by school quality - English Language Arts
CAHSEE.

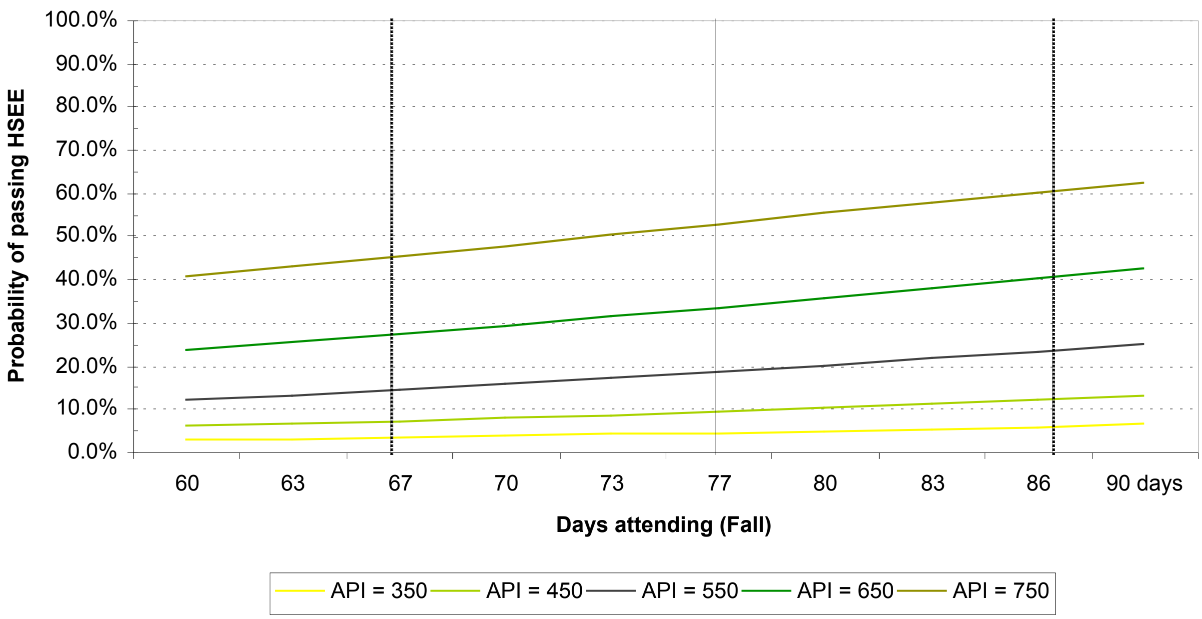

Figure 5. Probability of passing the Mathematics CAHSEE.

