School Accountability in California: An Early Evaluation

Julian R. Betts

University of California, San Diego, and Public Policy Institute of California

Anne Danenberg

Public Policy Institute of California

July 2001

Prepared for the Brookings Institution's conference "Accountability and Its Consequences for Students: Are Children Hurt or Helped by Standards-Based Reforms?", May 15-16, 2001. We thank Robert Hauser and Robert Rothman for helpful comments.

Abstract

We assess the three main components of California's recently adopted school accountability program. First, new content standards are highly specific and comprehensive. Second, the state has introduced a new norm-referenced statewide test, is phasing in criterion-referenced tests aligned with the content standards, and has introduced a high school exit exam which the graduation class of 2004 must pass. Third, California has implemented a complex series of rewards and punishments for school staff and students, including financial rewards to schools and teachers at schools that improve markedly, and scholarships to students who excel on the state test or Advanced Placement or equivalent tests. Notably, many of the financial incentives are aimed particularly at schools that initially rank poorly in student achievement. However, it is too soon to know how these incentives will affect both average student performance and inequality in student performance in California in the long run.

We analyze recent trends in both the level and distribution of test scores and of school resources. We find that average student achievement in California has risen markedly over a three-year period. Schools that initially had the lowest test scores appear to have improved the most, although some of this could result from growing student and teacher familiarity with the state test. We find evidence that some of this improvement merely reflects volatility in test scores and "regression to the mean," but alternative methods confirm that inequality in student achievement in California is falling.

Trends in school resources are less reassuring. We find little evidence that the high school curriculum is being diluted by the new emphasis on preparing for tests. However, overall in California, teacher education, experience, and credentials have fallen over the three-year period since the accountability reforms began. More troubling, schools that originally scored in the bottom fifth in student achievement have experienced a far greater decline in teacher preparation than have schools originally in the top fifth of student achievement. We hypothesize that teachers' concerns about the sanctions that await bottom-performing schools that fail to improve may have contributed to these adverse patterns in teacher mobility. Consequently, the state may have to redouble its efforts to improve teacher preparation and overall resources at schools with low achievement.

Introduction

Over the last decade, virtually every state has launched a school accountability program in light of public concern over failing schools. A state accountability program ideally consists of at least three components: a content standard or framework that stipulates what students should know and when they should know it, an assessment system that tracks student progress against the content standards, and a menu of responses by the state. These responses are typically directed at schools that excel and those that lag behind, but sometimes also allocate rewards for students who excel and extra educational resources for students who are struggling.

Assessments and accountability are not new. Indeed, as Linn (2000) notes, they have been included in many education reform programs over the last 50 years. He also points out that what *is* new in the current reform effort is the emphasis on content standards, setting challenging content standards while including all students, and the element of high-stakes accountability. [1]

Proponents of school accountability argue that standards can energize school systems, by putting pressure on schools to help *all* students. The public release of tests can bring public attention to bear upon schools that lag behind. A carefully designed system of consequences related to state assessments can in turn focus additional resources and oversight on those schools that are struggling the most. After almost a decade of accountability reforms, content standards and testing remain very popular in opinion polls.

At the same time, detractors of the standards movement point to a number of failings of typical accountability programs. First, opponents charge, the system of incentives and consequences can be unfair if it does not take into account the predominant role of family socioeconomic status in determining student achievement. A system that channeled all the financial rewards to schools in the suburbs, while imposing financial sanctions on inner-city schools, could be quite unfair. The "opportunity-to-learn" movement claims that holding all schools accountable to the same content standards makes sense only if schools serving disadvantaged students first receive an infusion of resources. Second, the assessment system must be closely aligned with the content standards. Otherwise, teachers and students will receive mixed messages about what it is students are expected to learn during the course of a year. Third, many critics worry that states' accountability systems will lead to significant changes both in school resources and the curriculum that may not benefit all students. For instance, Betts and Costrell (2001) note that in some states the most vociferous complaints about standards and testing have come from well-to-do suburbs, apparently on the grounds that state accountability systems interfere with schools that were already meeting affluent parents' goals for their children.[2] Inner city parents in the same cities often strongly support standards and testing.

In this paper we will address these issues in the context of California public schools. California has recently implemented an accountability system that encompasses the three fundamental components listed above: content standards, student assessment and a system of responses for schools at both ends of the achievement spectrum. We will pay particular attention to trends in overall student achievement and school resources and the *distribution* of student achievement and school resources since the introduction of the accountability system. We call our analysis an "early evaluation" because it is far too soon to know whether the new system will succeed. However, a retrospective of the accountability reforms a decade or more from now, while valuable, will be of no use to policymakers who are quite literally changing the accountability recipe on a year-by-year basis. It is essential that neutral outside observers monitor early trends in achievement and school resources to inform this policy debate.

A Summary of the Challenges Facing California Schools Prior to the Introduction of the Public School Accountability Act

The creation of a full-blown accountability system in California occurred in several stages during the late 1990's, with several key components preceding the capstone legislation, the Public School Accountability Act of 1999. A new statewide test was implemented in spring 1998. In this section we set the stage by discussing the level of school resources and initial student achievement before the accountability system was fully put into legislation in 1999. This summary provides the context in which we can assess the challenges facing the policymakers who pieced together California's accountability system in the closing years of the 1990's.

A first important fact to realize is that California's schools and students differ from those in the rest of the country in a number of dimensions. Each of these variations arguably heighten both the need for uniform educational standards *and* the risks of imposing them. First, California school expenditures have grown over time but beginning in the 1980's began to lag behind national averages. Figure 1 illustrates this trend. Sonstelie, Brunner and Ardon (2000) argue that a combination of legislative responses to the *Serrano v. Priest* school finance equalization case and the 1979 passage of Proposition 13, a voter initiative that effectively removed the ability of local authorities to raise property taxes for local needs, caused the growth in school spending in California to begin to trail the national average.[3] This suggests that standards in California, if they were pitched too high, might fail because of the relatively low levels of resources devoted to public education in California.

Second, California's student population differs from that in the rest of the nation. The state has a far higher number of Limited English Proficient (LEP) students than other states. This heterogeneous population implies that performance on standardized tests will be equally heterogeneous. Figure 2 shows the percentage of California's public school students scoring at or above national norms in math during the first (spring 1998) administration of the new state test. The middle line shows that in most grades the percentage of *all* students at or above the national median in math was in the low forties. (If California students were identical to those in the nation as a whole, then exactly 50% should have been at or above national norms.) However, the other two lines in the figure show that LEP and non-LEP students scored quite differently from each other at each grade. Indeed, non-LEP students in California still lagged behind national norms but only by a few percentage points, while LEP students were far below national norms. Clearly, limited experience with English in turn poses considerable language and academic challenges for many of California's students. (The gap between LEP and non-LEP students in reading is, understandably, even larger than the gap in math.)

The considerable gap in achievement between fluent English proficient and LEP students in California increases the importance of setting educational standards in a bid to bring up the achievement of California's LEP students. At the same time, this heterogeneity poses serious risks for the integrity of educational standards.

Which of these two variations from national averages, in spending per pupil or in students' ability in the English language, matters more in the design of educational standards? While both are potential landmines, we believe that variations in language pose the greater difficulty. If California's educational standards did not take into account variations in the number of LEP students in each school, the system could in theory unfairly penalize schools that served unusually large numbers of English learners.

Lower spending per pupil in California probably also drives a wedge between average student achievement in California and other states. This could lead to perennial underperformance of California's schools. However, it is important not to overestimate the size of the impact of spending and student outcomes. See Hanushek (1996) for a review of the limited impact of school resources on student test scores and Betts (1996) for a review of the link between school resources and years of education obtained and earnings after leaving school.[4] In the California case, Betts, Rueben and Danenberg (2000) find that variations in school resources explain only a small portion of the variation in achievement among the state's schools that existed in 1998.[5] Throughout this chapter, the school resources to which we refer are school-level teacher characteristics as measured by experience, education, and credentials; average class size; and high-school curriculum as measured by the percentage of college prep class sections across all subject areas. Although per-pupil spending is often the measure of resources that researchers use, in California the lowest available level of per-pupil spending is at the district level. Another reason to prefer our non-financial measures of school resources is that variations in spending per pupil to some extent mirror variations in the cost of living across the state.

Setting aside differences between California and the rest of the nation, there exists a remarkable variation in both test scores and certain school resources *among* California's schools. For example, Figure 3 shows that schools, when ranked into quintiles by the percentage of students receiving free/reduced-price lunch, varied radically in student achievement in spring 1998. The figure shows the percentage of *non-LEP* students scoring at or above national norms among schools in each quintile. Clearly, even setting aside the issue of LEP students, disadvantage is strongly related to variations in student performance across schools.

Variations in teacher characteristics and, at the high school level, curriculum, also appear among California's schools. Table 1 illustrates using data from California high schools. Teachers without full credentials, inexperienced teachers, and teachers with at most a Bachelor's degree are distributed highly unequally among schools. For instance, when we rank schools by the percentage of teachers without a full credential, at the tenth percentile school and the 90th percentile school 0% and 17.6% of teachers lack a full credential respectively. Curriculum measures in the table include the percentage of classes that are Advanced Placement (AP) and the percentage of courses that help students to meet the "a-f" requirements for admission to the University of California and the closely related requirements for the California State University system. The table shows meaningful variations in these measures of high school curriculum. Betts, Rueben and Danenberg (2000) show that these variations in teacher characteristics and high school curriculum exhibited in the table are strongly related to student disadvantage.[6]

In summary, in the 1997-98 school year in California both LEP students and students fluent in English lagged behind national norms. California's spending per pupil fell behind the national average in the 1980's, and in 1997-1998 considerable inequalities in school resources, especially teacher qualifications, existed among schools. Schools also varied dramatically in average student achievement at this time. However, variations in poverty and the percentage of students who were LEP at each school explained far more of the variations in student achievement across schools than did variations in teacher preparation and other resources.

The large variations in student achievement among schools, and the gap between California students' performance and national norms both suggest that the rationale for statewide educational content standards, assessment, and intervention is strong. At the same time it raises important questions about the capability of schools to bring all students up to common norms of achievement, especially in light of variations in resources among schools. Compounding matters, even if resources could be equalized across schools, this equalization by itself would do little to close the test-score gap related to variations in student disadvantage. [7]

For these reasons, it becomes essential to understand how the accountability system sets standards, what allowance the accountability system gives to existing differences across socioeconomic strata in initial variations in achievement, the allowances made for Limited English Proficient students, and the way in which school resources have changed at schools over the last few years.

The School Accountability System in California: An Overview of the Public Schools Accountability Act (PSAA) of 1999 and Related Legislation

Introduction of a State Test and Content Standards

The standards being set in California today come in the form of a series of measures enacted at the state level. [8] In 1997, legislation based on Senate Bill 376 passed, by which California re-instituted a statewide test in spring 1998. This was not the first state test in California by any means. Matrix tests that assessed average achievement at the school level were used in the 1980s; in the early 1990's a written test was introduced statewide, only to be cancelled in 1994. The new testing system that began in spring 1998, the California Standardized Testing and Reporting program (STAR), initially relied on one test, the Stanford 9. The Stanford 9 is a nationally normed multiple- choice achievement test. School districts are required to administer the Stanford 9 to all students in grades 2 through 11, except for those students whose Individual Education Plans (IEPs) explicitly exempt them, or students whose parent or guardian submits a written request to exempt the student. (Exemption rates to date have been very low.) In grades 2 through 8 students are tested in reading, spelling, written expression, and mathematics. In grades 9 through 11 testing is required in reading, writing, mathematics, science, and history/social science.

At roughly the same time, the state developed a series of content standards stipulating what students should learn in a variety of subjects, on a grade-by-grade basis. Specifically, California adopted English/language arts and mathematics standards in late 1997, science and history/social science content standards in late 1998, and content standards in visual and performing arts in early 2001. The State Board of Education (SBE) typically has published two companion volumes for each subject, a set of content standards that lists skills that should be mastered in each grade, and a much longer volume containing a "content framework" that provides specific examples of each type of content. [9]

Given the inability of an off-the-shelf test like the Stanford 9 to reflect these content standards fully, beginning in 1999, additional test items in language arts and in mathematics were included as part of the STAR testing program. These additional items, known collectively as the California Content Standards or the STAR Augmentation, are intended to address California content standards that are not addressed by the Stanford 9. Also beginning in 1999, Spanish-speaking English language learners (limited-English proficient or LEP) who have been in California public schools fewer than 12 months must be administered the Spanish Assessment of Basic Education, 2nd Edition (SABE/2).

The Public School Accountability Act of 1999

The re-introduction of statewide testing and the creation of content standards paved the way for a statewide accountability system. Such a system was approved in 1999, through Senate Bill (SB) 1X, known as the Public Schools Accountability Act of 1999 (PSAA), as well as amendments and additions to it in SB1552 in 2000. The PSAA is motivated by the fact that many of California's students are not progressing academically at a satisfactory rate, and contains three key elements to reform the assessment and accountability of schools across the state in order to raise student achievement. These three elements are (1) the Academic Performance Index (API), (2) the Immediate Intervention / Underperforming Schools Program (II/USP), and (3) the High Achieving / Improving Schools Program (HA/IS) which includes the Governor's Performance Award Program (GPAP).

The API

The first element--the API--provides a method for indexing and ranking schools on a number of indicators, including performance-based indicators.[10] In the 1998-1999 and 1999-2000 academic years, the performance component of the API was based solely on the norm-referenced Stanford 9 test published by Harcourt Brace, Inc., which is administered in the spring. Other measures such as criterion-referenced standards-based tests (the aforementioned STAR Augmentation), a high school exit exam, attendance rates, and graduation rates are to be phased into the API in subsequent years.

Achievement tests constitute at least 60 percent of the value of the API, and exclude the test scores of students enrolled in a district for less than one year. Only comprehensive high schools, middle schools, and elementary schools with more than 100 students are included in the index.[11] All "numerically significant ethnic and socioeconomically disadvantaged subgroups" are included in the API for the purpose of measuring the progress of these student groups within schools. The legislation defines *numerically significant* as "a subgroup that constitutes at least 15 percent of a school's total pupil population and consists of at least 30 pupils." In addition, the API disaggregates pupil data collected from the tests by gender, ethnic or racial group, English language learners, socioeconomic status (SES), and special education status.

The API has been used to establish two types of performance targets. First, annual growth targets for schools were set based on the July 1999 API baseline score.

The minimum growth target is 5 percent annually, however the SBE may set differential targets for schools at the top and bottom of the distribution, with particular attention paid to those at the bottom "because they have the greatest room for improvement."[12] Second, statewide performance targets are being developed as the state content standards are implemented. This type of target represents a basic proficiency level as deemed appropriate by the SBE.