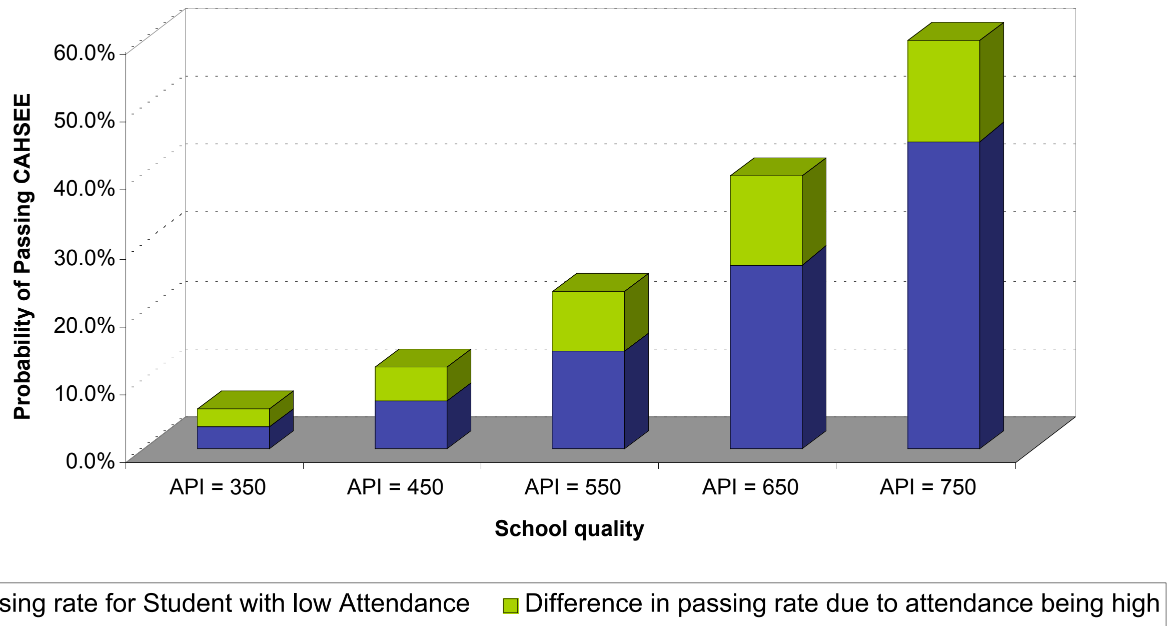

Figure 6. Marginal effect of OTL on passing rates by school quality - Mathematics CASHEE.

Continuous Model

The ANOVA unconditional models are presented first for reference and comparison to subsequent models. These models allow estimation of the proportion of variance at each of the three levels in the model—which is not possible with a logistic model. The results for both Language Arts and Mathematics, presented in Table 9, indicate most of the variability lies between the students within classrooms and schools (47.5% Language Arts and 58.5% Mathematics). At the second and third levels, the results reflect the same patterns observed in the logistic models, with a large classroom variance component, and a considerably smaller but still sizeable school component.

The results confirm that placing too much emphasis on differences between schools to account for differences in achievement is misleading. Variation among classrooms represents a larger portion of the variation in scores. For example, the mean Language Arts CAHSEE score in a classroom is about 345, but we would expect this to vary (95% CI) between 301 and 389. Given that we can parse the variation in CAHSEE scores into the appropriate level of aggregation in the model, we can examine whether magnet schools bring additional heterogeneity into schools. These results are displayed in Table 10. As expected, there is much greater variation attributable to classrooms in schools hosting magnet schools than to classrooms in schools not hosting magnet schools. Still the within-classroom variation is relatively similar between the two school types.

Table 9

Three-Level Multilevel Linear Model—Unconditional Model Results

CAHSEE Language						
Fixed Effect		Estimate	Standard Error	T-ratio	d.f	*approx p*
Baseline	G000	345.16	2.44	141.31	42	0.00
Random Effect		Standard Deviation	Variance	c2	d.f	*approx p*
School	U00	14.65	214.50	15.90	42	0.00
Classroom	R0	22.19	492.52	36.50	775	0.00
Student	E	25.29	639.43	47.50		
CAHSEE Mathematics						
Fixed Effect		Estimate	Standard Error	T-ratio		*approx p*
Baseline	G000	331.53	1.90	174.79	42	0.00
Random Effect		Standard Deviation	Variance	c2	d.f	*approx p*
School	U00	11.22	125.92	11.50	42	0.00
Classroom	R0	18.05	325.88	29.90	775	0.00
Student	E	25.25	637.44	58.50		

Table 10

Three-Level Unconditional Linear Model—Summary of Variance Components in Magnet and Non-magnet Schools

Random Effect	Variance Components (%)		
	Overall	Magnet Schools	Non-magnet
Reading			
Schools, (U00)	214.5 (15.9)	92.01 (6.46)	301.0 (20.6)
Classrooms, (R0)	492.5 (35.5)	667.8 (46.9)	522.3 (35.8)
Students, (E)	639.4 (47.5)	663.7 (46.6)	632.2 (43.4)
Math			
Schools, (U00)	125.9 (11.5)	78.39 (6.59)	157.7 (15.5)
Classrooms, (R0)	325.8 (29.9)	436.0 (36.6)	280.6 (27.6)
Students, (E)	637.4 (58.5)	674.1 (56.7)	576.0 (56.7)

Including student covariates at the student level explains about 13% of the student-level variance in the Language Arts CAHSEE and about 11% of the variance in Mathematics. The values of the fixed coefficients for the student covariates are presented in the Appendix. In general, the student level results based on a continuous outcome are not substantively different from those presented earlier for the logistic models. The student covariates have effects consistent with the literature; including large disadvantages for African American and Hispanic students, as well as English learners and special education students. Female students perform better than their male counterparts in Language, while the opposite is true in Mathematics. As before, OTL has a significant impact on the chances of passing both exit exams.