Elementary, middle, and high schools are ranked into deciles based on the API. In recognition of the role that socioeconomic status and language ability play key roles in determining student achievement, the state also ranks schools into ten deciles after first grouping schools together with other schools with similar characteristics. The "similar characteristics" used to derive each school's "Similar Schools Rank" include student ethnicity, student SES, student mobility, percentage of students who are English language learners (ELL), average class size, percentage of teachers who are fully credentialed, percentage of teachers on emergency credentials, and whether the school is on a multi-track, year-round calendar. Beginning in June 2001 schools will also be ranked by their actual growth rates and how their growth rates compare to schools with similar characteristics.

The II/USP

The second element—the II/USP -- provides a carrot-and-stick method for helping schools not performing well. In fall 1999, any school performing below the 50th percentile on the STAR test in *both* 1998 and 1999 were eligible to participate in the II/USP, which provides for both a state planning grant ($50,000) and a federal Comprehensive School Reform Demonstration (CSRD) implementation grant of at least

$50,000. External evaluations and an action plan arising from the evaluation are requirements for the participants in the II/USP. Improving academic performance, involving parents or guardians, and improving effective and efficient allocation of resources and school management are among the key elements of the action plan.

Among the possible consequences of an II/USP school's failure to meet performance goals and to show substantial growth within two years are state reassignment of school management, reorganization of the school, or closure of that school.[13] In addition to these direct state actions, the state may allow parents to apply for the creation of a charter school at the existing school site. Before any of these actions may be taken, however, a public meeting must be held and the state must publicly find the school to be a failure. In the 1999-2000 academic year, a first cohort of 430 schools was chosen for external evaluation, and in 2000-2001 a second cohort of 430 schools participated in II/USP.

The HA/IS and GPAP

The third element—HA/IS--and its subsection—GPAP—provide mechanisms for rewarding schools that meet or exceed their growth targets or state performance targets. All schools, including those in the II/USP, are eligible to participate in the GPAP, because the monetary rewards of up to $150 per pupil are based on meeting or exceeding growth targets rather than the school's absolute rank in the API. Other non-monetary awards such as classification as a distinguished school, placement on a public list of honor roll schools, and public commendations from the governor and/or legislature may be made in addition to or in lieu of monetary awards.

This emphasis on rewarding schools based on their rate of improvement, rather than their absolute level of achievement, represents a sensible solution to concerns that accountability systems could automatically funnel all the rewards to schools with the highest level of achievement, which typically are in affluent areas, have fewer LEP students, and often have more highly educated and experienced teachers. There is a second factor that somewhat mitigates problems in comparing schools with high and low percentages of students who are LEP: the API calculation excludes scores of students who are new to a district in a given year.

In 1999, only the STAR test results were used to identify II/USP and HA/IS schools. Since June 2000, the growth and statewide performance target components of the API have also been used to identify schools that are under-performing and those that are high achieving. Beginning in June 2001, the additional component of similar schools will be used to rank and identify schools for II/USP and HA/IS. Because our analysis compares outcomes across years, for issues of consistency we focus our analysis on the Stanford 9 test results only. However, the need to measure outcomes consistently across time is important for any researcher or policy-maker to consider when analyzing the API rankings and their implications for schools and children.

Recent Legislative Initiatives

The II/USP program channels additional funds to schools with low test scores, but in a rather limited way. Several other education bills passed during the 1999 emergency session of the legislature or in 2000 pay particular attention to funneling additional resources to the schools that have the lowest test scores or high proportions of low-SES students. For example, section 52247 of the California Education Code provides funding to increase college-going rates in high schools that have low college participation rates. In addition, the Awards and Evaluation Unit of the Policy and Evaluation Division administers a series of monetary incentive-based reward programs.

Incentive Programs for Schools and Staff

The PSAA and related legislation contain three monetary-based incentive programs for schools and staff: (1) Chapter 3 of the 1999 PSAA—the Governor's Performance Awards (GPA), (2) Senate Bill (SB) 1667, Chapter 71 of 2000—the School Site Employee Performance Bonus (SSEPB), and (3) Assembly Bill (AB) 1114, Chapter 52 of 2000—the Certificated Staff Performance Incentive Act (CSPIA). These pieces of legislation appropriated monetary rewards to schools and staff totaling $677 million. The next few paragraphs rely heavily on the description on the CDE website.[14]

The GPA rewards a *school* that has an API ranking if it meets all of the following conditions: the 2000 API shows 5% growth, 80% of the growth target is met by each subgroup, K-8 schools have 95% and 9-12 schools have 90% Stanford 9 participation rates, and schools already at or above 800 on the API have at least a 1-point gain. Schools meeting all of these conditions are eligible for rewards based on student enrollment at $63 per pupil, and the funds are for school-wide use to be determined by the site governing team and ratified by the local school board.

The SSEPB rewards ***all*** *staff at a school site* that has an API ranking and meets the same conditions that apply to the GPA above. It is funded at $591 per full time equivalent (FTE), all staff receive it based on their FTE, and an equal amount of money is to be given to the school and used under the same determination as above. In other words, if a school's staff consists of 10 FTE, the school receives $5,910 for the staff, and matching funds for the school. This matching amount is in addition to any monies received under the GPA above.[15]

The CSPIA rewards ***certificated** staff* in schools with API rankings in deciles 1 to 5 (i.e. low rankings) in 1999 who meet the following conditions: 1998 to 1999 Stanford 9 test score growth (unspecified growth amount), the 2000 API is twice the annual growth target of the school (10%), in K-8 schools 95% and in 9-12 schools 90% of the teacher's students are tested, and all subgroups make 80% of the growth target. The monies are distributed across the state as follows: 1,000 staff in schools with the largest gains receive $25,000 each; 3,750 staff receive $10,000 each, 7,500 staff receive $5,000 each. The district and teachers' union jointly decide which personnel receive the rewards at each school site. To put this latter program in perspective, there were over 292,000 teachers who taught in the classroom (rather than holding administrative positions) in California in the 1999-2000 academic year, so that just over 4% of teachers are likely to win awards under this program in any given year.

Clearly these three incentive programs do provide additional motivation for schools and staff to raise test scores in their schools and classrooms. However such high-stakes may also lead to abuses of the system. Indeed, there have been reports in the popular media of teachers changing answers on the test forms to create fraudulent gains in their students' test scores. The latest API scores on the CDE site in fact list a small number of schools at which testing irregularities were reported.

In early 2001 the legislature began to consider additional means by which it might improve resources at schools with low test-scores. In April the legislature's education committees again began consideration of bills that would attract highly qualified teachers to schools with low test-scores. For instance, Senate Bill 572 would pay a $15,000 salary bonus to credentialed teachers who teach in schools that rank in the bottom fifth of the state in test scores.[16] In this section we have described in some detail the "carrot" side of the current carrot-and-stick system in California as it relates to schools and staff. We now turn to the implications for students, and how the new system may affect them.

Accountability for Students

As mentioned before, testing is nothing new in education. There is now, and has been for many years, a battery of tests students face throughout their educational careers. In California, the Stanford 9 is not alone, although it has garnered much of the recent attention from the press, the public, and researchers. The new accountability system prompts the question: Just how high are the high-stakes for students in California?

Two other statewide tests exist—the Golden State Exam (GSE), and the newer High School Exit Exam (HSEE). The GSE was established by Senate Bill 813 in 1983, and was reauthorized twice during the 1990s. This is a rigorous exam given in key academic areas in grades 7-11. Unlike the HSEE, which tests for basic proficiency in particular content areas, these exams are used to recognize students for outstanding academic achievement and to determine student eligibility for the Golden State Seal Merit Diploma.[17]

In April 1999, Senate Bill 2 authorized a new graduation requirement: seniors in the class of 2004 must pass an exit examination to receive a high school diploma. The HSEE was administered statewide for the first time in spring 2001 after a trial run in 2000. Only students in grade 9 could volunteer to take the test in 2001. Beginning in the

2001-2002 academic year, students in grade 10 will be **required** to take the exam. Because the test is so new, it is uncertain what the effects of this new requirement will be. Heubert and Hauser (1999) find that there is little research on the specific effects of graduation testing, but caution that possible links to consequences such as dropouts must be considered by a well-designed accountability system.[18]

Similar to the incentives for schools and staff, there are specific monetary rewards for high test-scores in California. Senate Bill 1688 (Ch 404, 2000) established the Governor's Scholarship Programs. The legislation appropriated $118 million for 2000-2001, to be divided between two programs—the Governor's Scholars Program and the Governor's Distinguished Mathematics and Science Scholars Program.[19] The former awards $1,000 scholarships based on high test scores on the Stanford 9 and may be earned up to three times, for grades 9, 10, and 11. The latter awards $2,500 scholarships to students who win both a Governor's Scholars Award *and* who earn high test-scores on a math and science AP test, GSE test, or International Baccalaureate (IB) test. This award may only be earned once.

From the above information about the incentives and penalties that the PSAA and related legislation places in California's new accountability system, it is clear that there *is* much at stake—for school, for staff, and for students at either end of the performance distribution. The next section provides an empirical examination of trends in the distribution of test scores and school resources in California.

Trends in California's Test Scores and Gaps Between Groups

Summary Level Trends

We begin our analysis of achievement trends in California with summary level test scores for all students, limited English proficiency (LEP) students, and non-LEP students in the state. Because the state test has only been administered since spring 1998, we examine three years of test scores. The PSAA was passed just before the spring 1999 test was given, and most of its elements were not implemented until after the 1999 test. We therefore focus much of our attention on differences between the 1997-1998 and 1999-2000 academic years, interpreting this as a before/after analysis of the reforms.

Figures 4, 5 and 6 show the three-year trend in math scores for the percentages of students scoring above the national 75th percentile, at or above the national median, and above the national 25th percentile, respectively. The data for these figures as well as corresponding scores in reading are in appendix tables A-2, A-3 and A-4. There, we take a simple average across grades as well as an average across grades weighted by the numbers of students who were tested in each grade level; the figures show the latter. We tested the significance of the difference—that is, that the one-year changes in test scores were zero.[20] These tables show the test results as well as results by individual grades.

The largest absolute gain from spring 1998 to spring 2000 is for LEP students scoring above the 25th percentile of national norms, and the difference of means is statistically significant. In fact, all of the differences of means for math are statistically significant for all three groups. It appears that math achievement has risen significantly over the three years in question. As the appendix tables show, the results for reading are not as strong. Two-year gains in the share of students scoring at or above the 25th

percentile of national norms are significant, but gains in the share at or above the median or 75^{th} percentiles are not statistically significant.

Before concluding that reading and especially math achievement have risen in California we should warn the reader that it is common for test scores to rise in the first few years after the introduction of a new test, quite independently of trends in the true underlying achievement of students. Koretz (1996) recounts evidence that rising test scores in one school district reflected growing student (and teacher) familiarity with the test form over several years. [21] He and co-authors found that when they administered an older test form, test scores reverted to the levels observed in the first year that the current test form was adopted. Because California has used the same test form since spring 1998, it seems quite likely that at least some of the improvement in test scores reflects growing student and teacher familiarity with the format of the test and, a more serious possibility, growing teacher familiarity with the specific questions that are asked on the test form for a given grade. We doubt that all of the improvement in scores reflects growing knowledge of the Stanford 9 test form, but it certainly is likely to explain a portion of the gains.

Changes in Inequality in Student Achievement

We present the same achievement data in a slightly different way in Table 2. Here, we present the overall percentage of students across all grades in each quartile of national norms for reading and math. In both subjects, for all students, LEP, and non-LEP students examined separately, between 1997-1998 and 1999-2000 in all cases but one the percentage point drop in the number of students in the bottom quartile of national norms exceeded the percentage point rise in the share of students in the top quartile of national norms. [22] In other words, at the same time that average achievement has been rising, there has been a compression in the distribution of test scores. To re-state this a third way, students who were the furthest behind national norms have indeed improved more quickly than students who were initially above national norms.

A Comparison of Test Score Growth in Schools at the Bottom and Top of the Test Score Distribution in 1998

In addition to examining overall trends in student achievement, it is crucial to study whether the test-score gap between the bottom-performing and top-performing schools has narrowed or widened over time. Indeed, the legislature passed the PSAA specifically to address the large variations in student performance across schools.

As was discussed in an earlier section, the vast differences between the proportion of LEP students in California and the nation makes comparisons of California's "all student" and LEP student categories with student performance in the nation as a whole quite problematic. Indeed in the national norming sample for the Stanford 9 test only 1.8 to 2% of students were LEP, while California has approximately 20% LEP students. Therefore, for a more accurate comparison we examine the percentage of non-LEP students scoring at or above the national median from this point forward. We ranked schools by the percentage of non-LEP students scoring at or above the national median in the Stanford 9 math test results at each school in 1997-1998, and identified the schools that ranked in the bottom quintile, or fifth, of performance, and those in the top quintile of performance *in that year*. We then followed these schools through 1999-2000. We also identified the bottom and top quintiles of schools when ranking based on reading scores. We ranked elementary, middle and high schools separately.

Table 3 shows the percentage of students at or above the 25th, 50th and 75th percentiles of national norms in these two samples of schools in 1997-1998 through 1999-2000. The top panel shows results for math; the bottom panel shows results for reading. The three rightmost columns in the table show changes in the percentage of students at or above each cutpoint between 1997-1998 and 1999-2000 for schools that initially had low and high test scores, and the change in the test-scoregap over the two-year period.

The results are clear. The percentage of students at or above the 75th percentile of national norms has grown in both types of schools – those that were low-performing in 1997-1998 and those that were high-performing. The gains are broadly similar in the two types of schools, although in most cases the gains have been modestly higher in the top-performing schools. One major exception to this rule occurs for reading achievement among high school students. Here, the percentage of students in the top quarter of national norms surged upward by 16.1 points in the top schools compared to a rise of only 1.9 points in the bottom schools.

When we examine the percentage of students scoring at or above the 50th and 25th percentiles of national norms, the story is reversed. That is, typically the bottom-performing schools in 1997-98 appear to have improved considerably faster than the top-performing schools in 1997-1998. The percentage-point increase in the share of students above these two cutoff points in bottom-performing schools is typically at least twice as big as the corresponding increase in top-performing schools. Again, there is one exception to this overall pattern: in top-performing high schools the increase in the percentage of students at or above the 50^{th} percentile in reading exceeded the increase in the schools that were initially low-performing. Apart from this exception, schools that initially had the lowest achievement do appear to have had more success in increasing the share of students in the top half and top three-quarters of national norms. For example, in elementary schools, the percentage of students at or above the national median in the bottom-performing schools for math increased from 18.3% to 36.7%, an increase of 18.5%. In contrast, the schools that were top performers in 1997-98 increased their percentage of students at or above national norms from 74.1% to 83.1%, an increase of just 9.1%.

Of course, the bottom-performing schools had more room for improvement; so one might have expected results like this even if the school accountability reforms had not directly caused the narrowing in the gap in achievement among schools that is apparent in the table. This argument is not persuasive though. We note that it is mathematically possible that top-performing schools could have matched the gain in the percentage of students at or above the 50^{th} or 75^{th} percentile of national norms at bottom-performing schools. For instance, in the case of elementary school math results, the 18.5% gain in the share of students at or above the national median at the bottom schools could theoretically have been matched at the top-performing schools: this would have required an increase in the share at or above the national median at the top schools from 74.1% to 92.6%. In reality, the share of students above the median grew from 74.1% to only 83.1%. It seems clear, then, that schools originally in the bottom fifth of achievement have had more success at increasing the share of students in the top 50%

and 75% of national norms than have the top-performing schools. This finding strongly supports claims that testing and accountability may help to narrow the achievement gap among schools. The fact that low-performing schools have "multiple incentives" to raise performance through eligibility for II/USP, GPAP, GPA, SSEBP, and CSPIA funds lends even more credibility to this argument.

Is the Narrowing of the the Test Score Distribution Genuine?

Although this conclusion appears to be quite strong, we were concerned that a statistical oddity known as "regression to the mean" might have driven the narrowing in the achievement gap. To understand regression to the mean, consider the following extreme hypothetical example. Suppose that all schools in California in truth have equally proficient students, but that there is a great deal of random variation from one year to the next in the test scores. Then our "top-performing" and "bottom-performing" schools in 1997-1998 really were identical, but some had better "random shocks" than others. If this were true, then by 1999-2000 the two sets of schools should have had reverted to the mean on average, producing roughly equal student achievement. We would have incorrectly interpreted this as a complete elimination of a "real" test-score gap in the two-year period, when the test-score gap was not meaningful in the first place.

We undertook three steps to test this possibility. First, if the schools in our low- and high-peforming groups simply had bad or good luck in spring 1998, then we should see no evidence that compression has occurred in the overall distribution of student test scores. As our earlier discussion of table 2 showed, for the state's overall student population there is clear evidence that the number of California students who moved out of the bottom quartile of national achievement exceeded the number who moved into the top quartile. This pattern corroborates the claim that student achievement in California has grown relatively more quickly at schools that initially had the lowest test scores.

As a second test for regression to the mean, we identified fresh samples of schools in the bottom and top fifth of achievement in each school year. This procedure is quite different from what we did in table 3 where we selected these schools just once, in 1997-1998, and followed the same subsets of schools over time. If regression to the mean were driving our result of convergence in student achievement, we would expect to see essentially no narrowing of the achievement gap between the bottom and top schools over time when we choose fresh samples of bottom and top schools each year.

Table 4 shows the results, using the same format as for table 3. Strikingly, the overall patterns we reported in table 3 persist: there has been a slightly higher increase in the share of students at or above the 75th percentile of national norms at the top schools, but the increase in the share of students at or above the 50th or 25th percentiles of national norms has been considerably larger in the bottom fifth of schools than the top fifth of schools. In other words, when we focus on students performing in the top half or top three-quarters of national norms, there has undeniably been a narrowing in the achievement gap among schools.

It is noteworthy that the gains made by bottom-performing schools relative to top-performing schools are much greater when we follow the same schools over time, as in table 3, rather than allowing the schools making up the top and bottom groups to change each year, as in table 4. This fact suggests that there could be some regression to the mean at play in table 3.

However, there is another interpretation more favorable to the view that school accountability provides strong incentives to those working in the bottom-performing schools to improve student achievement. Specifically, it could be the case that schools at the very bottom of the 1997-1998 test score results implemented various reforms to improve over time. That means that by 1999-2000 many of these schools had overtaken other schools that were originally ranked higher. If it is true that the public release of test score results creates stronger incentives to improve at the very weakest schools, then we should expect exactly what is observed in the two tables. First, schools that were originally bottom performers should have gained markedly relative to the top schools. Second, because these "bottom" schools should have overtaken some other schools, the overall variation in test scores statewide from one year to the next should have narrowed, as shown in table 4, but not as much as did the variation in test scores for the original bottom-performing schools.

As a third test that the apparent narrowing in achievement among schools merely represents regression to the mean, we should see the opposite pattern – a widening in the test-score gap – if we define our samples of low- and high-scoring schools using end-of-period scores rather than beginning-of-period scores. [23] To understand this, suppose that random errors affect each year's data on student achievement at each school, and that these random errors are the only reason why two schools' rate of improvement might differ between years. One way of detecting this is to re-analyze the school data after defining low- and high-scoring schools using end-of-period test scores. If all that was different between these schools was that they had had different amounts of random luck in the final year, then at the start of the period, their test scores should have been quite similar to each other. Figure 7 illustrates with hypothetical data, under the assumption of regression to the mean and, for simplicity, no real growth in average achievement. According to this theory, the narrowing in test scores across schools that we reported earlier would be reversed if instead we defined high- and low-scoring schools based on their end-of-period scores. This is shown by the solid lines in Figure 7. In the case illustrated, we would see declines in achievement at schools that initially had unusually high scores, and at schools that at the end of the period had unusually low scores. Conversely, the schools that would show gains in achievement would be those that initially had unusually low scores and those that at the end of the period had unusually high scores. Most important of all, the gap between low- and high-scoring schools should move in opposite directions, depending on whether we defined these schools using beginning- or end-of-period test scores. In fact, if there is no correlation in the errors over time, the change in the gaps should be opposite sign but equal in absolute value. (Note that even if all schools are truly improving, at the same rate over time, but with random errors that regress to the mean, then this statement should still be true.)

Table A-6 in the appendix performs this exercise, using samples of high- and low-performing schools selected based on 1999-2000 test scores. The results are quite similar to our results in table 3, where we selected the subsamples based on 1997-1998 scores. First, and most important, we do not find that the change in the gaps between low-scoring and high-scoring schools are equal and opposite to those found in table 3. Thus, pure regression to the mean without any true trends in achievement gaps can be ruled out. Second, as our second test suggested, there probably is some regression to the mean at work along with true trends in achievement gaps across schools. The narrowing in the gap in the percentage of students at or above the 25th percentile of national norms in table 3 also occurs, but to a lesser extent, in every case in table A-6, suggesting that the narrowing achievement gap is real. In most cases in table A-6, the change in the percentage of students scoring at or above the national median is very close to zero, while in table 3 the gap narrowed substantially in five out of six cases. The most logical interpretation is that there is both genuine narrowing in the achievement gap between low- and high-performing schools and regression to the mean, which work in opposite directions, nearly canceling each other out. That is, the estimates of test-score compression in table 3 overstate the true extent of narrowing while the estimates in table A-6 understate the narrowing. Finally, table 3 suggested that when we measure achievement by the share of students at or above the 75th percentile of national norms, in two cases the gap did not change, and in four cases it rose, by modest to meaningful amounts. The changes in this test-score gap in table A-6 are all positive, and in most cases are a few percentage points higher than in table 3. Together, these results suggest that a modest but genuine widening in the test-score gap may indeed have occurred at the 75th percentile of national norms.

In sum, these three tests suggest that although regression to the mean is present to some extent in the data, there has been a genuine narrowing of the test-score gap when measured by the gap in the percentage of students scoring at or above the 25th or 50th percentile of national norms, and a slight widening of the test-score gap when measured as the percentage of students scoring at or above the 75th percentile of national norms.[24]

Trends in School Resources and Dropout Rates

This section addresses the questions of whether accountability in California has been matched by an overall increase in resources, what the distribution of resources was before and after the PSAA and related reforms, and what the trend in dropout rates has been.

School Resources and Course Offerings in High Schools

There were over 270,000 teachers who taught in the classroom (rather than holding administrative positions) in California in the 1997-1998 academic year. By the fall of 1999, that number had risen to over 292,000. At the same time, the numbers of AP courses increased in high schools pursuant to an order in 1999 by Governor Davis. By these measures, school resources did increase pre-PSAA to post-PSAA. However, California's student population also increased from over 5.7 million to over 5.9 million during that time. The question of how these additional resources are distributed across California and how measures such as average class size and teacher characteristics have changed in recent years naturally follows.

Table 5 presents statewide averages for each of the three grade-spans, weighted by student enrollment at each school in each of the three years. In order to facilitate comparison with the earlier samples of schools we limit our sample to the schools with valid math or reading data. (In practice, only about 3% of schools and approximately 1% of students in regular schools statewide were excluded from our sample by this restriction.) Notice that resource levels are virtually identical whether we limit the sample to either schools with math or reading scores. Thus, any trends we observe are observed across all schools for which there were math and reading data. Asterisks in the final column in the table indicate whether the change in each resource level between fall 1997 and fall 1999 is statistically significant. Corresponding columns of asterisks to the right of the earlier columns show the results of tests that the one-year changes in resources were zero; we focus our attention on tests of the two-year changes shown in the final column of the table. Although the average student in elementary and middle schools attends a school in which class size did go down, the change between 1997 and 1999 is so small as to be virtually unnoticeable and is not statistically significant. In high schools, class size also went down, and this change is significant at the 0.01 level.

In all three grade-spans, the average student attends a school in which average teacher experience and the percentage of teachers with a Master's degree have declined, while the percentages of teachers with at most a Bachelor's and lacking full credentials have risen. In elementary schools the percentage of teachers with at most two years of teaching experience has decreased, however in middle and high schools it has increased. Of these five resource measures, in elementary and middle schools all differences in the means from 1997 to 1999 are significant except the percentage with at least a Master's, while in high schools all of these measures are significant at levels equal to or less than 0.05.

Because low experience and lack of credentials are relatively highly correlated (with a correlation of 0.46 in 1997-1998), we would expect the percentages to change in the same direction rather than inversely as they do in elementary schools. This finding suggests that it may be taking teachers longer to earn a full credential in the elementary schools and that middle and high schools may be hiring novice teachers who begin teaching without full credentials, but who earn their credential before they reach their third year of teaching.

In high schools we also examine the levels of college-prep and AP course offerings and look at changes since the PSAA of 1999. Table 5 shows that the average high school student attends a school in which the percentages of both types of curricula have increased, but the difference in the averages from 1997 to 1999 are significant only for the "a-f" series of courses. We will explore all of the measures discussed in this section for specific subgroups of schools and students in more detail later.

In addition to course-taking patterns, we also examine recent trends in the dropout rate in California. Theoretical work suggests that any increase in academic standards is likely to cause students on the borderline of success to give up or even drop out, unless the increase in standards is accompanied by additional resources targeted at the students at risk. [25] With the introduction of rigorous new content standards in the late 1990's and the new statewide test that began in spring 1998, with results going both to a student's teachers and parents, one might wonder whether the dropout rate has increased in California. Figure 8 shows official calculations by the California Department of Education on the annual dropout rates among California high school students. If anything, the trend in dropout rates appears to be downward.

Although it appears that the development of the statewide test and content standards has not adversely affected the dropout rate, in truth it may take a decade to see the full effects of the accountability reforms on the dropout rate. We say this because the component of the reforms that is most likely to affect dropout/graduation decisions is just now being phased in. The high school class of 2004 will be the first that is required to pass a new state high school exit examination before graduating from high school. Thus, if this new graduation requirement lowers the probability of high school graduation for *marginal* students, it may take several years to become apparent.

We also should caution the reader that the measurement of dropout rates in California has become a highly contentious issue, mainly because the state has not adopted a longitudinal database that tracks students as they move between schools, leave the state, or in the case of immigrants "reverse migrate" to their country of origin. We have reported the one-year dropout rates calculated by the state. Using enrollment and graduation data, Betts (forthcoming) shows that the ratio of high school graduates to grade 9 enrollment 3 years earlier in California has consistently hovered around 0.69-0.74 throughout the last half of the 1990's, suggesting a four-year dropout rate of about 25-30%. [26] This stands in stark contrast to the one-year dropout data in figure 8, which imply a dropout rate over the four years of roughly 12%. [27] For this and the other reasons mentioned above, it is too soon to conclude whether the new school accountability program has affected the dropout rate.

The Gap in School Resources and Course Offerings Before and After the Public Schools Accountability Act

This section poses two main questions. First, in 1997-1998, how did resources vary between schools with unusually high and low test scores? Second, since California began its accountability program are low test-score schools receiving more resources and high test-score schools receiving fewer resources? We calculate the size of the resource gap between high and low test-score schools, and measure how much the gap has changed from 1997 to 1999. We also test that any changes in the resource gap between schools are statistically significant.

Resource Levels across Years

As in our earlier section on test scores, we rank schools by the percentage of students scoring at or above the national median in math and reading tests in spring 1998. We then choose a cohort of low test-score schools and a cohort of high test-score schools in each grade-span and follow them across the three years for which data are available. In the interest of saving space, we discuss only the resource levels for schools ranked by math tests. However, the results for reading, which are shown in the tables, are similar.

Table 6 shows that students in high test-score schools in each of the three years attend schools, on average, with lower percentages of students receiving free/reduced lunch. Students in these schools also have higher levels of experienced teachers and more-educated teachers than do students in low test-score schools. On the other hand, the average student in a low test-score school attends a school with higher percentages of teachers with low education, low levels of experience, and who lack a full credential. We can observe these patterns in all three years in all grade-spans. Average class size appears to differ little between low and high test-score schools, and in fact, is slightly higher in high test-score schools. In high schools in each of the three years, the average student attending a low test-score school has a lower percentage of college-prep and AP courses than the average student in a high test-score school.

An analysis of 1997-1998 data showed that relationships between student socioeconomic status (SES) and resource levels are similar to the relationship we have just shown between average test scores and resource levels.[28] This is not surprising given the high correlation between SES and test scores.

Average Resource Differences across Years

We now examine whether the gap in resources between schools that in 1997-1998 had low and high test scores has changed meaningfully. Table 7 presents several measures that compare the 1997-1998 and 1999-2000 academic years. The first two pairs of columns in the table explore whether the changes in levels of resources at low and high test-score schools before and after the PSAA are likely to be random or systematic. The last three columns of the table calculate the 1997 high-low gap, the 1999 high-low gap, and the change in the size of each gap for each resource between 1997 and 1999.

We perform a paired t-test on the average difference between the resource levels observed in the 1997-1998 (pre-PSAA) data and the resource levels observed in the 1999-2000 data (post-PSAA) for the two cohorts of schools. The first two columns compare enrollment-weighted mean differences between the resource level in 1997-1998 and 1999-2000. We weight the average difference by 1999 enrollment to illustrate the average change over the last three years at a school attended by the average student attending a school of either type in 1999. This weight results in slightly different results than if we simply calculate the difference between 1997 and 1999 resources in Table 6.