The results displayed in Tables 11 and 12 summarize the cross-level interaction effects of classrooms and schools on both the mean CAHSEE scores and specific student covariate effects. That is, we can determine whether school quality has a direct impact on specific performance gaps. The first step in determining whether student covariates might be affected by classroom or school contexts is to determine the extent to which the parameter estimates vary among both classrooms and schools. The results in Tables A1 and A2 indicate that each of the estimated parameters for all of the student background characteristics vary significantly among classrooms, within schools. For example, the average performance gap between ELL and non-ELL students on the Language Arts CAHSEE (Model 4 in Table A1) is –17 points. This performance gap between ELL and non-ELL varies among classrooms, however. In fact, comparing classrooms where this performance gap is one standard deviation above average with a classroom where it is one standard deviation below average demonstrates that these classrooms differ by about 14.6 points in the gap (-17+7.3-(-17-.7.3) = 7.3+7.3 =14.6).

The effectiveness of OTL varies significantly among classrooms as well. The difference, calculated as above, in the effect of OTL is .46 points per day attended. This translates into about a 9-point difference at the end of the semester between a high and low attending student. Given the substantively and statistically significant variation in student level effects, we next examine whether class or school context can account for any of this variation. The context variables we focus on, of course, are the school quality indicators.

Consistent with the logistic models, the effect of school quality (both magnet and API) exerts considerable positive influence on the likelihood that students will pass the exit exams, even after taking student background characteristics into account. However, these school indicators have only minimal impact on the student level effects. The results are presented in Tables 11 and 12. The results in Table 11, pertaining to Language Arts, indicate that the benefit of attending magnet schools accrues to all students equally. That is, a student attending a magnet school is expected to score about 13.5 points higher, ceteris paribus. But the Language Arts achievement gaps that exist are not diminished in magnet classrooms. The results in Table 11 indicate that school quality, as measured by the API, does not improve student performance equally. For example, a non-ELL student is expected to score about 18 points better if he attends a high quality school.[8] In contrast, there is only a 5.6 ((8.9-6.09)x2) point benefit accruing to ELL students. This means that although an ELL student benefits from attending a higher quality school, the gap between non-ELL and ELL probabilities of passing the CAHSEE increases.

The results indicate that OTL generates a greater benefit in magnet schools. On average, each day attended in a magnet classroom translates into a 0.22 point gain more per day than in a non-magnet classroom. There is no additional direct effect due to the API. Figure 7 displays this result graphically. The figure highlights that a student with

---

[8] We define a high quality school as one that has an API 1 standard deviation above average and a low quality school one that has an API 1 standard deviation below average. The standard deviation for the API is about 100 points.

low attendance attending a magnet school is expected to out-perform a student with high attendance not attending a magnet classroom.

Table 11

Summary of Level-2 and Level-3 Effects (Language Arts CAHSEE-Model 4)

	Classroom Predictors				School Predictors	
	Magnet	**Class Size**	**Teacher Ed. (6-10)**	**Teacher Ed. (10+)**	**Average Parent Ed.**	**API (100 Pts)**
Intercept	13.54 (2.44)	0.11 (0.01)	-	6.63 (2.61)	-	8.9 (0.02)
Level-1 Parameter						
Days Attended (OTL) (γ_{100})	0.22 (0.10)					-
ELL(γ_{200})	-					-6.09 (0.02)
RFEP (γ_{300})	-					-3.32 (0.01)
American Indian (γ_{400})	-					-
Asian (γ_{500})	-					-5.45 (0.02)
African American (γ_{600})	-					-
Hispanic (γ_{700})	-					-
Other (γ_{800})	-					-
Special Education (γ_{900})	-					-
Low SES (γ_{1000})	-					-
Gender (γ_{1100})	-					

The effects for Mathematics are presented in Table 12 and are generally the same as for Language Arts, except for the role of school quality on OTL. As with Language Arts, each day of attendance generates additional performance benefits above those accruing to non-magnet students. However, school quality, as measured by the API also has a direct effect on OTL. That is, on the Mathematics CAHSEE students attending higher quality schools benefit more in terms of improved performance for each day attended than students attending low quality schools.