Teacher Characteristics in Elementary Schools. The patterns of change in teacher characteristics at the elementary schools with the highest and lowest test scores in spring 1998 are complex. Overall, the patterns suggest that schools with high scores have fared better over the two-year period. The average K-6 student attends a school in which the average experience level of teachers has declined slightly, but the decline is significant only in low test-score schools. At the same time, the percentage of teachers with very low levels of experience has also decreased by 3 percent in the bottom-scoring schools and 5 percent in the top-scoring schools; the decrease is significant in both types of schools. However, the percentages of teachers with low education levels and lacking full credentials have increased at both types of schools, and are significant for all but credentials in the high test-score quintile. The percentage of highly educated teachers has decreased in the low quintile and is significant, while in the high quintile it has increased slightly but is insignificant, suggesting that in the low quintile the change is systematic, while in the high quintile it is random. Of the five measures of changes in teacher preparation that we focus on, all are significant in the low quintile of schools, while only three in the high quintile are.

Teacher Characteristics in Middle Schools. We see a similar pattern to that observed for K-6 schools, with the exception that of the five measures of change, only one—declining average teacher experience level—is significant in the high quintile of schools, while all five changes are significant at the 0.001 level in the low quintile of schools. These changes at schools that initially had low test scores are uniformly disturbing: over time teacher experience, education and credentials have deteriorated in a statistically significant fashion.

Teacher Characteristics and Course Offerings in High Schools. We examine the above teacher characteristics in high schools as well as the percentage of classes that are college-prep and AP. The average high school student in these groups of schools attends a school in which average teacher experience and the percentage of teachers with at least a Master's degree have declined slightly, whereas the percentage of teachers with at most a Bachelor's degree, at most two years of experience, and who lack full credentials have risen. In the low test-score quintile of schools, changes in all of these measures are statistically significant, while in the high quintile of schools, changes are statistically significant for only three of the five measures of teacher preparation, and when significant, are less significant than in the low quintile of schools.

The change in the percentage of classes that are AP at either type of school is significant. This last finding that the increase in AP classes is unlikely to be by chance is unsurprising, given Governor Davis's order in 1999 to increase the proportion of AP classes in the state.

Average Class Size. The mean differences in the table suggest that average class size has changed very little for the average student attending elementary and middle schools at either end of the test-score spectrum, and that any changes that have taken place are likely to have been by chance. For the average high school student in a low test-score school, there has been an average decrease of less than one student per class, and the change is not statistically significant. However, in high schools with high test-scores class size has decreased by slightly more than one student per class with significance of 0.05. Thus, the change in class size is likely to be systematic only in high test-score high schools.

Overall, these results suggest that several of the changes in resources are unlikely to be by chance and are more likely the result of systematic changes in the composition of the teaching force in these schools. We now compare the resource levels in the quintile of schools that initially had to the lowest test scores to those in the highest quintile and whether the gap we observe in 1997 has narrowed or widened in 1999.

Gaps between High and Low Test-score Schools, and Change in the Size of the Gap

The last three columns of Table 7 compare the average resource gaps between low and high test-score schools in 1997-1998 and 1999-2000 and the change in the gap measured as the high-low gap in 1999 minus the high-low gap in 1997. Remarkably, all resource gaps in all grade-spans have widened, with the exception of average class size in middle and high schools, and the percentage of college-prep classes in high schools.

These findings, together with data on the levels of resources from Table 6, suggest that resources are not being taken away from high test-score schools in order to increase resources at low test-score schools. Nor does it seem to be the case that resources are increasing at low-test score schools while resources are maintained at the top-scoring schools. Indeed, low test-score schools appear to be receiving lower levels of teacher resources over time despite additional funds being directed toward raising the quantity of highly qualified teachers. At high test-score schools, the rise in the proportion of teachers who are unqualified is many times less than that seen in low test-score schools for many of the measures. For example, in middle schools, the increase in the share of uncredentialed teachers is almost six times as high at low test-score schools as at high test-score schools. In elementary and high schools, a similar divergence in trends appears. Given the high correlation between student SES and test-scores, these findings imply that the resource gaps are also rising between low SES and high SES students.

What is unclear from these data is exactly what causal mechanisms for these patterns are operating in California. However, we note that the widening gap in resources between schools that initially had low and high test scores is related primarily to various measures of teacher preparation, and not class size or high school curriculum. Therefore it seems plausible that teacher mobility between schools, and perhaps different exit rates from the teaching profession across schools, are the driving forces behind the increase in resource inequality. One possibility is that the accountability system, with its sanctions for bottom-performing schools, has discouraged highly educated and experienced teachers from moving to, or staying at, such schools. Another possibility is that the class size reduction initiated in 1996 in elementary schools has lured highly qualified elementary teachers from low-performing schools to high-performing schools. However, this second hypothesis does little to explain why we find the same disturbing trends in not only elementary schools but also middle and high schools.

Policy Implications and Outlook

The Early Impact of School Accountability on Student Achievement and the Distribution of School Resources

Our analysis of test scores suggests that since the introduction of the state test in the spring of 1998, student achievement in California has grown overall and among LEP and non-LEP students. Of course, it is difficult to determine the cause or causes of this gain. Some of the gains could reflect growing familiarization with the test form, and some of the gains among LEP students might be related to the ending of bilingual education in California, although it is too early to know yet whether the latter is the case. It seems plausible that the introduction of testing itself, combined with publication of school rankings and reports to schools and parents about individual students' performance, might itself have spurred some of the improvement.

Patterns in the test scores over time can also inform the debate about who might be helped most by school accountability, and who might be hurt by it. A plausible scenario is that the implementation of school accountability could *aggravate* inequality in test scores, because schools in relatively affluent areas likely have greater resources—both in the school and in the home—to bring to bear upon improving student achievement. We compared growth in test scores among the schools that ranked at the bottom fifth and the top fifth of achievement in spring 1998, and found that the bottom-performing schools increased their shares of students above the 25th and 50th percentiles of national norms significantly more than did the top-performing quintile of schools. This finding is meaningful because mathematically the top schools had just as much room for improvement but did not improve to the same extent. Turning to the share of students at or above the 75th percentile of national norms, this measure of elite performance grew at both types of schools, but growth was slightly higher at the schools that were originally top performers. We undertook three checks to determine whether these trends in achievement gaps merely reflect regression to the mean. We found some evidence of regression to the mean but the overall patterns appear to be genuine.

Another concern about raising standards is that it can discourage students who hover at the borderline between success and failure. It is far too soon to know for sure if such effects will occur to a large extent. We note however, that one-year dropout rates in California high schools have been declining in recent years. Thus there is no evidence yet of a significant negative impact of the new system on dropout rates. California also needs to improve its measurement of the dropout rate if we are to test this hypothesis accurately.

A second set of issues surrounds resources and curriculum. Earlier we outlined a number of programs designed to bring additional resources to schools with low test scores. Have the schools that initially had the lowest test scores enjoyed a measurable increase in resources? At the top end of the test score distribution, at schools that largely in affluent areas, have resources gone up? One concern that seems quite plausible is that the need to devote additional time to test preparation might have in fact *diluted* the curriculum, especially in the top performing schools. Has this happened?

Our analysis of trends in school resources and curriculum reveal some important patterns. Most disturbingly, we find that the level of preparation of teachers in the bottom-achieving quintile of schools has *declined* between 1997-1998 and 1999-2000, both relative to teacher preparation among the top-fifth of schools and in most cases, in absolute terms. For instance, in the bottom-performing middle schools, between 1997-1998 and 1999-2000 the percentage of teachers lacking a teaching credential rose from 19.2% to 26.3%, while in the top-performing fifth of schools, the percentage of teachers lacking a credential barely budged, rising from 5.4% to 6.7%.

However, we cannot be sure of the reason for this and other trends in teacher preparation, and whether they are related to the school accountability program. Betts, Rueben and Danenberg (2000) voiced the concern that after the passage of the PSAA, teachers at the bottom-performing schools might leave these schools because of the additional public scrutiny such schools now receive. As we have outlined in the paper, the state has taken several steps to improve the level of resources at schools with low achievement. But it seems highly possible that these measures might be swamped by movements of teachers between schools in response to the sanctions that await bottom-performing schools that fail to improve adequately. Further research into this question is urgently needed. Unfortunately, it will be difficult to examine this issue at the state level due to the lack of a longitudinal database on California's teachers.

On the issue of possible dilution of the curriculum, we were able gain some insights into this question by examining trends in the percentage of high school courses that were Advanced Placement (AP) or college preparatory. Statewide, the percentage of courses that fit either description has risen between 1997-1998 and 1999-2000, although only the increase in the share of courses that are AP is statistically significant. We also looked for divergences in curriculum between the schools that were initially in the bottom or top fifth of the test score distribution in spring 1998. In both types of schools the percentage of classes that were AP rose significantly, and by roughly the same amount. If anything, the percentage of courses that were college preparatory seemed to rise more at the schools that initially had the lowest test scores. We conclude that at least by these measures, testing and the related aspects of accountability have not diluted the high school curriculum, nor has it widened inequality in the curriculum between top- and bottom-performing schools.

One of our discussants raised the following conundrum: how could it be that low-scoring schools have done more to improve student achievement at the 25^{th} and 50^{th} percentiles of national norms at the same time that their teacher resources have been declining both in absolute terms and relative to high-scoring schools? We cannot answer this question with any certainty, but two possibilities spring to mind. First, Betts, Rueben and Danenberg (2000) find that teacher qualifications are related to student achievement, but only rather weakly. Second, it seems somewhat plausible that less experienced teachers might have adapted to the new regime to testing and content standards more quickly than did their more experienced colleagues. For example, inexperienced teachers who have not yet fully prepared lecture notes and class materials might be able to meld their teaching more closely to the new standards.

In sum, we have found two important trends in the data. Test scores in California have risen significantly, especially for schools that originally had the lowest test scores. At the same time, teacher resources have if anything declined, especially at low-scoring schools. Given that student achievement in these low-performing schools has risen while resources have decreased, and given the monetary incentives for individual teachers in the low-performing schools, a skeptic could argue that these patterns reflect teaching to the test, or at the very least, growing test familiarity. An alternative interpretation is that the growth in achievement in spite of the significant drop in a number of measures of teacher preparation in low-performing schools reflects genuine achievement growth spurred by the accountability reforms and public scrutiny. We think that the latter hypothesis deserves to be taken seriously. It certainly should surprise critics of the standards movement who claim that additional resources are the only way to improve achievement at the bottom schools to learn that the bottom achieving schools have shown the most pronounced improvement in spite of a reduction in average teacher preparation at these schools.

Overall, we cannot know for sure which of these two explanations – teaching to the test or genuine improvement -- is the more important. One method of distinguishing between these divergent hypotheses is to analyze California's results on the National Assessment of Educational Progress (NAEP) over the next decade. If test scores on the state test continue to surge while NAEP scores do not improve significantly, it would suggest that our first, less sanguine, hypothesis was the dominant explanation for rising test scores on the Stanford 9.

The Outlook for the Future of School Accountability in California

After several years of rapidly introduced accountability reforms, what has California achieved, and what are the prospects for the future? We began this paper by defining a state school accountability system in terms of content standards, a testing program, and a system of actions that state agencies would take in reaction to specific outcomes at the student or school level. We could begin by assessing how well the current system fulfills each of these goals.

California has probably made the most progress in defining content standards, with a detailed set of standards of what students should know in each grade, with sample questions to guide teachers, students and parents alike. On the other two foundations of reform, testing and intervention, California has also made significant progress, although significant problems clearly remain.

First, the current Academic Performance Index, by relying solely on an off-the-shelf test, ranks schools based on subject matter that is not particularly closely aligned to the new content standards. By introducing supplementary test questions that specifically target the content standards, the state has advanced in this regard. However, the state now faces a difficult dilemma: while it seems apparent that at least theoretically all tests should link directly to the content standards, if California abandons the Stanford 9 test completely, it will once again have cancelled an existing statewide test, making it extremely difficult to know whether student achievement is truly improving over time. Another problem with abandoning the Stanford 9 test outright is that it provides the lone nationally normed test used by the state. It makes sense to provide some external yardstick against which to compare student achievement in California. Officials in Sacramento seem to be acutely aware of these issues. In late April the Senate Education Committee began considering a proposal from the governor to shorten the Stanford 9 test in all subjects but language arts, while increasing the length of new tests in various subjects that are more closely linked to the state's content standards. [29] The proposal has been approved and will apply to testing in 2002. The key to such a reform will be to ensure that a statistically valid "cross-walk" exists between the streamlined Stanford 9 tests in math and science that are being implemented and the more detailed test given in 1998 through 2000.

In this unsettled environment, it becomes extremely difficult to know what sort of testing system, if any, will be in place ten or even five years from now. If what emerges is a new testing system that links questions closely to the state's detailed subject framework, while providing a means of estimating whether student achievement is changing meaningfully over the years, it would represent an improvement on the current system. Clearly, the high school exit examination, already in place, will also become an important content-specific gauge of achievement for high school students.

The most difficult question facing the state's accountability system right now is whether it has the right mix of responses for schools where student achievement lags. The most compelling dilemma here relates to the High School Exit Examination. If the experience of other states is any guide, no matter what passing grade the state decides on, a large percentage of students in California will fail the test initially. To prevent this almost certain outcome from spiraling into higher dropout rates in California, it will become imperative to target additional assistance to students at risk and to their teachers. There is some reason to be optimistic. Both the legislature and the governor's office have taken steps recently to increase further the assistance going to the bottom schools. To give one example, in late April 2001 the governor's office announced that a plan to extend the school year across ***all*** middle schools would be scaled back, with the savings being devoted to additional resources, of an unspecified nature, for schools in the lowest two deciles of student performance. [30]

While the future of the school accountability system in California is not yet assured, it would be a mistake to underestimate the scope of what has already been achieved. In the space of half a decade, rigorous content standards have been developed. A statewide testing system has been introduced, and is being used to help schools identify weaknesses in the preparation of individual students, and to help administrators identify schools that are underperforming given the resources they receive and the socioeconomic status of their students. A high school exit examination has been introduced, and provides incentives for schools, parents and students to focus on mastery of basic skills. This test provides a worthwhile complement to the Golden State Exams and scholarships that currently reward students at the top end of the achievement spectrum. This subtle move to create a system of multiple incentives for students at different points in the achievement distribution represents a move toward the sort of multiple credential system that Betts and Costrell (2001) espouse as a way of improving incentives for all schools and all students.

At the same time, many measures have been enacted in order to improve teacher training and enrich the curriculum at bottom-performing schools. This trend represents a sea change in political thinking. It is true that Title 1 and similar state programs have for many years directed additional resources to schools in disadvantaged areas. But the multiple new programs in this vein represent significant movement beyond the notion, embodied in the *Serrano v. Priest* court decision of 1971, that all that California needed was equal access to funding across districts. Time will tell whether the current reforms in spending, together with the newly created incentives for teachers, administrators, and students suffice to bring meaningful and lasting changes to the level of student achievement and the dispersion in achievement in California.

We close by reiterating what may be our most important finding, that inequality in teacher preparation as measured by teacher credentials, education and experience has risen between the bottom- and top-performing schools in California. California needs to continue the search for powerful new incentives to retain highly qualified teachers at the schools in the greatest need. Meeting this challenge could prove pivotal to the long-term success of school accountability in California.

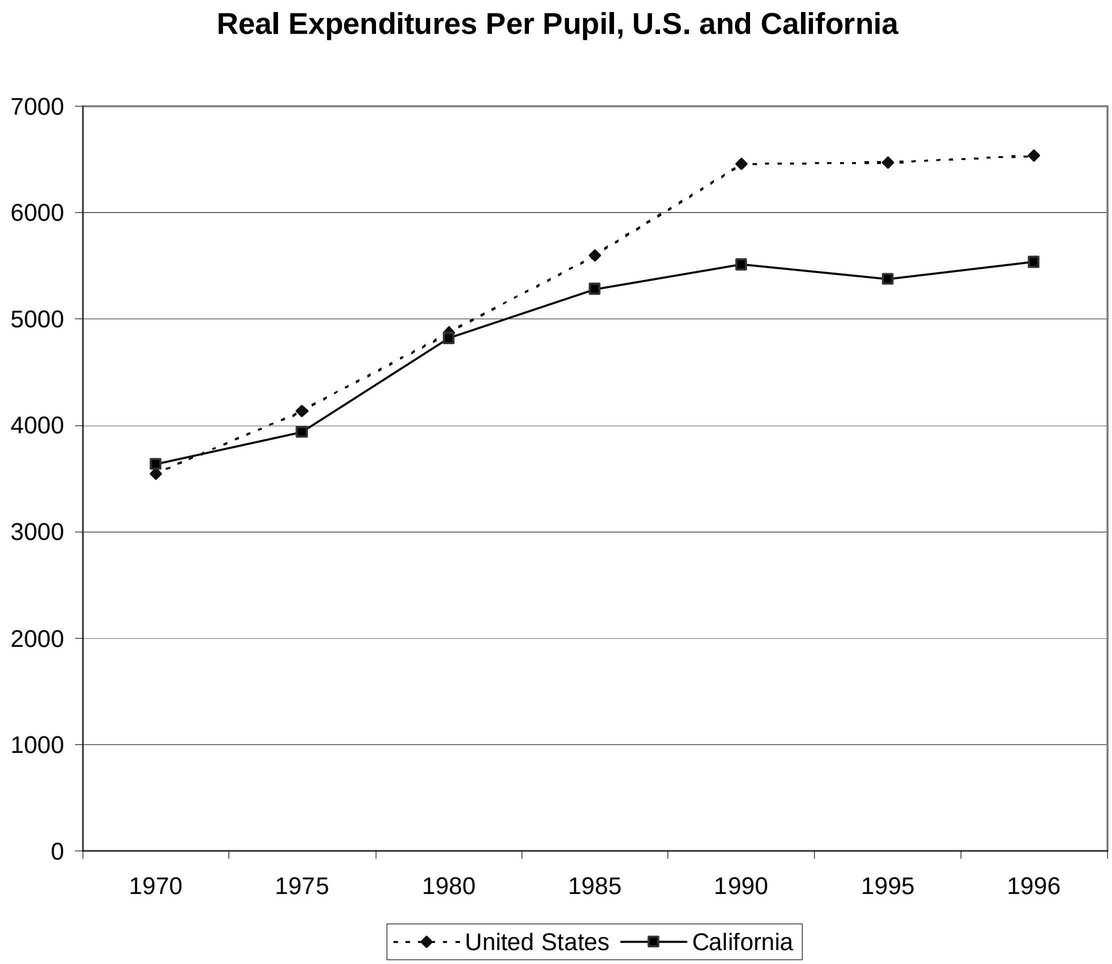
Tables and Figures

Table 1

Percentiles of California High School Characteristics, 1997-98, Weighted by Student Enrollment					
Variable	Distribution Percentile				
	10th	25th	50th	75th	90th
Class size	20.0	21.6	23.3	26.1	31.6
% a-f classes	40.8	47.2	54.9	61.1	67.0
% AP classes	0.6	1.4	2.2	3.4	4.8
% Teachers with 0-2 years of experience	5.2	8.4	12.2	16.8	21.7
% Teachers with at most Bachelor's	2.4	5.4	10.4	20.6	37.5
% Teachers not fully certified	0.0	2.3	6.3	12.2	17.6

Source: Betts, Rueben and Danenberg (2000), Table A.2.

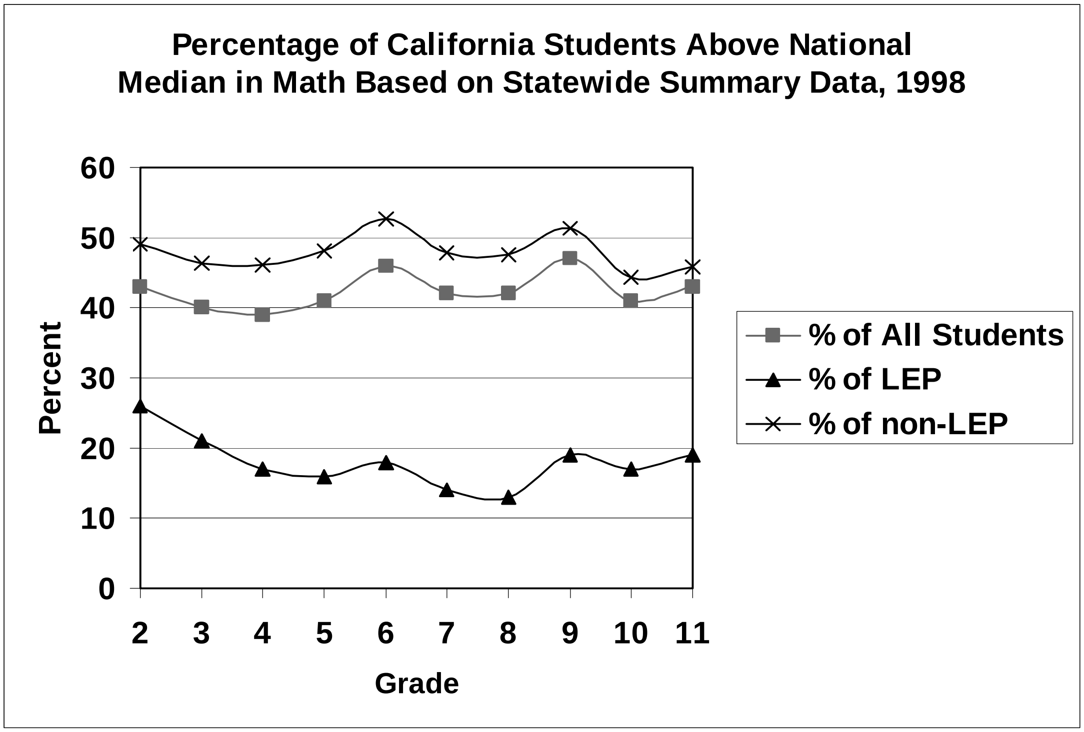
Figure 1



Notes: 1) Data are adjusted to 1997 prices using the CPI, obtained from the Bureau of Labor Statistics, U.S. Department of Labor.

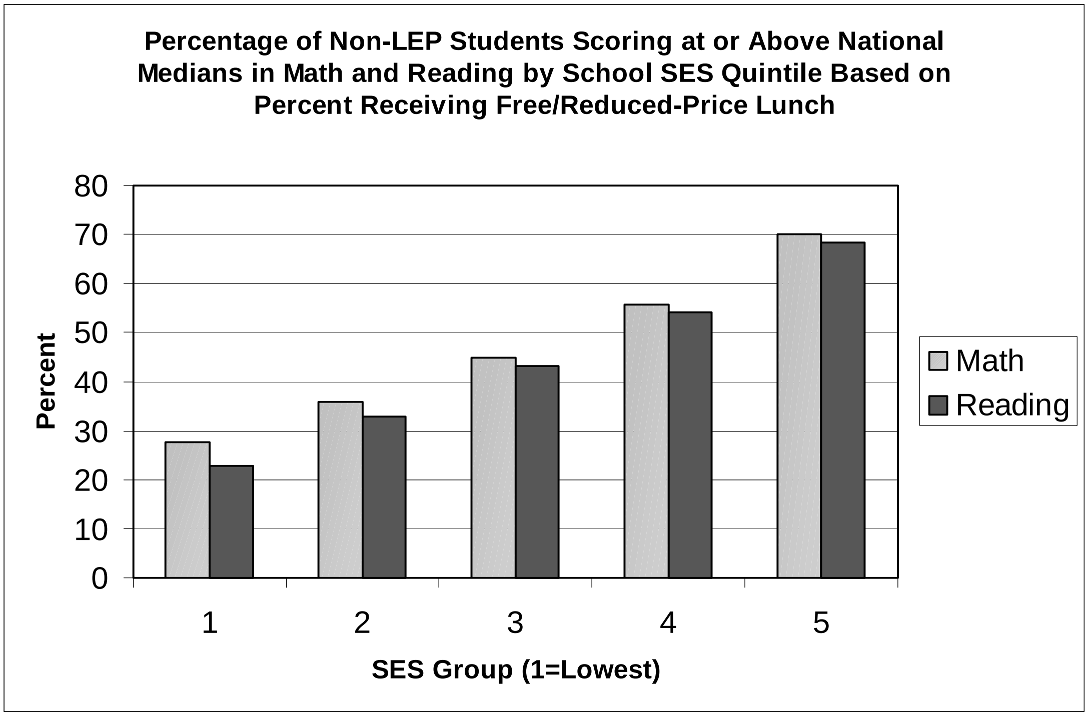
2) Data sources: National Center for Education Statistics (1999, Table 171, 1976 Table 75, 1971, Table 78.)

Figure 2



Source: Betts, Rueben and Danenberg (2000).

Figure 3



Source: Betts, Rueben and Danenberg (2000).

Figure 4. Average percentages of California students scoring in top quartile in math, 1997-1998 to 1999-2000

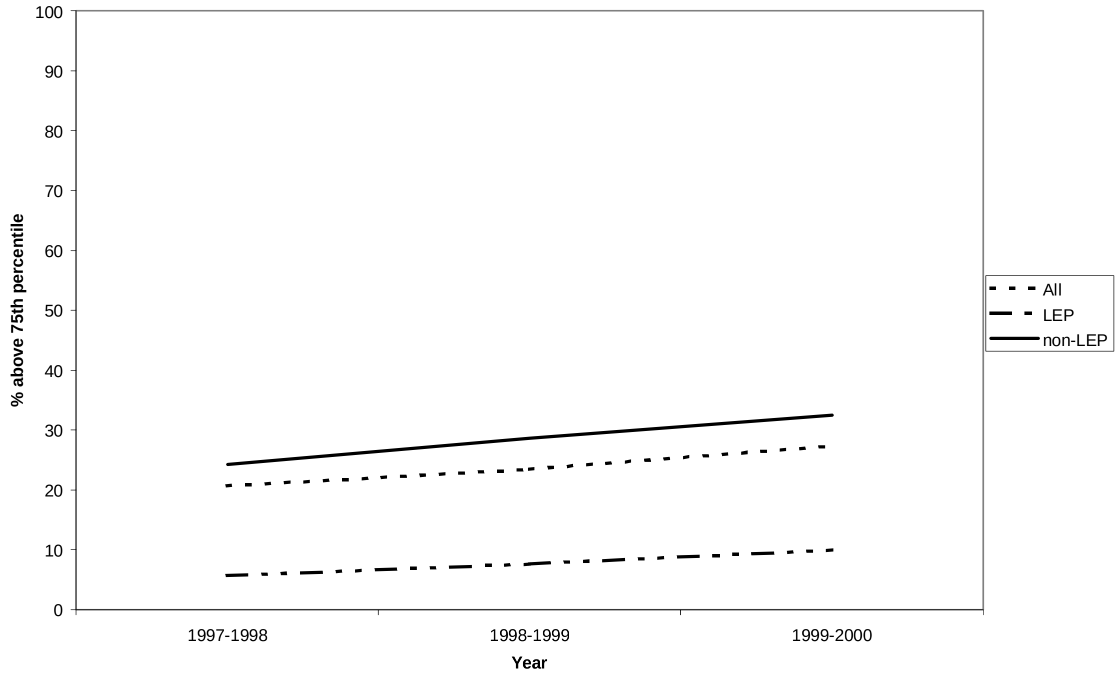


Figure 5. Average percentages of California students scoring at or above national median in math, 1997-1998 to 1999-2000

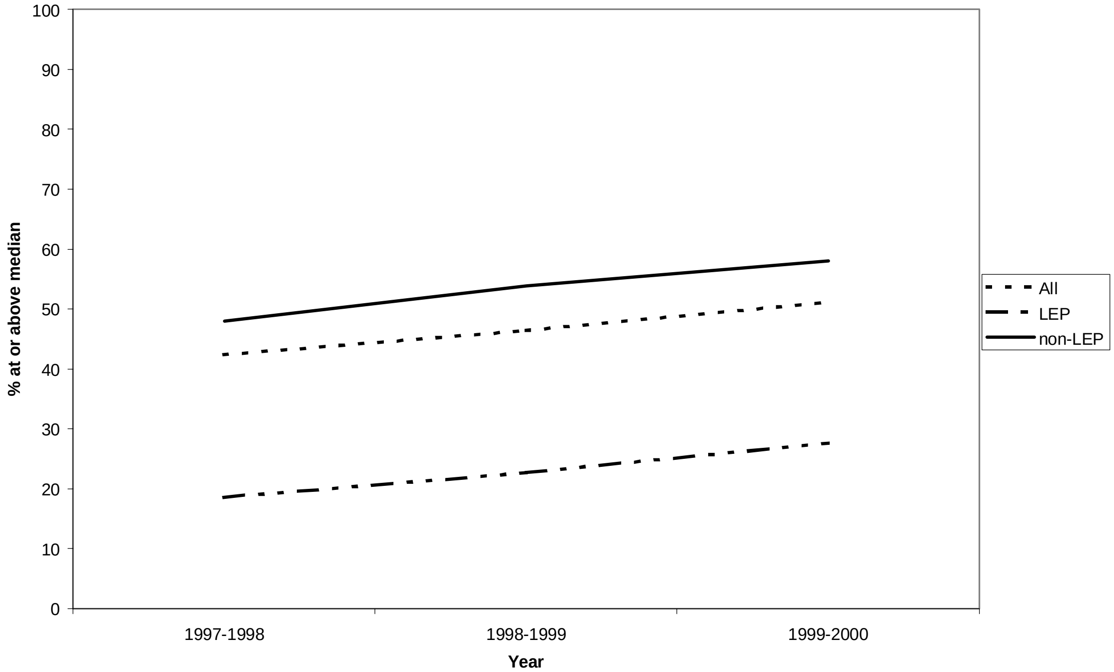


Figure 6. Average percentages of California students scoring above national 25th percentile in math, 1997-1998 to 1999-2000