The results in Figure 8 are similar to those in Figure 7, except that high-attendance students in non-magnet classrooms are expected to outperform low-attendance students in magnet classrooms. This occurs as a result of the direct impact school quality has on the effect of attendance. That is, general school quality can mitigate the effect of not being a magnet school student when the school hosts a magnet school.

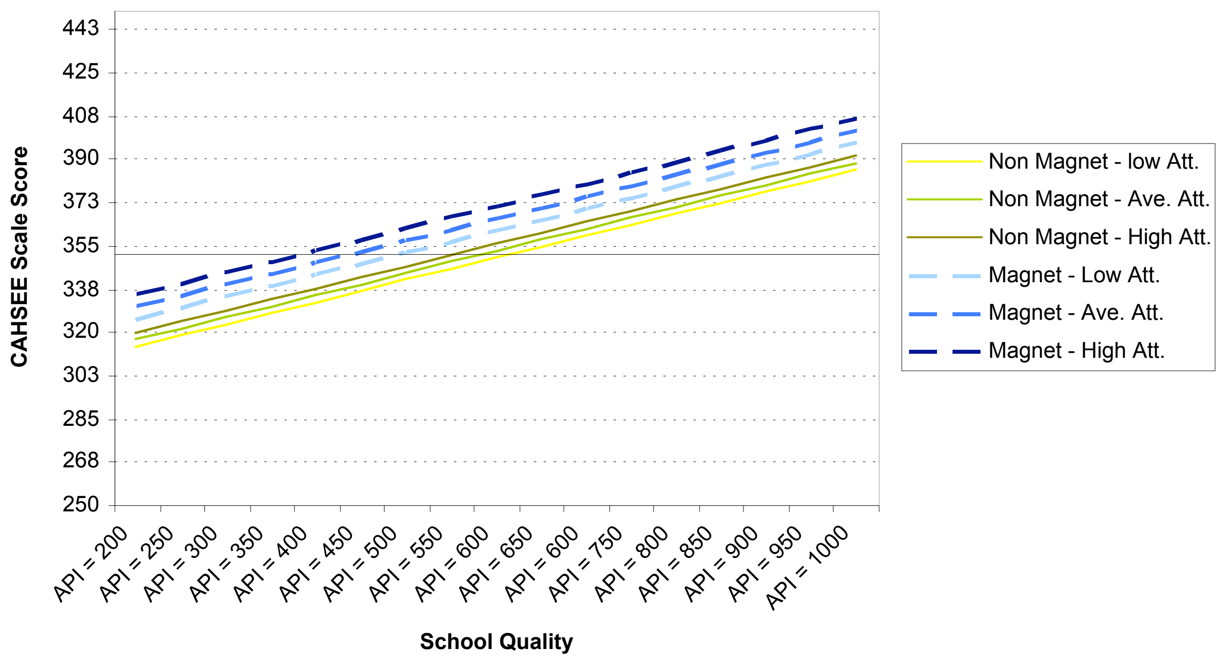

Figure 7. Effect of school quality on the benefit of OTL - Language Arts.

Table 12

Summary of Level-2 and Level-3 Effects (Mathematics CAHSEE—Model 4)

	Classroom Predictors				School Predictors	
	Magnet	**Class Size**	**Teacher Ed. (6-10)**	**Teacher Ed. (10+)**	**Average Parent Ed.**	**API (100 Pts)**
Intercept	10.10 (1.98)	0.07 (0.01)	-	-	-	8.4 (0.01)
Level-1 Parameter						
Days Attended (OTL) (γ_{100})	0.34 (0.11)					0.08 (0.00)
ELL (γ_{200})	-7.45 (2.9)					-5.94 (0.01)
RFEP (γ_{300})	-					-
American Indian (γ_{400})	-					-
Asian (γ_{500})	-					-
African American (γ_{600})	-4.27 (2.00)					-
Hispanic (γ_{700})	-					-
Other (γ_{800})	-					-
Special Education (γ_{900})	-13.42 (6.46)					-
Low SES (γ_{1000})	-5.53 (1.44)					-
Gender (γ_{1100})	-					-2.40 (0.00)