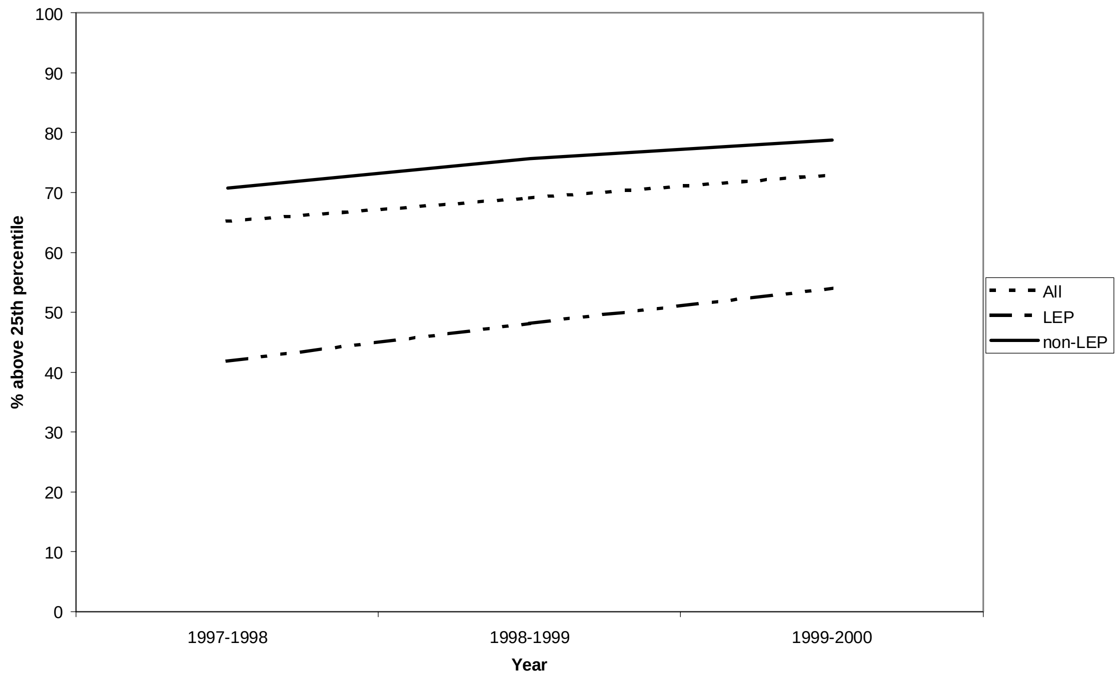


Table 2. Percentage of Students in Each Quartile of National Norms, by Student Group, 1997-1998 to 1999-2000.

	All students			LEP students			Non-LEP students		
	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
Math									
Quartile 1	34.78	30.87	27.02	58.21	51.80	46.03	29.27	24.34	21.24
Quartile 2	22.85	22.74	21.83	23.22	25.48	26.35	22.76	21.84	20.71
Quartle 3	21.67	22.91	23.81	12.87	15.10	17.65	23.74	25.23	25.64
Quartile 4	20.70	23.47	27.34	5.70	7.63	9.96	24.23	28.59	32.40
Reading									
Quartile 1	37.67	35.87	33.18	74.00	69.70	65.58	29.49	25.61	23.67
Quartile 2	22.90	23.42	23.75	17.85	20.39	22.32	24.04	24.23	24.53
Quartile 3	21.45	21.88	23.01	6.40	7.76	9.39	24.85	26.43	26.74
Quartile 4	17.97	18.82	20.06	1.75	2.15	2.70	21.62	23.74	25.06

Note: Quartile 1 is the Lowest Achievement Group.

Table 3 Average percentage of non-LEP students in 1997-1998 cohorts of high test-score schools and low test-score schools scoring at or above national norms in reading and math, by gradespan, 1997-1998 to 1999-2000

Math

		1997-1998		1998-1999		1999-2000		Changes, 1997-1998 to 1999-2000	
Gradespan	Percentile	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools
K-6	% above 75th	5.04	46.79	10.62	53.49	15.24	59.80	10.20	13.01
	% at or above 50	18.28	74.09	28.97	78.67	36.74	83.14	18.47	9.05
	% above 25	40.74	89.46	53.09	92.07	61.45	94.18	20.71	4.71
# observations		853	869	853	869	853	869	853	869
6-8	% above 75th	5.65	47.77	9.18	51.05	10.80	54.34	5.15	6.57
	% at or above 50	20.76	74.74	28.04	76.92	31.09	79.65	10.33	4.91
	% above 25	45.88	89.91	55.28	90.94	58.03	92.48	12.15	2.57
# observations		191	192	191	192	191	192	191	192
9-12	% above 75th	5.80	43.76	9.37	46.17	10.36	48.10	4.56	4.34
	% at or above 50	23.72	72.61	31.15	73.70	33.09	75.78	9.38	3.17
	% above 25	55.25	88.79	62.45	89.11	63.92	89.84	8.67	1.06
# observations		151	157	151	157	151	157	151	157

Table 3, continued.
Reading

		1997-1998		1998-1999		1999-2000		Changes, 1997-1998 to 1999-2000	
Gradespan	Percentile	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools
K-6	% above 75th	4.00	44.31	7.45	48.66	8.91	51.76	4.91	7.45
	% at or above 50	17.00	73.85	25.59	77.81	29.79	80.49	12.79	6.65
	% above 25	40.64	90.25	52.04	92.81	58.17	94.44	17.54	4.19
# observations		854	863	854	863	854	863	854	863
6-8	% above 75th	5.77	44.11	8.22	44.69	8.66	46.98	2.89	2.87
	% at or above 50	22.64	75.86	29.22	76.38	30.45	78.56	7.82	2.69
	% above 25	50.55	92.06	59.77	92.41	61.72	93.74	11.16	1.68
# observations		192	193	192	193	192	193	192	193
9-12	% above 75th	3.82	32.17	5.65	46.29	5.76	48.27	1.94	16.10
	% at or above 50	16.23	62.43	20.75	73.76	21.20	75.83	4.98	13.40
	% above 25	43.41	84.70	50.26	89.12	51.09	89.83	7.68	5.12
# observations		152	153	152	153	152	153	152	153

Note: All difference in means are significant at < 0.05.

Table 4

Average percentage of non-LEP students in high test-score schools and low test score schools scoring at or above national norms in reading and math, by gradespan, 1997-1998 to 1999-2000

Math

		1997-1998		1998-1999		1999-2000		Changes, 1997-1998 to 1999-2000	
Gradespan	Percentile	Low-score schools	High-score schools	Low-score schools	High-score schools	Low-score schools	High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools
K-6	% above 75th	5.04	47.22	9.53	55.27	13.50	62.06	8.46	14.85
	% at or above 50	18.34	74.38	26.75	80.63	33.67	85.24	15.33	10.86
	% above 25	40.94	89.56	51.35	93.09	59.20	95.19	18.26	5.63
# observations		895	896	888	890	914	926		
6-8	% above 75th	5.66	48.36	8.00	52.02	9.78	55.39	4.12	7.03
	% at or above 50	20.57	75.18	25.44	78.08	29.03	80.97	8.46	5.79
	% above 25	45.38	90.11	52.36	91.76	55.76	93.32	10.38	3.20
# observations		202	203	199	205	212	222		
9-12	% above 75th	4.98	43.74	7.10	46.77	7.50	47.96	2.52	4.22
	% at or above 50	21.86	72.70	25.28	74.97	27.34	76.51	5.48	3.81
	% above 25	52.33	88.83	57.19	90.00	58.18	90.40	5.85	1.57
# observations		168	169	158	162	175	186		

Table 4, continued

Reading

Gradespan	Percentile	1997-1998 Low-score schools	1997-1998 High-score schools	1998-1999 Low-score schools	1998-1999 High-score schools	1999-2000 Low-score schools	1999-2000 High-score schools	Changes, 1997-1998 to 1999-2000: 1997-1998 Low-score schools	Changes, 1997-1998 to 1999-2000: 1997-1998 High-score schools
K-6	% above 75th	4.04	44.89	7.15	50.22	8.30	53.61	4.26	8.72
	% at or above 50	17.03	74.20	23.85	79.46	27.71	82.37	10.67	8.17
	% above 25	40.68	90.29	50.59	93.48	56.11	95.23	15.43	4.95
# observations		895	896	884	892	907	917		
6-8	% above 75th	6.03	44.85	7.51	46.21	8.15	48.65	2.13	3.80
	% at or above 50	23.03	76.35	27.25	77.89	28.95	79.97	5.91	3.62
	% above 25	50.22	92.17	57.28	93.11	58.95	94.39	8.73	2.23
# observations		202	203	200	205	212	215		
9-12	% above 75th	3.66	32.58	4.42	32.53	4.71	33.77	1.06	1.19
	% at or above 50	15.73	62.81	17.62	63.53	18.35	64.91	2.61	2.10
	% above 25	42.17	84.91	44.71	85.83	47.15	86.51	4.97	1.60
# observations		168	169	157	161	177	183		

Note: This table, unlike table 3, defines low-score and high-score schools based on the test score rankings from the stated school year. Thus, the identity of schools in each of these quintiles varies by year, whereas in table 3, the composition of the low-score and high-score cohorts is identical across years and is based on 1997-1998 rankings. All differences of means are statistically significant at <0.05 except reading in high test-score high schools, which is not significant..

Table 5: Weighted average school and teacher characteristics for subsamples of schools with reading and math test scores, by gradespan, from 1997-1998 to 1999-2000

Math

		1997-1998		1998-1999		1999-2000	
Gradespan	School Characteristic	Mean	97-98 Pr |T|	Mean	98-99 Pr |T|	Mean	97-99 Pr |T|
K-6	% Free/reduced lunch	56.66		56.23		55.79	
	Avg. Class Size	24.85	***	24.28	*	24.60	
	Avg. Teacher Exp.	12.30	***	11.87		11.84	***
	% at most BA	21.08	***	23.24		22.87	***
	% at least MA	26.23		25.95		25.96	
	% low exp. (0-2 yrs)	21.13	***	19.78	***	16.73	***
	% lacking credential	12.10	***	13.20		13.63	***
Number of observations		4477		4442		4587	
6-8	% Free/reduced lunch	47.81		48.52		47.53	
	Avg. Class Size	28.68		28.98		28.64	
	Avg. Teacher Exp.	13.70	***	13.14	*	12.81	***
	% at most BA	18.33	***	21.05		22.38	***
	% at least MA	33.14		32.49		32.41	
	% low exp. (0-2 yrs)	14.13	***	15.67		16.40	***
	% lacking credential	10.27	**	12.03	***	14.66	***
Number of observations		1011		1012		1071	
9-12	% Free/reduced lunch	30.86		32.30		32.27	
	Avg. Class Size	29.70		29.76	**	28.77	**
	Avg. Teacher Exp.	15.27	***	14.73	***	14.24	***
	% at most BA	15.60	***	18.15	*	19.63	***
	% at least MA	39.73		38.70		37.85	**
	% low exp. (0-2 yrs)	12.98		13.63	*	14.52	***
	% lacking credential	8.35	**	9.75	***	12.40	***
	% "a-f" classes	59.42	*	60.89		61.46	**
	% AP classes	2.97		3.11		3.21	
Number of observations		843		794		896	

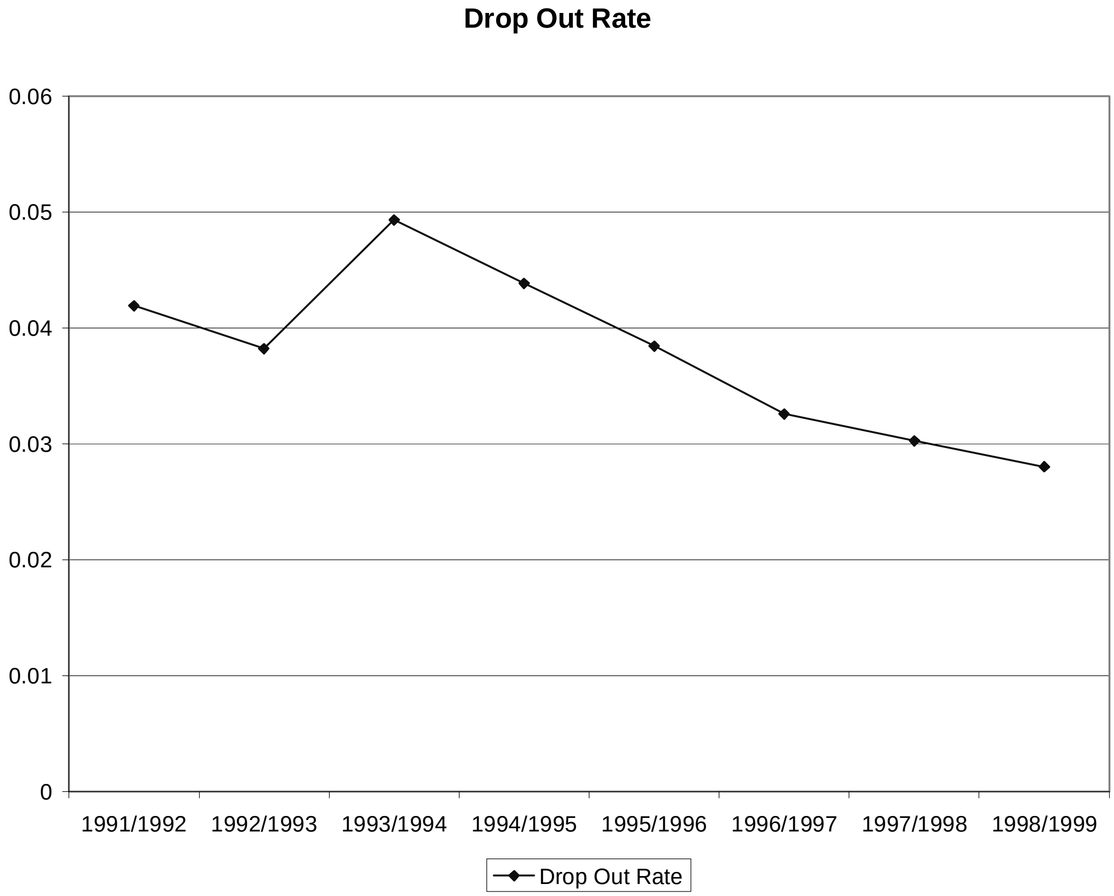
Probability > |t|: * =0.05, ** =0.01, *** =0.001