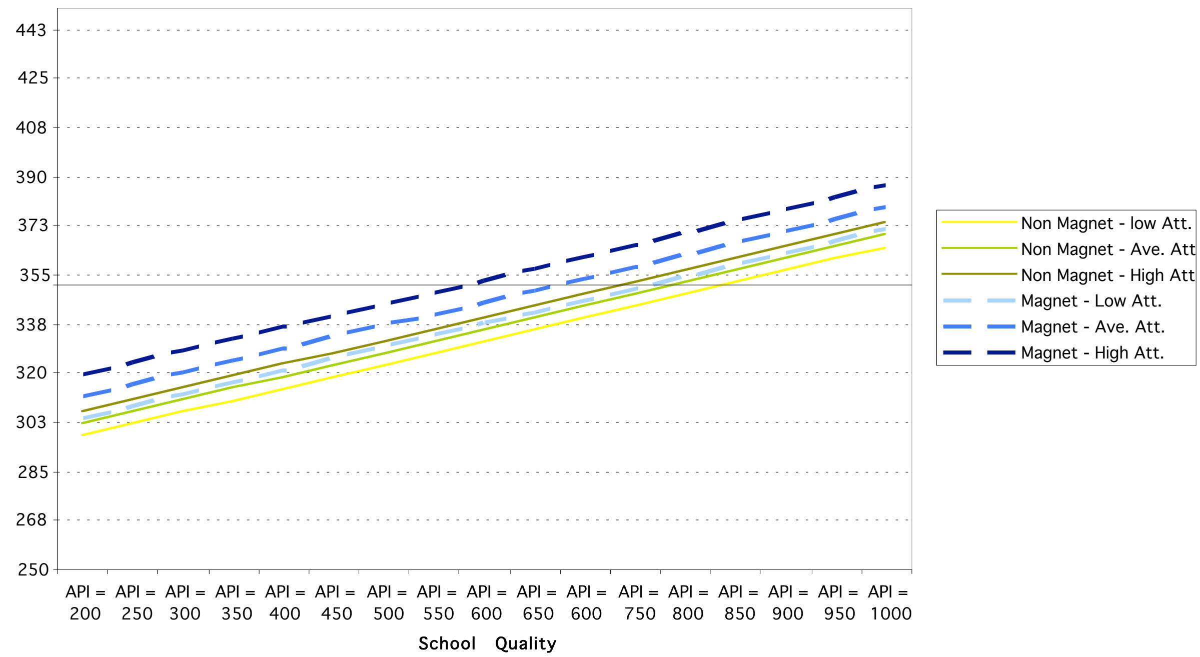

Figure 8. Effect of school quality on the benefit of OTL – Mathematics.

Discussion

Within the current climate of accountability—both student and school, it is increasingly important to identify whether available information regarding schools can lead to beneficial decisions by parents for their children. If students are to be held accountable for passing a standards-based performance assessment, then schools need to be held accountable for enabling students to pass such an exam. But holding schools accountable necessarily implies that schools vary in the success with which they can facilitate student progress towards successfully completing the high school exit exam

The evidence demonstrates that schools do, indeed, vary in terms of the passing rates of their students, but that a greater amount of the variation in passing rates is attributable to classrooms and students within classrooms than to schools. This means that policy aimed at alleviating differences among schools (whether measured in this analysis or not), will only marginally impact differences in student passing rates. However, school quality as measured by the API does increase the mean passing rates of students in a school. This means that if schools are failing students, then parents can use the API to choose a high school that will provide a better opportunity to pass the exit exam. Unfortunately, school quality is not synonymous with egalitarianism. While all students benefit from attending a better quality school some traditionally disadvantaged students, such as ELL, RFEP, and girls (in the case of mathematics), do not benefit as much as their classmates which results in larger performance gaps between groups.