Table 5, continued.
Reading

Gradespan	School Characteristic	1997-1998 Mean	97-98 Pr |T|	1998-1999 Mean	98-99 Pr |T|	1999-2000 Mean	97-99 Pr |T|
K-6	% Free/reduced lunch	56.66		56.21		55.79	
	Avg. Class Size	24.85	***	24.28	*	24.60	
	Avg. Teacher Exp.	12.30	***	11.87		11.84	***
	% at most BA	21.08	***	23.25		22.87	***
	% at least MA	26.23		25.94		25.96	
	% low exp. (0-2 yrs)	21.13	***	19.79	***	16.74	***
	% lacking credential	12.10	***	13.21		13.63	***
Number of observations		4477		4437		4585	
6-8	% Free/reduced lunch	47.81		48.52		47.53	
	Avg. Class Size	28.68		28.98		28.64	
	Avg. Teacher Exp.	13.70	***	13.14	*	12.81	***
	% at most BA	18.33	***	21.05		22.38	***
	% at least MA	33.14		32.49		32.41	
	% low exp. (0-2 yrs)	14.13	***	15.67		16.40	***
	% lacking credential	10.27	**	12.03	***	14.66	***
Number of observations		1011		1013		1071	
9-12	% Free/reduced lunch	30.86		32.34		32.27	
	Avg. Class Size	29.70		29.77	**	28.77	**
	Avg. Teacher Exp.	15.27	***	14.73	***	14.24	***
	% at most BA	15.60	***	18.14	*	19.63	***
	% at least MA	39.73		38.72		37.85	**
	% low exp. (0-2 yrs)	12.98		13.63	**	14.52	***
	% lacking credential	8.35	**	9.74	***	12.40	***
	% "a-f" classes	59.42	*	60.88		61.46	**
	% AP classes	2.97		3.11		3.21	
Number of observations		843		795		896	

Source: California Department of Education datasets
Probability > |t|: * =0.05, ** =0.01, *** =0.001

Figure 8 One-Year Dropout Rates in California



Source: California Department of Education

Table 6: Comparison of average school and teacher characteristics in 1997-1998 cohorts of high test-score schools and low test-score schools for reading and math, by gradespan, from 1997-1998 to 1999-2000

Math

		1997-1998		1998-1999		1999-2000	
Gradespan	School Characteristic	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools
K-6	% Free/reduced lunch	86.35	17.72	85.66	17.22	85.00	16.95
	Avg. Class Size	24.34	25.38	23.83	24.85	24.31	25.41
	Avg. Teacher Exp.	11.16	13.43	10.37	13.18	10.27	13.28
	% at most BA	30.46	12.69	33.35	14.74	32.93	14.58
	% at least MA	23.37	30.76	22.89	30.96	22.34	31.48
	% low exp. (0-2 yrs)	24.89	17.47	25.25	15.13	21.75	12.38
	% lacking credential	20.97	4.82	23.56	4.53	24.10	5.39
Number of observations		853	869	853	869	853	869
6-8	% Free/reduced lunch	76.51	16.37	78.69	15.60	76.17	14.96
	Avg. Class Size	27.88	30.35	28.09	30.18	27.51	29.82
	Avg. Teacher Exp.	12.64	14.96	11.72	14.72	11.25	14.60
	% at most BA	27.95	10.89	30.56	14.14	33.53	14.83
	% at least MA	29.96	37.21	29.18	36.88	27.87	37.65
	% low exp. (0-2 yrs)	16.59	11.71	19.78	12.24	21.72	11.80
	% lacking credential	19.21	5.42	22.35	4.74	26.31	6.73
Number of observations		191	192	191	192	191	192
9-12	% Free/reduced lunch	53.31	11.05	57.83	10.68	59.16	10.16
	Avg. Class Size	28.43	30.32	28.47	29.53	28.03	28.82
	Avg. Teacher Exp.	14.42	16.63	13.92	16.04	13.36	15.61
	% at most BA	24.99	11.40	28.26	12.71	30.73	13.78
	% at least MA	38.03	45.72	36.66	45.02	35.59	44.52
	% low exp. (0-2 yrs)	13.88	11.08	14.64	11.71	16.70	12.07
	% lacking credential	13.93	6.13	15.79	5.20	20.28	7.13
	% "a-f" classes	53.88	66.43	57.26	69.08	57.59	68.28
	% AP classes	2.14	4.29	2.31	4.54	2.51	5.08
Number of observations		151	157	151	157	151	157

Table 6, continued.

Reading		1997-1998		1998-1999		1999-2000	
Gradespan	School Characteristic	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools	1997-1998 Low-score schools	1997-1998 High-score schools
K-6	% Free/reduced lunch	88.45	14.59	87.91	14.12	87.15	13.63
	Avg. Class Size	24.60	25.48	23.85	24.91	24.31	25.36
	Avg. Teacher Exp.	11.01	13.51	10.31	13.28	10.13	13.41
	% at most BA	32.26	12.54	35.42	14.14	35.19	13.67
	% at least MA	23.60	30.02	22.95	30.21	22.22	30.65
	% low exp. (0-2 yrs)	25.51	16.99	25.78	14.61	22.48	11.87
	% lacking credential	22.20	4.04	24.80	3.79	25.33	4.54
Number of observations		854	863	854	863	854	863
6-8	% Free/reduced lunch	78.73	14.64	81.06	14.28	77.99	13.51
	Avg. Class Size	27.87	30.09	28.19	29.73	27.60	29.62
	Avg. Teacher Exp.	12.81	14.90	11.94	14.65	11.41	14.56
	% at most BA	29.72	10.76	32.55	13.85	35.40	14.17
	% at least MA	30.05	36.26	29.09	35.72	27.75	36.77
	% low exp. (0-2 yrs)	16.00	11.37	19.06	12.08	21.59	11.23
	% lacking credential	18.66	5.20	21.81	4.64	25.99	6.13
Number of observations		192	193	192	193	192	193
9-12	% Free/reduced lunch	53.95	9.17	58.26	8.84	59.44	8.39
	Avg. Class Size	28.28	29.95	28.44	29.66	28.19	28.86
	Avg. Teacher Exp.	14.67	16.48	14.12	15.87	13.51	15.50
	% at most BA	24.45	10.29	27.66	10.58	30.24	12.08
	% at least MA	38.36	45.47	37.11	44.73	35.66	44.43
	% low exp. (0-2 yrs)	13.45	11.07	14.45	11.68	16.77	11.75
	% lacking credential	13.33	5.35	15.10	4.73	19.81	6.53
	% "a-f" classes	54.74	67.96	58.10	69.36	58.13	68.84
	% AP classes	2.27	4.40	2.41	4.65	2.60	5.00
Number of observations		152	153	152	153	152	153

Note: Weights used for means in each year that year's enrollment.

Source: California Department of Education datasets.

Table 7: Average change in school and teacher characteristics between 1997-1998 and 1999-2000 for 1997-1998 high test-score schools and low test-score schools, gaps between low and high test score schools, and change in gap size, for reading and math, by gradespan.

Math

		Average resource level change from 1997 to 1999 for 1997-1998 high and low test-score schools (Weighted)				Gap between low and high test-score schools' average resources (from Table 6)		Change in size of gap, fall 1997 and fall 1999
Gradespan	School Characteristic	Low-score schools	PrT	High-score schools	PrT	Fall 1997 Gap	Fall 1999 Gap	99 Gap Minus 97 Gap
K-6	Avg. Teacher Exp.	-0.85	***	-0.09		2.28	3.00	0.73
	% at most BA	2.22	***	1.81	***	17.76	18.35	0.59
	% at least MA	-1.06	***	0.76	**	7.39	9.14	1.75
	% low exp. (0-2 yrs)	-3.17	***	-5.27	***	7.42	9.37	1.95
	% lacking credential	2.91	***	0.53		16.15	18.71	2.57
	Avg. Class Size	-0.02		0.00		1.04	1.10	0.05
6-8	Avg. Teacher Exp.	-1.43	***	-0.35	**	2.32	3.35	1.03
	% at most BA	5.59	***	3.84		17.06	18.71	1.64
	% at least MA	-2.18	***	0.39		7.26	9.78	2.53
	% low exp. (0-2 yrs)	5.11	***	0.03		4.88	9.91	5.03
	% lacking credential	7.16	***	1.22		13.79	19.58	5.79
	Avg. Class Size	-0.37		-0.29		2.47	2.31	-0.16
9-12	Avg. Teacher Exp.	-1.04	***	-0.99	***	2.20	2.25	0.04
	% at most BA	5.45	***	2.50	*	13.59	16.95	3.36
	% at least MA	-2.61	***	-1.25	*	7.70	8.93	1.23
	% low exp. (0-2 yrs)	2.76	***	0.97		2.79	4.64	1.84
	% lacking credential	6.14	***	1.12		7.80	13.15	5.34
	% "a-f" classes	3.24	*	1.72		12.54	10.69	-1.85
	% AP classes	0.36	***	0.79	***	2.14	2.57	0.43
	Avg. Class Size	-0.48		-1.39	*	1.89	0.79	-1.10

Table 7, continued.
Reading

Gradespan	School Characteristic	Average resource level change from 1997 to 1999 for high and low test-score schools (Weighted)				Gap between low and high test-score schools' average resources (from Table 6.1)		Change in size of gap, fall 1997 and fall 1999
		Low-score schools	PrT	High-score schools	PrT	Fall 1997 Gap	Fall 1999 Gap	99 Gap Minus 97 Gap
K-6								
	Avg. Teacher Exp.	-0.86	***	-0.02		2.49	3.28	0.79
	% at most BA	2.78	***	1.06	**	19.72	21.52	1.80
	% at least MA	-1.39	***	0.67	*	6.42	8.43	2.00
	% low exp. (0-2 yrs)	-3.10	***	-5.27	***	8.51	10.60	2.09
	% lacking credential	2.97	***	0.49		18.17	20.80	2.63
	Avg. Class Size	-0.25		-0.15		0.88	1.05	0.17
6-8								
	Avg. Teacher Exp.	-1.45	***	-0.33	**	2.09	3.16	1.07
	% at most BA	5.63	***	3.30	***	18.96	21.23	2.27
	% at least MA	-2.40	***	0.36		6.21	9.02	2.82
	% low exp. (0-2 yrs)	5.60	***	-0.18		4.63	10.36	5.73
	% lacking credential	7.36	***	0.83		13.46	19.86	6.39
	Avg. Class Size	-0.27		-0.41		2.23	2.02	-0.20
9-12								
	Avg. Teacher Exp.	-1.16	***	-0.95	***	1.81	2.00	0.18
	% at most BA	5.60	***	1.88	*	14.16	18.16	4.00
	% at least MA	-2.86	***	-1.07	**	7.10	8.77	1.67
	% low exp. (0-2 yrs)	3.27	***	0.67		2.39	5.01	2.63
	% lacking credential	6.34	***	1.30		7.98	13.28	5.31
	% "a-f" classes	3.06	*	0.90		13.21	10.71	-2.50
	% AP classes	0.34	***	0.61	***	2.13	2.40	0.27
	Avg. Class Size	-0.14		-1.00		1.66	0.67	-1.00

Probability > |t|: * =0.05, ** =0.01, *** =0.001

Note: Columns 1 and 2 are mean differences between fall 1997 and fall 1999, and the weight used is fall 1999 enrollment. Column 5 may have 0.01 rounding error.

Source: Authors' calculations using California Department of Education datasets.

Appendix

Data Sources

CBEDS

The California Basic Educational Data System (CBEDS) is a data system maintained and supported by the Educational Demographics Unit in the California Department of Education (CDE). It contains individual-level credentialed-personnel data and summary-level student and program data at the school and district levels. These data are collected through three report forms each October: the Professional Assignment Information Form (PAIF), the School Information Form (SIF), and the County/District Information Form (CDIF).

Variables available at the individual level (PAIF) include education level, experience, and types of credentials held for credentialed personnel in California's public schools. The PAIF also collects information on specific classes taught and student counts per section for each teacher. From these data we calculate average class size and course offerings and then take school-level, weighted means of these measures and overall teacher characteristics, such as the proportions for experience, education, and credentials for each school.

The school-level data (SIF) contain variables of two general types: (1) staff and student counts, and (2) program types. Staff and student counts include classified staff counts and student enrollment, including student counts in specific types of programs, as well as graduate and dropout counts. These variables are enumerated by gender and ethnicity. Program types include variables such as technology, educational calendar, magnet programs, and alternative education.

CalWorks (formerly Aid to Families with Dependent Children (AFDC))

This school-level dataset contains counts and percentages of California children in families receiving Aid to Families with Dependent Children (AFDC) and children enrolled in free and reduced-price meal programs. According to the CDE, these AFDC data are collected each October through the cooperative efforts of the schools, districts, county offices of education, and the county offices of health and welfare. Schools report their meal program enrollment data annually, based on their October meal program enrollment files. (Both sets of data are collected on the California Department of Education, Education Finance Division Form No. CFP-2 School Level AFDC Report)

STAR Test Results

The STAR file is maintained by the Standards, Curriculum, and Assessment Division of the California Department of Education. It contains results from the *Stanford Achievement Test* Series, Ninth Edition, Form T (Stanford 9) test series administered by Harcourt, Brace & Co. These results are reported at school level in two ways for each subject area and grade level (grades 2-11 only): first, for all students tested in the group, and second, for limited English proficient (LEP) students tested. From these two measures, we also calculated non-LEP students' test scores.

There are six subject-test areas: 1) reading, 2) math, 3) language (written expression), 4) spelling, 5) science, and 6) history/social science. Students in grades 2 through 8 are required by Senate Bill (SB) 376[31] to take tests in the first four subject areas above. Students in grades 9 through 11 were required take tests in areas 1, 2, 3, 5, and 6 above. Our analysis focuses on the first two subject tests.

The following six statistics were reported at school, district, county, and state level: total number valid in each subject and grade, mean-scaled score, percent of normal curve equivalency, percent scoring above the 75^{th} percentile (based on national norms), percent scoring at or above the 50^{th} percentile, and percent scoring above the 25^{th} percentile.

Sample

Using the SIF, we identified different types of schools and then focused on schools that have regular academic programs. Thus, our analysis excludes students in special education schools, juvenile hall, continuation schools, and adult schools. The sample also excludes schools for which Stanford 9, SIF, and PAIF information were missing. We grouped the selected schools into four grade-span categories, based on similar characteristics of the schools. Each of the K-6, 6-8, and 9-12 groups included all schools where the enrollment fell entirely within one of these grade-span boundaries.[32] Schools where the enrollment crossed the boundaries described above were placed into an "Other" category, which we do not analyze in this research.

Table A-1 shows the numbers of schools, teachers, and students in the state and in our sample in each year. Our math sample includes over 75% of all schools, comprising over 88% of all students in each of the three years.