School quality information is also conveyed by whether or not a student attends a magnet school (or magnet classroom within a school). Magnet school students are expected to perform better, and this benefit is independent of the host school's quality. Further, there is a positive externality accruing to students attending a school that hosts a magnet school, even if that student is not attending a magnet classroom.

An important aspect of school quality that needs to be considered is the logical extent to which school processes affect student outcomes—in this case it is how well schools can translate OTL into performance gains. We would naturally expect higher quality schools to be so, precisely because they can more effectively translate OTL into performance gains. But if the school quality index we used merely represents an aggregate of student quality, then school quality would have no effect on OTL. In fact, the evidence indicates that school quality does impact how effectively schools translate opportunities to learn into performance success.

The results also highlight the extent to which classrooms, and by extension teachers, play a role in student achievement gains—more variation exists between-classrooms within schools than between schools suggesting that educational experiences in the classroom are the most important determinant for student academic performance (and, consequently, success in the CAHSEE). However, more research is warranted with more classroom specific context variables, teacher characteristics variables, and specific measures of OTL.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in School Effectiveness Studies. *Journal of the Royal Statistical Society, A*, *149*(1), pp. 1-43.

Baker, E. L., Goldschmidt, P., Martinez, F., & Swigert, S. (2002). *In search of school quality and accountability: Moving beyond the California Academic Performance Index (API).* National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles. U.S. Department of Education Office of Educational Research and Improvement Award #R305B60002.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. C. Berliner (Ed.), *Review of Research in Education* (Vol. 8, pp. 158-233). Washington, DC: American Educational Research Association.

Burstein, L., & Winters, L. (1994). *Models for Collecting and Using Data on Opportunity to learn at the State Level: OTL Options for the CCSSO SCASS Science Assessment.* Paper presented at the National Conference on Large-Scale Assessment, Albuquerque, NM.

California Department of Education. (2003a). *Academic Performance Index (API).* Retrieved August 6, 2003, from the CDE Website: http://www.cde.ca.gov/psaa/api/

California Department of Education. (2003b). *California High School Exit Examination (CAHSEE).* Retrieved August 6, 2003, from the CDE Website: http://www.cde.ca.gov/statetests/cahsee/background/info.html

Goldstein, H. (1997a, March). Value added tables: The less-than-holy-grail. *Managing Schools Today*, (*6*), 18-19.

Goldstein, H., & Meyers, K. (1996). Freedom of information: towards a code of ethics for performance indicators. *Research Intelligence, 57*, 12-16

Hanushek, E. A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature, 24*, 1141-1177.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425. (2002). Retrieved October 31, 2002, from http://www.ed.gov/legislation/ESEA02/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

APPENDIX

Student level results of three-level continuous models

Table A1

Summary of Multilevel Linear Model Results (Language Arts CAHSEE)

	Model 0 (ANOVA)	**Model 1 (Student Covariates)**	**Model 2 (Magnet and API)**	**Model 3 (Slopes as outcomes. Fixed)**	**Model 4 (Slopes as outcomes. Random)**
Intercept	345.15	348.57	345.51	348.01	348.03
Fixed effects					
Days Attended (OTL) (γ_{100})		0.29		0.29	0.28
		(0.02)		(0.01)	(0.01)
ELL (γ_{200})		-27.7		-27.0	-27.99
		(0.80)		(0.66)	(0.79)
RFEP (γ_{300})		-5.38		-5.45	-5.33
		(0.44)		(0.45)	(0.46)
Ethnicity					
American Indian (γ_{400})		-		-	-
Asian (γ_{500})		-		4.62	3.64
				(1.05)	(1.12)
African American (γ_{600})		-16.54		-16.7	-15.45
		(1.03)		(0.90)	(1.23)
Hispanic (γ_{700})		-7.54		-7.26	-6.64
		(0.95)		(0.65)	(0.98)
Other (γ_{800})		-5.09		-	-
		(1.14)			
Special Education (γ_{900})		-28.09		-22.3	-24.66
		(1.41)		(1.83)	(1.51)
Low SES (γ_{1000})		-2.40		-2.62	-2.37
		(0.66)		(0.41)	(0.53)
Gender (γ_{1100})		5.76		5.95	5.85
		(0.31)		(0.32)	(0.32)
Magnet			20.3	13.5	17.9
Table A1 (*continued*)			(3.34)	(2.44)	(4.08)
API (100 Points)			14.9	8.80	8.9
			(1.49)	(0.02)	(0.02))
Random effects (Std Dev)					

	Model 0 (ANOVA)	Model 1 (Student Covariates)	Model 2 (Magnet and API)	Model 3 (Slopes as outcomes. Fixed)	Model 4 (Slopes as outcomes. Random)
Classrooms					
Days Attended (OTL) (R_{100})		0.18			-
ELL (R_{200})		8.88			8.92
African American (R_{600})		7.55			7.41
Hispanic (R_{700})		5.21			5.14
Special Education (R_{900})		7.36			6.15
Low SES (R_{1000})		1.65			1.71
Schools					
Days Attended (OTL) (R_{100})		-			-
ELL (R_{200})		-			-
African American (R_{600})		-			-
Hispanic (R_{700})		3.74			3.45
Special Education (R_{900})		-			-
Low SES (R_{1000})		3.24			1.93
Model Fit					
Deviance	213577	209466	213477	209650	209318

Note. Only Significant parameters (p <.05) are presented

Table A2

Summary of Multilevel Linear Model Results (Mathematics CAHSEE)

	Model 0 (ANOVA)	**Model 1** (Student Covariates)	**Model 2** (Magnet and API)	**Model 3** (Slopes as outcomes. Fixed)	**Model 4** (Slopes as outcomes. Random)
Intercept	331.5	332.7	332.0	332.75	331.98
Fixed effects					
Days Attended (OTL) (γ_{100})		0.33		0.29	0.35
		(0.02)		(0.01)	(0.02)
ELL (γ_{200})		-16.9		-16.3	-17.1
		(0.71)		(0.66)	(0.72)
RFEP (γ_{300})		-2.23		-2.64	-2.41
		(0.44)		(0.45)	(0.46)
Ethnicity					
American Indian (γ_{400})		-		-	-
Asian (γ_{500})		9.05		12.6	10.5
		(0.86)		(1.05)	(1.13)
African American (γ_{600})		-18.69		-19.0	-17.4
		(1.18)		(0.90)	(1.29)
Hispanic (γ_{700})		-11.11		-10.73	-9.61
		(1.02)		(0.65)	(1.01)
Other (γ_{800})		-5.87		-	-
		(1.15)			
Special Education (γ_{900})		-21.91		-19.3	-18.8
		(1.38)		(1.77)	(1.49)
Low SES (γ_{1000})		-		-1.16	-
				(0.41)	
Gender (γ_{1100})		-6.28		-6.40	-6.47
		(0.31)		(0.32)	(0.32)
Magnet			14.6	12.8	10.7
			(2.57)	(2.23)	(1.98)
API (100 Points)			12.1	8.81	8.08
			(0.09)	(0.01)	(0.01)
Random effects (Std Dev)					
Classrooms					
Days Attended (OTL) (R_{100})		0.27			0.23

Table A2 (*continued*)

	Model 0 (ANOVA)	**Model 1** (Student Covariates)	**Model 2** (Magnet and API)	**Model 3** (Slopes as outcomes. Fixed)	**Model 4** (Slopes as outcomes. Random)
ELL (R_{200})		7.24			7.32
African American (R_{600})		9.64			9.64
Hispanic (R_{700})		7.44			7.38
Special Education (R_{900})		6.04			-
Low SES (R_{1000})		2.47			2.12
Schools					
Days Attended (OTL) (R_{100})		-			-
ELL (R_{200})		-			-
African American (R_{600})		-			-
Hispanic (R_{700})		3.91			3.14
Special Education (R_{900})		-			-
Low SES (R_{1000})		2.46			1.99
Model Fit					
Deviance	213226	209313	213103	209608	208635

Note. Only Significant parameters (p <.05) are presented