TABLE A-1 HERE

Data Tables

Tables A-2, A-3, and A-4 show the underlying data and t-test significance levels for Figures 4, 5, and 6 and Table 2 in the main text.

TABLES A-2, A-3, and A-4 HERE

Test Score Quintile Ranges

Table A-5 shows the boundaries of the five test score quintiles for math and reading in each of the three grade-spans in the 1997-1998 academic year.

TABLE A-5 HERE

TABLES

Table A-1 Numbers and percentages of schools and students in state, and in math test sample, 1997-1998 to 1999-2000.

Academic Year	Statewide	Sample	Percentage
1997-1998			
Schools	8,179	6,331	77.41
Students	5,727,303	5,145,831	89.85
1998-1999			
Schools	8,331	6,248	75.00
Students	5,844,111	5,150,133	88.13
1999-2000			
Schools	8,563	6,554	76.54
Students	5,951,612	5,381,890	90.43

Table A-2: State % scoring above 75th percentile in math and reading, 1997-1998, 1998-1999, and 1999-2000

Math	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	21	27	33	9	13	18	25	33	40
3	19	24	31	7	10	14	23	31	38
4	20	23	29	5	7	10	25	29	36
5	20	22	27	4	5	7	25	28	33
6	25	28	32	5	7	9	30	35	39
7	21	22	25	4	5	6	25	27	30
8	20	22	24	4	4	5	23	26	28
9	22	23	25	5	5	5	25	27	29
10	17	19	20	5	6	6	19	22	23
11	22	24	25	7	9	8	24	27	28
Simple average across grades *	20.7 (2.11)	23.4 (2.59)	27.1 (4.09)	5.5 (1.65)	7.1 (2.81)	8.8 (4.24)	24.2 (2.72)	28.5 (3.72)	32.4 (5.68)
Weighted average across grades **	20.7	23.5	27.3	5.7	7.6	10.0	24.2	28.6	32.4
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test	*		***	*	*	*	**		**

Probability > |t|: * =0.05, ** =0.01, *** =0.001

Reading	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	18	21	24	4	5	7	23	28	32
3	17	18	20	2	2	3	22	25	28
4	21	22	23	2	3	3	27	29	31
5	20	21	21	2	2	2	25	27	27
6	21	22	23	1	2	2	26	27	29
7	21	21	23	1	1	2	25	26	28
8	19	20	21	1	1	1	22	24	25
9	12	12	13	0	0	0	14	15	15
10	13	14	14	1	1	1	15	16	17
11	17	16	17	1	1	1	19	19	19
Simple average across grades *	17.9 (3.25)	18.7 (3.56)	19.9 (3.93)	1.5 (1.08)	1.8 (1.40)	2.2 (1.93)	21.6 (4.54)	23.6 (5.08)	25.1 (5.99)
Weighted average across grades **	18.0	18.8	20.1	1.8	2.2	2.7	21.6	23.7	25.1
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test									

Probability > |t|: * =0.05, ** =0.01, *** =0.001

*S.D. in parentheses

**Weighted by numbers of students taking tests in each category and grade.

Table A-3: State % scoring at or above national median in math and reading, 1997-1998, 1998-1999, and 1999-2000

Math	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	43	49	57	26	33	40	49	57	65
3	40	48	56	21	28	37	46	57	65
4	39	44	51	17	21	27	46	52	60
5	41	45	50	16	19	24	48	54	59
6	46	50	55	18	22	26	53	59	63
7	42	45	48	14	16	19	48	52	55
8	42	45	48	13	15	17	48	52	55
9	47	48	51	19	19	21	51	55	57
10	41	44	46	17	20	21	44	49	50
11	43	45	47	19	22	23	46	50	51
Simple average across grades *	42.4 (2.50)	46.3 (2.21)	50.9 (3.90)	18.0 (3.68)	21.5 (5.40)	25.5 (7.51)	47.9 (2.56)	53.7 (3.27)	58.0 (5.37)
Weighted average across grades **	42.4	46.4	51.2	18.6	22.7	27.6	48.0	53.8	58.0
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test	**	**	***			**	***	*	***

Probability > |t|: * =0.05, ** =0.01, *** =0.001

Reading	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	40	44	49	15	19	25	48	56	61
3	38	41	44	9	12	15	47	53	57
4	40	41	45	9	11	13	49	53	56
5	41	42	44	8	9	10	50	53	55
6	42	44	46	7	9	10	50	54	55
7	44	44	46	7	7	9	52	53	55
8	46	47	49	7	8	9	53	57	58
9	34	34	35	3	3	4	39	41	41
10	32	33	34	3	3	3	36	38	39
11	36	35	36	4	4	4	40	41	41
Simple average across grades *	39.3 (4.37)	40.5 (4.84)	42.8 (5.67)	7.2 (3.55)	8.5 (4.86)	10.2 (6.51)	46.4 (6.06)	49.9 (7.02)	51.8 (8.13)
Weighted average across grades **	39.4	40.7	43.1	8.1	9.9	12.1	46.5	50.2	51.8
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test									

Probability > |t|: * =0.05, ** =0.01, *** =0.001

*S.D. in parentheses

**Weighted by numbers of students taking tests in each category and grade.

Table A-4: State % scoring above 25th percentile in math and reading, 1997-1998, 1998-1999, and 1999-2000

Math	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	65	71	77	50	58	65	70	77	83
3	64	71	77	45	55	64	70	78	83
4	62	66	73	39	46	54	69	74	80
5	61	65	70	35	41	48	68	73	78
6	67	71	74	40	46	51	73	78	81
7	65	68	71	37	41	45	71	75	77
8	64	68	71	34	38	42	70	75	77
9	71	73	75	48	50	53	75	78	80
10	68	70	71	46	50	50	71	75	75
11	66	68	69	43	48	49	69	73	73
Simple average across grades *	65.3 (2.91)	69.1 (2.51)	72.8 (2.86)	41.7 (5.53)	47.3 (6.31)	52.1 (7.43)	70.7 (1.96)	75.6 (2.01)	78.7 (3.30)
Weighted average across grades **	65.2	69.1	73.0	41.8	48.2	54.0	70.7	75.7	78.8
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test	**	**	***	*		**	***	*	***

Probability > |t|: * =0.05, ** =0.01, *** =0.001

Reading	All students			LEP students			Non-LEP students		
Grade	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000	1997-1998	1998-1999	1999-2000
2	61	66	71	33	42	50	70	76	81
3	59	63	68	28	34	41	69	76	80
4	62	64	68	29	34	38	72	76	80
5	63	64	67	26	29	32	73	76	78
6	66	68	70	28	32	34	75	78	80
7	66	67	69	24	27	30	75	77	79
8	70	72	73	29	31	34	78	81	82
9	59	60	62	16	17	19	65	69	70
10	55	55	56	12	13	14	61	64	64
11	62	61	62	17	19	20	67	69	69
Simple average across grades *	62.3 (4.27)	64.0 (4.71)	66.6 (5.13)	24.2 (6.86)	27.8 (8.95)	31.2 (10.97)	70.5 (5.06)	74.2 (5.16)	76.3 (6.24)
Weighted average across grades **	62.3	64.1	66.8	26.0	30.3	34.4	70.5	74.4	76.3
Change Significance	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000	1998 to 1999	1999 to 2000	1998 to 2000
T-test			*			*			*

Probability > |t|: * =0.05, ** =0.01, *** =0.001

*S.D. in parentheses

**Weighted by numbers of students taking tests in each category and grade.

Table A-5. Math and Reading Test Score Quintile Ranges for the Percentage of non-LEP Students Scoring at or above National Median, by grade-span, 1997-1998

Math Grade-span	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
K-6	0 to 25.9	26.0 to 37.2	37.3 to 49.4	49.5 to 63.2	63.3 to 100
6-8	0 to 28.2	28.3 to 39.5	39.6 to 51.3	51.4 to 64.7	64.8 to 100
9-12	0 to 31.3	31.2 to 43.1	43.2 to 52.7	52.8 to 62.9	63.0 to 100

Reading Grade-span	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
K-6	0 to 24.6	24.7 to 37.0	37.1 to 49.7	49.8 to 63.6	63.7 to 100
6-8	0 to 31.7	31.7 to 43.9	44.0 to 56.3	56.4 to 67.4	67.5 to 100
9-12	0 to 22.8	22.8 to 33.2	33.3 to 42.3	42.4 to 52.4	52.5 to 100

[1] Linn, Robert L., 2000. "Assessment and Accountability." *Educational Researcher*, Vol.29, No. 2, pp. 4-16.

[2] Betts, Julian R. and Robert M. Costrell, "Incentives and Equity under Standards Based Reform," in Diane Ravitch (ed.), *Brookings Papers on Education Policy 2001*, (The Brookings Institution, 2001).

[3] Sonstelie, Jon, Eric Brunner, and Kenneth Ardon, *For Better or Worse? School Finance Reform in California*, (San Francisco, California: Public Policy Institute of California, 2000).

[4] See Hanushek, Eric A., "School Resources and Student Performance", in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, (Brookings Institution, 1996), and Betts, Julian R., "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an .Old Literature," in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, (Brookings Institution, 1996).

[5] Betts, Julian R., Kim S. Rueben, and Anne Danenberg, *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California*, (San Francisco: Public Policy Institute of California, 2000).

[6] ibid.

[7] For evidence, see Betts and Danenberg, "Resources and Student Achievement: An Assessment", in Jon Sonstelie and Peter Richardson, (Eds.), *School Finance and California's Master Plan for Education*, (San Francisco: Public Policy Institute of California, 2001).

[8] The next three paragraphs are drawn in large part from the California Department of Education website.

[9] For example, the content standards for math stipulate in Item 2.1 under Algebra and Functions that in grade 3 students should be able to: "Solve simple problems involving a functional relationship between two quantities (e.g., find the total cost of multiple items given cost per unit)." (California Department of Education, *Mathematics Content Standards for California Public Schools, Kindergarten through Grade Twelve*, (Sacramento, 1999a).) The more detailed mathematics framework gives the following example of this content item: "John wants to buy a dozen pencils. One store offers pencils at 6 for $1. Another offers them at 4 for 65 cents. Yet another sells pencils at 15 cents each. Where should John purchase his pencils in order to save the most money?" (California Department of Education, *Mathematics Framework for California Public Schools*, *Kindergarten through Grade Twelve*, (Sacramento, 1999b). The entire content standards and subject frameworks can be found at http://www.cde.ca.gov/ci/.

[10] For a complete discussion of the API calculation in 2000, see the California Department of Education website http://www.cde.ca.gov/psaa/api/yeartwo/base/apicalcb.pdf

[11] The legislation provides for an "alternative accountability system" for schools with fewer than 100 students, alternative and continuation high schools, independent study programs, community day schools and other schools that are excluded from the API.

[12] See Article 2, Section 52052 (c) of the PSAA.

[13] SB 1552 added an additional year of funding for schools making significant progress towards their growth targets without actually meeting them.

[14]Side-by-Side Comparison of PSAA Award Programs: (http://www.cde.ca.gov/ psaa/awards/compare.htm [4/25/01])

[15] Kathleen Seabourne, CDE Consultant. April 25, 2001. Personal communication clarifying website language.

[16] Associated Press, "Qualified Teachers Sought for State's Worst-Off Schools: Legislative Bills Offer Incentives"," *San Diego Union Tribune* (April 19, 2001).

[17] Golden State Examination website: (http://www.cde.ca.gov/statetests/gse/gse.html [4/25/01]).

[18] Jay P. Heubert and Robert M. Hauser, Editors; Committee on Appropriate Test Use, National Research Council. 1999. "High Stakes: Testing for Tracking, Promotion, and Graduation." National Academy Press: Washington D.C.

[19] CDE website: http://www.cde.ca.gov/sfsdiv/budgetact/implementation300.htm [4/25/01]

[20] For the difference between 1998 and 2000 test score means we are actually comparing a 2-year change.

[21] Koretz, Daniel, "Using Student Assessments for Educational Accountability," in Eric A. Hanushek and Dale W. Jorgenson, eds., *Improving America's Schools: The Role of Incentives*, (Washington, D.C.: National Academy Press, 1996).

[22] The exception was math scores for non-LEP students where the gains in the share of students in the top quartile very slightly outweighed the decrease in the share in the bottom quartile.
[23] We thank Lorrie Shepard for this third way of testing for regression to the mean.
[24] Also, note the low-scoring schools do not display a decline in achievement in table A-6, as predicted in our hypothetical model in Figure 7 under the assumption that regression to the mean without any true change in test-score gaps was at play. While an important finding, it is important to remember that Figure 7 implicitly assumed no real growth in average achievement over time. If there has been growth in average achievement, then we would not necessarily expect low-scoring schools as defined at the end of the period to show declining test scores, but rather, slower growth than high-scoring schools.
[25] Robert M. Costrell, "A Simple Model of Educational Standards," *American Economic Review* 84 (4), 1994, 956-971; Julian R. Betts, "The Impact of Educational Standards on the Level and Distribution of Earnings," *American Economic Review* 88 (1), 1998, 266-275.
[26] Julian R. Betts, **A Critical Path Analysis of California's K-12 Sector**, (Irvine: California Council on Science and Technology, forthcoming).
[27] Analysis of decennial Census data for California by Betts (2000) finds that between 1970 and 1990 the percentage of 19-24 year old natives living in California who lacked a high school diploma was about 16.5%, while among immigrants in California in the same group the percentage rose from 40.6% in 1970 to 50.3% in 1990. Further, in 1990 among young California immigrants, 58.7% who had lived abroad in 1985 were high school dropouts compared to 46.7% of those who had lived in California in 1985. Overall, this suggests that the true dropout rate among teenagers who attend school in California is around 20-25%, given that by 1990 26.2% of adults in California were immigrants. See chapters 2 and 3 of Julian R. Betts, **The Changing Role of Education in California**, (San Francisco: Public Policy Institute of California, 2000).
[28] Betts, Rueben, and Danenberg (2000).
[29] The proposal would also eliminate the Stanford 9 social science test current given to high school students. See Sarah Tully Tapia, "Davis Proposes Streamlining Tests", *Orange County Register*, April 25, 2001.
[30] See Julie Tamaki, "Davis Scales Back Plan to Lengthen School Year", Los Angeles Times, April 25, 2001.
[31] California Department of Education site – http://goldmine.cde.ca.gov downloaded 3/24/99.
[32] Some 6th grade students are in elementary schools, and some are in middle schools. Because approximately half of elementary schools are K-5, approximately half are K-6, and the rest comprise some combination of grades between Kindergarten and 6th grade, our elementary school category comprises grades K-6. The majority of middle schools are 6-8, whereas others are 7-8, and a few are single-grade. Sixth-grade-only schools were included in K-6 schools rather than 6-8 schools.